THE DEMAND SIDE OF THE SHADOW ECONOMY: ESSAYS ON INFORMAL CONSUMPTION

ALVARO GERMÁN MONGE ZEGARRA

DOCTOR OF PHILOSOPHY (Ph.D.) in Economics

Department of Economics
University of Sussex

May 2015

UNIVERSITY OF SUSSEX

ALVARO GERMAN MONGE ZEGARRA

DEGREE OF DOCTOR OF PHILOSOPHY (Ph.D.) IN ECONOMICS

THE DEMAND SIDE OF THE SHADOW ECONOMY: ESSAYS ON INFORMAL CONSUMPTION

SUMMARY

The focus of this thesis is the study of the demand side of the shadow economy. To achieve this, the informal consumption of Peruvian families located in urban areas is studied. This is possible thanks to a household survey collecting information on where people acquired their goods. The main contribution of the research is that it identifies an unexplored area in the literature with limited theoretical discussion and few empirical applications. Information about why people purchase from informal markets will supplement wider knowledge of labour allocation on informal opportunities. The thesis uses an Almost Ideal Demand System in order to verify some demand properties of informal consumption: income and price elasticities, the existence of linkages between working and purchasing decisions and explore the effects of bargaining on expenditure allocation. Four robust results are encountered. First, the inferiority of informal consumption is rejected. Formal and informal expenditure are classified as normal, but income responses on the latter (necessity) are lower than on the former (luxury). Second, there are linkage effects between working and purchasing in the informal and formal sectors. These effects are stronger for informal consumption and among the self-employed. Linkages are also not equally applicable across all goods. Better results are found within quasi-substitutes with leisure. Third, formal and informal food consumption reveals elastic demand curves and imperfect substitution between them, with higher compensated own-price and cross-price elasticities for formal markets. Fourth, household members bargain in their allocation decisions across markets, with females' decisions being closer to less-informal purchasing baskets. This result is clearer in the case of food consumption. Public policy recommendations based on these results are derived, where it is found that formalization policies will need to take into account their negative distributional effects.

ACKNOWLEDGEMENT

There are so many people whom I need to thank for accompanying me in the completion of my Ph.D. thesis. Of course, this thesis would have been impossible without the encouragement, advice and help, in so many ways, of my supervisor, Prof. Andy McKay. He was always ready to discuss with me the details of my research from both empirical and theoretical points of view at every stage of my dissertation. I never doubted for one minute his willingness to help, even when I left Sussex and moved to Peru. I will never forget our long phone and Skype talks, where time difference was never an obstacle. His wise guidance and friendship are the main reasons why I find myself at the end of this long journey. Also, I want to thank Prof. Richard Dickens and Dr. Paolo Massela for their comments and advice at different stages of the thesis; Prof. Barry Reilly, who helped me to decide on the topic of my dissertation and who further helped me on several econometric issues; Jenny Money who corrected my awful English; and Prof. Robert Eastwood, who encouraged me to undertake this Ph.D. I also want to thank my Ph.D. friends during my period in Brighton (Max, Javier, Antonio and Jairo). With them, a stressful time was always better managed with a laugh, a drink, a cigarette or a football match!

When I moved to Peru and started a full-time job things became more complicated. However, I managed thanks to the help of my friends at Macroconsult. They supported me in so many ways, giving me enough time to work in my thesis and discuss many of my preliminary results. I must also thank my professor at the Universidad del Pacífico, Dr. Enrique Vasquez. He has been a prime source of support throughout most of my academic career. I also want to thank two of my best friends, Diego and Oswaldo. During our time in the UK and Lima we found the way to mutually encourage each other in the completion of our studies.

I also want to thank my family. My mother (Dora Luisa), who is responsible for every step I have taken in my live. She simply always put her sons first and decided to fight for us, even when everything was against her and it would have been easier to give up; my father (Francisco), because he taught me through his example; Oscar, because he was always there; my brother (Francisco) for his friendship; my grandmother (Dora) for her unconditional love and my grandfather (Germán), because he was my biggest inspiration and my best friend. Finally and most importantly, my wife (Carlita), because you are my definition of happiness. I cannot imagine life without you (and, of course, our children). Your love, your friendship, your smile, even in the most difficult circumstances, gave me the strength to start, to continue and to finish. I know that I will never be able to give you back the time my research has stolen from us. And I know that nothing will ever be enough, but let me just say that this thesis is simply dedicated to you.

CONTENTS

LIST OF TABLES

CHAPTER IV

CHAPTER V

CHAPTER VI

LIST OF FIGURES

LIST OF ABREVIATIONS

(in alphabetical order)

2SLS	Two Stage Least Squares
AIC	Akaike Information Criterion
AIDS	Almost Ideal Demand System
BIC	Bayesian Information Criterion
CC	Clothing and personal care
C-IV	Conditional Instrumental Variable estimator
CV	Compensating Variation
ED	Education and cultural goods and services
ENAHO	Encuesta Nacional de Hogares
ENAPROM	Encuesta Nacional de Propósitos Múltiples
EPP	Extra-household Environmental Conditions
ESSALUD	Seguro Social de Salud
Foff	Food to be consumed out of the house
Fon	Food to be consumed in the house
GDP	Gross Domestic Product
GMM	Generalized Method of Moments
GNI	Gross National Income
GRG	Generalized Reduction Gradient
HDI	Human Development Index
HEA	Health goods and services
ILO	International Labor Organization
INEI	Instituto Nacional de Estadística e Informática
IV	Instrumental Variables
LIML	Limited Information Maximum Likelihood
LogL	Log-likelihood
MINEDU	Ministerio de Educación
MINSA	Ministerio de Salud
OLS	Ordinary Least Squares
OT	Other goods
PIGLOG	Price Independent Generalized Logarithmic
PP	Percentage points
PPP	Purchasing Power Parity
REIR	Régimen Especial de Impuesto a la Renta
RENAMU	Registro Nacional de Municipalidades

RUC	Registro Único de Contribuyentes
RUS	Registro Único Simplificado
S&Y	Shonkwiller and Yen (1999) estimator
S/.	New soles
SD	Standard Deviation(s)
SUNAT	Superintendencia de Administración Tributaria
SUR	Seemingly Unrelated Regression
SUR–GLS	Seemingly Unrelated Regression adjusted for heteroscedasticity
TC	Transport and Communication
TCOL	True Cost of Living
UNDP	United Nations Development Program
USA	United States of America
US$	USA Dollars
VAT	Value Added Tax
VCE	Variance Covariance Matrix

I. INTRODUCTION

Recent estimations by Schneider and Buehn (2007) reveal that the Peruvian economy is one of the most informal economies around the world. According to their comparative study, the size of the shadow economy in 2006 was around 60% of GDP, with figures that have changed only slightly during the last decade. Looking at the ranking constructed in that study for a total of 120 developed and developing countries, Peru is one of the top five informal countries, with higher figures shown only in the cases of Georgia (69%), Bolivia (68%) and Panamá (65%). The consequences of informality in the economy have been investigated by Loayza (2007). Using cross-country evidence for developed and developing countries, Loayza concludes that increasing one standard deviation in the indicators of informality produces a reduction of between 1 and 2 percentage points in GDP growth. Using a similar approach, but restricted only to Latin American countries, Loayza (1997) found similar results.

Loayza explains that the main drivers of the negative correlation between informality and economic growth are the sub-optimal allocation of production factors, the congestion of public infrastructure and lower tax payments reducing the quality of public goods and service provision. Perry *et al*. (2006) also attribute lower incentives to invest to informality and, in general, a reduction in the productivity ratio. Other authors have found additional negative impacts of the informal economy, mainly in terms of higher corruption levels (Schneider and Enste, 2000) and negative welfare effects (Perry *et al*., 2007). In this last case, the authors indicate that, in highly informal countries, most households remain unprotected from economic and other external shocks, family incomes are more volatile and earning sources less secure (see also Ferrer-i-Carbonell and Gërxhani, 2004). When the impacts of informality on household welfare are tested, a positive correlation is generally found between working in the informal sector and self-rated poverty (Ferrer-i-Carbonell and Gërxhani, 2004; Beuran and Kalugina, 2005; Arias and Lucchetti, 2007).[1]

Therefore, it is possible to recognize social and individual costs attached to informality, exacerbated in those developing countries where informal ratios tend to be extremely high. In this context, it is not surprising that, over the last fifteen years, and once democratic regimes return to

[1] Most of these results are driven not only by the earnings gap between formal and informal opportunities, but also by non-pecuniary factors like worse working conditions in the informal sector or less access to social benefits (Ferrer-i-Carbonell and Gërxhani, 2004). However, it is difficult to generalize as long as most of these results are country-specific and depend on particular economic contexts and the type of informal job considered (Perry *et al*., 2007). In fact, depending on the informal earning opportunity under analysis, positive welfare effects associated with informality could be encountered, such as more flexible schedules or financial advantages, given that there is generally no payment of most regulation costs like taxes or contributions (see also Ferrer-i-Carbonell and Gërxhani, 2004 and Perry *et al*., 2007).

the Peruvian government, the formalization of economic activities in the country has been at the top of the public agenda.[2] However, the results are far from satisfactory. Although in that period the economy has seen growth at sustainable levels (6.2% on average)[3] and this expansion has partly resulted in a reduction in informality, as reported in Chacaltana and Infante (2014), the Peruvian economy is still characterized by severe productivity gaps where small-scale and highly informal firms are still the dominant sector.

One possible explanation for this underperformance is the limited number of policy tools available and the incomplete understanding of informality as an economic phenomenon. As long as informality has been seen only as a problem on the supply side of the economy (why formal firms form or why people work in the informal sector), little scholarly attention has been paid to the demand considerations around it (Fortin *et al*., 2000; Pisani, 2013a). However, the only reason for an informal firm to exist is because it is selling its products, maybe some of them to other informal and formal firms (Böhme and Thiele, 2012a), but most probably to consumers.[4] The question overlooked by the empirical literature and public policy design is why households decide to purchase from informal or formal outlets. The research presented here is an attempt to fill this gap, focusing on the study of the demand side of the shadow economy in order to empirically explore the main theoretical properties of informal consumption and derive complementary public policy recommendations oriented towards the formalization of economic transactions in developing countries. From the literature review undertaken during this research, this is one of the first investigations that uses formal demand theory in the study of informal consumption. This is the major contribution of this thesis.

The closest precedents encountered are Fortin *et al*. (2000), Gardes and Starzec (2002, 2009), Reilly *et al*. (2006), Böhme and Thiele (2012b) and Pisani (2013b), where the main focus of the research was testing specific hypotheses around informal consumption using a demand framework. In general, the focus of attention of these studies has been on the income effects on consumption allocation between formal and informal markets, the quantity effects of labour supply (formal and informal) on the same allocation process, the relationships between formal and informal consumption and the effects of demographic variables on market participation. These applications are carried out in sub-regional contexts like the South Texas borderlands in the case of Pisani (2013b), in country studies in Canada, Poland and Serbia in the cases of Fortin

---

[2] Taken from the presidential discourse of Alejandro Toledo, Alan García and Ollanta Humala at the beginning of their 5 year mandate available at www.congreso.gob.pe.
[3] See www.bcrp.gob.pe
[44] Böhme and Thiele (2012a) provide some estimates for African countries and reveal that around 80% of informal firms sell to households. Also, Dasgupta (1992) and de Soto (1986) comment that an important destination of informal commerce is household consumption.

et al. (2000), Gardes and Starzec (2002, 2009) and Reilly *et al*. (2006), respectively, and in cross-country evidence for four African countries in the case of Böhme and Thiele (2012b). The thesis presented here departs from these studies, and extends them in several ways (see below).

As explained above, the focus of attention is Peru, one of the most informal countries in the world. The specific questions to be answered during the research are related to the main drivers of informal consumption. First, an in-depth exploration of expenditure effects is provided in order to derive the income elasticities that characterize market allocation. The objective is to verify if informal consumption could be classified as normal or inferior (a question which, right across the limited literature, has produced mixed results when tested) and see whether different responses to income shocks between formal and informal expenditure can be identified. The results in this case will be crucial for policy design, since it will inform how the resource allocation of households responds to economic cycles – a key aspect to be considered when other actions are undertaken in order to formalize the economy or protect families from recessive periods. Also, in line with Böhme and Thiele (2012b), it will be possible to inspect quality differences between formal and informal consumption using their income elasticities (with lower values in the case of informal markets if quality there is lower). Using this evidence, as stated by the authors, it will be possible to conclude whether informal goods are constrained from the demand side in the sense that the demand of goods using informal distributional channels is income-inelastic (i.e. mainly bought by the poor).

Second, the existence of linkages between working in the informal sector and purchasing from informal markets is also tested. There are reasons to believe that this linkage exists. As originally noted by Fortin *et al*. (2000), informal workers might have better and more timely information about the supply or quality of informal goods increasing the likelihood to purchase there. Also, as other authors have noted using observational studies (to be presented later), informal workers may have strong reasons for preferring informal transactions if they want to remain hidden or if they have characteristics that facilitate these kinds of transaction: eventuality in their incomes, payments in cash and so on. If these linkages exist, then public policies that focus on the formalization of the supply side of the economy acquire an additional meaning, given their potential effect on the demand side, generating second-round effects to be accounted for. The limited empirical research investigating this issue in developed and European transitional economies concludes that both activities are in fact linked. However, other efforts in African countries conclude that no effect is encountered, suggesting that the existence of such a linkage is possibly country-specific. The added value proposed in this research is to re-test the hypothesis in a different context in order to contribute to the construction of an empirical regularity. Also, in contrast to previous studies, the estimations presented here allow for some heterogeneity in the working options of the family members. Instead of using aggregated formal and informal labour supply in the empirical model, five earning opportunities are implemented (formal and informal in the wage-earning or self-employment sectors) in order to inspect in which labour segments linkages could be stronger or weaker.

Third, own-price and cross-price elasticities for formal and informal food consumption are estimated. The objective here is to verify the degree of responsiveness of formal and informal consumption to price shocks and to compute the substitution effects between them. The literature review has revealed that this is the first time that a result like this is provided and, as recognized by Reilly *et al*. (2006) in their suggestion for further research, this is a key point in the understanding of informal demand that must not be overlooked. Own-price responses and the degree of substitutability among consumption profiles are key elements in the design of any policy intention that tries to take advantage of the incentive structure of economic agents.

Finally, the effect of bargaining power inside the family in terms of market allocation is the last empirical question. For this purpose, informal consumption is modelled using a collective utility approach. The objective is to verify whether some members' expenditure is closer to formal or to informal consumption in order to exploit redistribution policies in the formalization objective of the government. There are reasons to believe that a result like this is plausible. The research undertaken using bargaining demand models (see, for example, Haddad *et al*., 1997) and the anthropological traditions (some of which are quoted in Hoddinott and Haddad, 1995), have attributed to mothers' better expenditure patterns in terms of the type of goods they prefer (food, schooling, child health). This research takes these results as its point of departure and asks whether better expenditure means not only the type of goods consumed, but also the sources used to purchase them. If quality is also a concern, mothers´ expenditure could be less biased towards informal markets. From the literature review it is possible to conclude that this is the first time that a question like this is answered.

The data to be used in the thesis come from the National Household Survey (ENAHO, in Spanish) for the year 2006, conducted by the National Statistics Office (INEI, in Spanish), restricted to urban households in order to obtain a better identification of the main variables of interest. In fact, the main property of this database is that it contains detailed information on consumption (which goods the family purchases and where the expenditure takes place), detailed information on labour supply (the condition of informality of the worker and the firm where he or she works), detailed information on incomes (who specifically receives income from which source) and permits the computation of prices (or unit values) for food consumption. To construct expenditure variables, the information on outlets was exploited in order to classify them between informal and formal markets. Therefore, the approach is closer to Böhme and Thiele's (2012b) and Reilly *et al*.'s (2006) identification of informal channels of commercialization than to those of Fortin *et al*. (2000), Gardes and Starzec (2002, 2009) and Pisani (2013b), who use responses on unregistered purchases made by households.

However, one critical point must be mentioned: informal activities are, by definition, hidden, so data construction is not free from arbitrary choices made when both labour supply and consumption were classified between formal and informal sectors. The key point was to answer the question as to which signals are the data giving in order to infer where an informal relation is arising or among which agents could be clearer. Experience of the particular economic framework being analysed, the previous literature reviewed, alternative data sources and anecdotal evidence collected during this research helped to make better choices, but it was impossible to ensure the exact ones. Thus, important issues to consider were to make explicit the assumptions, recognize the degree of precision and control for possible deviations changing the definition of informality (when possible). This is what has been undertaken in the thesis.

The document presented here is organized as follows. In Chapter 2, a summary of the literature reviewed during the research is presented in order to provide an adequate theoretical framework for the rest of the thesis. This chapter is further divided into three main sections. The first discusses the definitions of informality to be adopted during the research. As mentioned there, informality will be understood at the level of market relationships maintained by agents lacking adequate institutional enforcement. The second briefly presents the most representative literature on the supply side of the shadow economy. The objective there is to understand the main stylized facts around firms' formation and labour allocation that will help in the interpretation of the empirical results to be presented in subsequent chapters. The third section is the main focus of this chapter and contains a discussion of the literature on informal consumption. As shown there, this literature is scant in terms of formal models and empirical applications. Most of the research has been undertaken based on qualitative evidence and observational studies and these constitute the main sources of stylized facts to be considered during the research.

In Chapter 3, based on the literature review, a simple theoretical model for informal consumption is proposed in order to formally present and justify the hypothesis to be tested. The objective is to study three main demand properties of informal and formal consumption: expenditure effects, own-price and cross-price effects and quantity effects related to labour supply (both formal and informal). The structure of the model responds to a partial equilibrium short-run unitary maximization process under two different budgeting assumptions and preferences' ordering. In the first one, the household allocates its resources between formal and informal markets, facing transactional costs in each and conditioned to budget constraints and predetermined leisure time. This model is referred to as the aggregate version, since it produces two conditional demand functions for estimation purposes: one for overall formal market consumption and the other for overall informal market consumption. In the second one, a two-stage maximization process is considered for households deciding in the first stage which goods they will purchase and, in the second, where these goods will be bought (formal or informal markets). This model is referred to as the disaggregated version and will produce several demand systems (for informal and formal consumption), one for each good (or consumption group) considered.

In Chapter 4, the database used for estimation purposes is presented, and the general characteristics of the main sample as well as the complementary data sources included are considered. At the same time, in this chapter the assumptions used in the construction of the main variables of interest to be used during the estimation are discussed: expenditure, prices and labour supply. The chapter also provides some preliminary inferences, using descriptive statistics obtained from the main variables.

In Chapters 5 to 7 the model presented in Chapter 3 is estimated. For that purpose, given its attractive empirical and theoretical properties, the Almost Ideal Demand System (AIDS) of Deaton and Muellbauer (1980a) is used as the main specification for fitting the demand equations. In Chapter 5, the estimations for the aggregate version of the model are discussed. The objective here is to study both income effects and quantity labour effects in a context of overall formal and informal consumption allocation. This application is the closest to the previous empirical applications encountered in the literature, so comparisons with Fortin *et al*. (2000), Gardes and Starzec (2002), Reilly *et al*. (2006) and Böhme and Thiele (2012b) could be carried out. Two empirical strategies are followed. The first considers a Seemingly Unrelated Regression (SUR) model with robust (White/Huber) standard errors, referred to as the SUR–GLS estimator under the assumption of exogenous expenditure and labour supply. Given the simplicity of the model, most of the robust checks in terms of specification issues, sub-population estimations and definitions of informality are done. The second empirical strategy drops the assumption of exogeneity of the main variables of interest (expenditure and labour supply) and estimates several IV models. These estimations are more austere in their specification and are specifically constructed to re-test the hypotheses of interest in a better empirical context. The chapter then presents the main implication using policy simulations of the effect of income and labour supply on the demand of formal and informal markets.

In Chapter 6, the estimations for the disaggregated model are presented and the same effects as earlier are inspected, but for specific consumption groups. The objective is to analyse whether the income effects prevail when detailed consumption profiles are considered (as was the case in Böhme and Thiele, 2012b) and to verify whether the linkage hypothesis is more plausible in some types of consumption than in others (as suggested for further research in Fortin *et al*., 2000). In this case the main econometric problem faced was the relatively high incidence of zero-consumption values in most expenditure categories. In order to control for possible censoring bias using system estimation, the procedure of Shonkwiller and Yen (1999) is used for estimation, based on a first-step PROBIT model and a second-step SUR estimation for the whole sample, with correctional terms included. The within-standard errors are computed using robust techniques (White/Huber) and the between-equations variance covariance is computed with the algorithm developed by Weessie (1999). This is henceforth referred to as the S&Y estimator. This procedure helps to construct the conditional and unconditional elasticities for income and hours of work, considering the parameter estimates of both steps using the formulas provided by Su and Yen (2000) and Carpentier and Guyomard (2001). Policy simulations are carried out and compared with those in the previous chapter.

In Chapter 7, prices (or unit values) are included in the analysis. The model is restricted to food consumption, which is the expenditure group where unit values were identifiable in the survey. Again, the main econometric problem handled was censoring, so the S&Y procedure is again used. The model estimates the Slutsky matrix for different definitions of prices and different ways to impose the theoretical demand restrictions. Then both compensated and uncompensated price elasticities are computed. The analysis again provides policy simulations in order to verify the effectiveness of formalization for different price-shock combinations. The welfare effects of these policies are investigated using compensated valuation techniques.

In Chapter 8 the assumption of unitary maximization process used in previous chapters is dropped in order to test the influence of bargaining power on market allocation. The final specification of the model follows the collective demand framework used by Chiappori's (1988, 1992) and Bourguignon *et al*.'s (1993) theoretical insights around sharing rules. Three empirical strategies are followed. In the first, as undertaken by Lancaster *et al*. (2006) and Monge (2007) for consumption-groups systems, a theoretically plausible non-linear specification for bargaining power is estimated considering a SUR–GLS technique under the assumption of exogenous total expenditure and bargaining power. This assumption is dropped in the second empirical strategy, where the specification is simplified to a less theoretical, but more tractable linear version, as Hoddinott and Haddad (1995) for consumption groups. In this case, standard IV estimation methods are used to re-test the hypothesis. The third strategy replicates the first, but for the disaggregated version of the model, in order to inspect the effects of bargaining inside consumption groups.

In Chapter 9 the main conclusions of the thesis are summarized and policy implications discussed in a broader way. Similarly, the main limitations of the research undertaken are explicitly mentioned and the agenda for further research is outlined.

II. LITERATURE REVIEW

II.1. Introduction

In this chapter the literature review on informal consumption is presented in order to define the theoretical framework for the research. The discussion begins with a brief presentation of the definitions around informality, focusing on the concepts of both exit and exclusion that explain the emergence of this phenomenon in developing countries. Next, the revision of the literature on informal labour and goods markets is considered. In the first case, the objective is not to yield a complete review of the huge discussion around informal labour markets produced throughout decades of research. On the contrary, only some representative models and theoretical approximations are revised in order to take into account the main stylized facts. The focus is placed on the second case, where the literature on informal consumption is presented using the limited theoretical and empirical discussions around the issue and the larger insights provided by the qualitative research. The chapter finishes with some concluding remarks.

II.2. What is informality?

A distinct aspect in developing countries is the existence of a broad spectrum of informal transactions.[5] These are economic relationships maintained by individuals outside the legal or established markets, where traditional explicit norms no longer hold and are replaced by implicit regulations or conventions. In the words of Feige (1990), informality can be understood as 'those actions of economic agents that fail to adhere to the established institutional rules or are denied their protection'; in those of Castells and Portes (1989), informality includes 'activities that are not regulated by the state in social environments where similar activities are regulated' (taken from Centeno and Portes, 2006). Using this last definition, then, it is possible to argue that any formal market has its informal parallel in some sense. Informal labour markets, informal goods markets and informal insurance or credit markets are some of the examples encountered in the literature.

Following Thomas (1992) and Tokman (2000), the first characterization of the informal economy was provided by Hart (1970) and the social scientists of the International Labour Organization (ILO, 1972). For these authors, the concept of informality was introduced to deal with the particularities observed in the labour markets in developing countries. In general in these

[5] In this thesis the terms informal, shadow, irregular, unregulated, underground, unregistered and undeclared economy are all used as synonyms to describe the phenomenon being analysed: informality. However, from a theoretical point of view, they can refer to different aspects of it (see Thomas, 1992 or Portes, 2000).

countries, in the absence of unemployment benefits, few can afford to be unemployed, so most of them will need to find alternative ways of earning money, not necessarily via the production or distribution of illegal goods. Some of these alternative ways of earning money were found in economic activities that avoid recognition, regulation and protection from the government. In particular, rural migrants to the cities who experienced difficulties engaging with the formal economy, found a way of life in this sector.

However, as argued by de Soto (1986), this conceptualization of informality has a pejorative connotation, seeing it as the result of some kind of exclusion process. For de Soto this is incorrect; he prefers to explain the emergence of informality as the natural response of (possibly poor) individuals to a state unable to satisfy their demands and ambitions. As explained by Portes (2000), in de Soto´s view, informality is seen as a rational decision and a genuine expression of market forces gaining space in an over-regulated context, with governments incapable of enforcing regulations. As a result, informality is no longer just a means of subsistence; it is also a way to accumulate capital in developing countries outside the legal framework.

Perry *et al*. (2007) recognize the validity of both approaches and propose an intermediate (or more general) interpretation, preferring to call informality a multidimensional phenomenon where agents interact with the government in some dimensions and not in others, thus creating a large grey area between compliance and non-compliance. Furthermore, compliance in some cases will be determined by exit strategies (the rational choice of agents given the high costs or lower benefits of formality) and in others by exclusion patterns (where access is explicitly or implicitly denied). For that reason, if all the possible relationships maintained by firms, workers and consumers are considered, it is difficult to think in terms of a homogenous, well-defined informal sector. On the contrary, it is better conceptualized as a web of informal relationships that emerge in a particular market, socially determined, continuously changing and mixed up with the formal ones.

II.3. Informality in the labour market

In their original works, Harris and Todaro (1970) saw informality as a transitional phenomenon in the urban labour market, generated by the rural-urban migration implied in the first stages of development. The authors proposed the idea that rural workers' migration to the city (attracted by higher wages) produces an increase in urban labour supply. However, given some market imperfections (fixed wages set above the equilibrium level), only a fraction of these new workers find a job in the formal sector, while the others must accept some kind of informal arrangements. Although simple, the argument implies several unsatisfactory hypotheses. For example, in this context, informal jobs must be non-permanent, a result which is difficult to verify in reality, as noted by Sethuraman (1981), Rauch (1991) and Thomas (1992). Moreover, Williamson (1988) observes that both the little attention paid to informal labour market behaviour and the simplistic story of the rules embodied in the formal one, undermine the Harris–Todaro model as an adequate explanation of how the urban labour markets in developing countries actually work (see also Kannappan, 1985).

Given these difficulties, several authors have tried to formalize the emergence of the informal sector in the labour market. One such attempt was made by Rauch (1991), who extended the Lucas (1978) model of firm size distribution. Following his argument, the production function of the economy depends on entrepreneurial ability and there are both managers with different managerial abilities and homogeneous workers. Under universally enforced minimum wages (set above the competitive level) labour supply increases, labour demand falls and unemployment emerges. However, if this minimum wage is not universally enforced and can only be imposed on large firms (probably due to the government's lack of enforcement power) then small-scale firms offering wages below the minimum level (and below the clearing market wage level) will form. At the equilibrium, there is a group of managers with a high enough entrepreneurial ability who decide to work at the profit-maximizing level of labour, paying the minimum wage (the profits of producing at this level compensate the higher costs), while others who are less talented will not be able to afford the higher wage and will be producing at a low scale. The first group (workers and firms) constitutes the formal sector and the second group the informal one.

The structure of Rauch's model has proved to be flexible enough to include other stylized facts of the shadow economy without violating its original structure. For example, Fortin *et al*. (1997), Straub (2005), de Paula and Scheinkman (2006) and Galiani and Weinschelbaum (2007) introduce tax payments, evasion costs, benefits of producing in the formal sector (better access to credit markets or to public goods), penalties of being informal and the enforcement ability of the government. At the same time, Rogers and Swinnerton (2004) modified the behaviour of the labour market, switching to a 'search model', while Levenson and Maloney (1998) propose, from a dynamic perspective, a characterization of informality based on institutional participation. The diagnosis of the urban labour market provided by these extensions to Rauch's model amplified the scope of the analysis on the supply side of the economy and introduced the concept of net gains on participation for each earning opportunity.

The theoretical implication of Rauch's model is the emergence of two kinds of dualism in the urban labour market or, more precisely, in the supply side of the economy: size dualism and wage dualism. In the first case, small firms will be informal and large firms formal while, in the second case, formal workers will earn more than their similar informal counterparts. At the same time, the theoretical and empirical applications lead to the assumption that the emergence of informality is a problem only of entrepreneurial ability, where only a firm's choices matter and worker choices are not taken into consideration. This implies, for the labour market as a whole, a non-competitive (or segmented) structure where workers cannot choose their preferred allocation of time.

Arguments against this assumption were proposed originally by Rosenzweig (1988), who describes the labour market as competitive; however, given workers' characteristics, some will choose formal jobs and others informal. Similar ideas are developed by Amaral and Quintin (2006), who replicate most of the stylized facts of Rauch's model without the assumption of wage dualism. The general idea is that agents are heterogeneous in their managerial abilities and education, and choose between becoming formal or informal entrepreneurs, or workers. In the same way, workers can move freely between both sectors, so the wages are the same for similarly skilled workers in informal or formal firms. Other characteristics of the model are that formal firms pay taxes and obtain credit while informal ones do not. In the equilibrium model, more able managers operate with a higher capital:labour ratio (size dualism), hire skilled workers and choose to become formal (returns to external financing exceed the additional tax cost). Informal firms are credit-constrained and hire unskilled labour, since this is a better substitute for capital shortages.

Mixed evidence around the segmentation hypothesis[6] led Fields (1990, 2007) to argue over two narratives about urban labour markets: neither purely segmented nor competitive and possibly country-specific. In fact the author proposes to focus on two types of informal labour supply: 'upper-tier jobs' which are good jobs in the informal sector and which some workers will freely choose, and 'free-entry jobs', which are bad jobs in the informal sector which people choose only because of the difficulties in access to formal or 'good informal' positions. For the Peruvian case, Yamada (1996) found evidence in favour of Field's arguments, with the informal self-employment sector better described as a competitive one and the informal wage-earning sector better described as a segmented market. A similar result is commented on by Perry *et al*. (2007) for other Latin American countries.

More complex structures for exploring the relationship between formal and informal labour supplies are considered in Lacroix and Fortin (1992) and Lemieux *et al*. (1994) using an extended

[6] See Magnac (1991), Maloney (1999, 2004) and Pradhan and van Soest (1997).

version of the Cowell (1985) model for tax evasion. In this approach, agents choose to work in the regular or irregular sector and can hold down more than one job. The decision process is defined by tax avoidance attitudes but at a cost of risk detection. Then the individual follows some kind of strategic behaviour and different regimens arise in equilibrium: only formal, only informal, formal/informal and not working. The model also allows for constrained and unconstrained situations in access to formal labour markets. The empirical application strongly rejects perfect substitutability between formal and informal jobs, but supports the idea that labour supply in both sectors is negatively related; however, informal labour supply is more responsive to formal earnings than conversely.

II.4. Informality in the goods market

Despite some remaining debates, firm behaviour and worker behaviour have been modelled under most of the stylized facts of the shadow economy in a developing country. Thus, the emergence of informal firms or labour relations and the linkages between them and formal ones is not an under-researched area in the literature. However, as noted by Fortin *et al*. (2000), while these studies provide useful information on the supply side of the irregular sector; little is known about the factors influencing the demand for goods and services produced in that sector. In fact, the academic literature around the demand side of the shadow economy is not as deep as that in the supply side. This view is shared by Pisani (2013a), who recognizes that informal consumption has received scant scholarly attention.

One of the first references found to informal consumption is by Cermeño (1987). The author develops a model where large and small firms interact with poor, middle-class and rich consumers. The structure presented is very simple and assumes that (formal) goods produced by bigger firms are of higher quality and price in comparison to those (informal) goods produced by smaller firms. Then, using lexicographic preferences, the model establishes that formal goods are only purchased by rich consumers and informal goods only by poor ones. This is explained using a critical-income level not attained by this last group, which prevents them from being able to afford the more expensive formal consumption. However, when the income of the poor increases and they become middle-class consumers, they start substituting informal goods with their formal counterparts, up to the point where they become rich and their purchasing basket is completely composed of formal goods.

As a result, the model predicts the existence of two demands. One for formal goods, negatively related to its own price and positively related to income (for rich and middle-class consumers) and one for informal goods negatively related to its own price, positively related to poor consumers' income and negatively related to middle-class consumers' income. Therefore, formal goods tend to be normal (as long as only middle-class and rich consumers can afford them), while informal goods can be normal or inferior depending on the size of the income elasticities for poor and middle-class consumers and the relative importance of the consumption share of each type of individual on total informal consumption. Finally, cross-price elasticities are positive in the model, revealing that formal and informal goods will behave as (imperfect) substitutes.

Cermeño does not provide concrete empirical evidence for his arguments; however, it should be emphasized that this two-demand framework is in some way supported by a traditional line of thinking by other authors in developing countries, where informal commerce is associated with the needs satisfaction of low-income families (Portes, 2000) and motivated by the lower prices obtained there (Pisani, 2013a). For example, Tokman (1978) argues that informal firms can compete successfully in developing countries with large supermarkets, since only a relatively small proportion of households have a sufficiently high level of consumption, storage capacity and transportation facilities to buy in bigger outlets. Similarly, Yamada (1996) argues that informal repair services for cars or electrical appliances (typical informal services) emerge where only a few households can afford to replace these goods. As a result, as stated by Livingstone (1991), informal goods are mainly the target of low-income consumers.

Quantitative evidence around this issue is provided by Pisani (2013a) for the South Texas borderlands, where he reports that poor people allocate 22% of their income to informal consumption, with middle-class and richer individuals allocating 15% and 5%, respectively. A similar relationship is found in developed countries, where Williams and Windebank (2005) report, in their study on Leicester, that poor families are responsible for between 70% and 85% of the acquisition of informal goods. Both results are in line with the location decision of informal and formal sellers reported in de Soto (1986) for Lima. De Soto shows that, in general, petty traders and city markets (typically informal outlets) are located in low-income neighbourhoods with formal outlets concentrated in the richer or more modern ones.

Although there is an evident relationship between poverty and informal consumption (or, more precisely, the liquidity constraints faced by households), this does not necessarily mean that informal consumption is inferior. When formally tested, the hypothesis has produced mixed evidence. For example, Gardes and Starzec (2002) and Reilly *et al*. (2006) all show that informal expenditure is normal in their studies of Poland and Serbia, respectively, whereas Fortin *et al*. (2000) in Canada found that it is inferior or unrelated (depending on the specification used). The differences could be interpreted in terms of the countries used for estimation. In Canada, a developed country with a large formal sector, inferiority would be a reasonable result while, in

Eastern and Southern European countries with a large informal sector, it would not. Similar results were obtained by Böhme and Thiele (2012b), who show for a sample of African countries that the inferiority of informal markets is rejected, but they recognize differences in the income elasticities that characterize informal and formal channels of commercialization: the former lower (below 1) than the latter (above 1). Therefore, as they propose, a better characterization is to consider informal markets as normal necessities and formal ones as normal luxuries.

This result, as Böhme and Thiele explained, is consistent with quality dualism in the demand structure of African countries and occurs in a context where demand segmentation (as suggested in Cermeño, 1987) is not necessarily verified. As reported, there appears to be a strong overlapping consumer base between the formal and the informal sectors. This result is consistent with the observations of Arellano (2010) and Arellano and Burgos (2010) in Lima, who show that, although formal supply is still limited or inadequate in most of the poorest neighbourhoods in the city (in line with de Soto, 1986), formal and informal markets are more evenly distributed than in the past, with formal and informal outlets expanding to new markets. Similarly, Dasgupta (1992), in his study of petty traders in Calcutta, reports how these sellers are not exclusively tied to low-income markets. In fact, thanks to their greater ability to adapt to consumers' needs, they have been increasingly capturing the middle-income and richest segments defining a spatial configuration throughout the city.

This has generated changes in consumer behaviour, as reported in Arellano and Burgos (2010) and a call for additional explanations for informal/formal consumption beyond a simplistic segmentation approach where consumers' choices rather than exclusion patterns explain their demand. A similar interpretation is provided by Williams (2002) and Williams and Windebank (2002) using the market structure of two English cities (Southampton and Sheffield) and Williams (2008) using information on other European cities. For these authors, in a context where informal channels of commercialization are not the dominant ones, they identified that in at least 70% of the transactions carried out there is a matter of the free choice of clients rather than of exclusion, with positive attributes attached to these outlets beyond reduced prices – for example, more-flexible supply conditions and greater product variety (also mentioned in Böhme and Thiele, 2012b).

Therefore, consumers could be better described in terms of having a strategic behaviour approach, where households are in permanent search of optimal combinations of quality (lower at informal markets) and price (higher at formal markets) given their preferences and economic possibilities. Also, in a context of quality differentials where quality is inadequately signalled in informal markets (as in the lemon market in Akerlof, 1970), searching costs (in the form of Nelson, 1970) will be also a driver of consumption.

This last interpretation is close to the models developed by Kesselman (1989) and Anbarci *et al.* (2012). In the firs case, goods in the formal and informal sector are imperfect substitutes, but consumers will prefer to purchase formal goods given their better reputation, follow-up service, warranty or return policy, lower search costs and so on. Therefore, to successfully compete, informal vendors will need to sell at lower prices, even if the goods are practically identical. As a result, Kesselman imposes a demand behaviour where households of any kind purchase both types, with demand functions depending on relative prices (negatively) and total income (positively). In the second case, the coexistence of informal and formal markets is investigated in a context where buyers and sellers can move freely between sectors depending on their relative net participation gains. In the model, and given the quality differentials between markets, buyers' payoffs will depend on the expected utilities of consumption, prices and transactional costs generating differences in access to the preferred market.

Fortin *et al*. (2000) extend the demand side of Kesselman's model and incorporate the possible influence of labour supply on informal demand. In their model, workers behave in the way used by Lacroix and Fortin (1992) and Fortin *et al*. (1997); however, instead of only choosing labour allocation, they now have preferences in the ordering of goods: formal (produced and sold in the formal sector) and informal (produced and sold in the informal sector). The distinctive feature of the model is that consumers face transactional costs in the informal sector (negatively related to informal hours of work) which are part of the price structure of such goods and some consumers will pay in order to avoid taxes.

This framework allows the authors to derive a consumption function for informal goods conditional on labour supply (as in Pollak, 1969; Browning and Meghir, 1991) which will depend on the hours worked in each sector, prices, total income and exogenous variables. The influence of the informal labour supply is then deconstructed into three effects: income effect (positive as long the informal good is normal); non-separability effect (zero, positive or negative if labour supply in the irregular sector and consumption of irregular goods are weakly separable, complements or substitutes, respectively); and network effect (positive as long as the irregular good is not a Giffen one). The authors suggest that this network effect is a key element in the joint behaviour of labour allocation and consumption in the shadow economy. They argue that there are strong reasons to believe that participating in the informal sector as a worker leads to a higher consumption of informal goods. In fact, informal workers will have better access to the output of this sector, given their better and 'more timely' access to information on the availability and quality of goods and services produced there.

The empirical application by Fortin *et al*. (2000) in Canada and others like Gardes and Starzec (2002) and Reilly *et al*. (2006) in Eastern Europe countries revealed evidence in favour of this network effect. However, Böhme and Thiele (2012b) for African countries found that informal labour supply and informal consumption are unrelated. This is probably a country-specific characteristic (possibly driven by the overlapping of consumers reported by the authors) that must be taken into consideration when testing Fortin *et al*.'s hypothesis in developing countries.

An additional source of transactional costs is proposed by Phumpiu and Triveño (2007). In their unpublished paper, they develop a model to explain the emergence and consequences of the size of the informal sector but from the demand side. In short, the authors consider that 'doing business' in both the formal and the informal sector is costly, but the cost will be negatively correlated with the skills or the ability to reduce the probability of being cheated or deceived developed by consumers. The skills in this case are obtained from interaction and experience in each sector, so two types of consumer emerge: those with enough skills to 'do business' in the formal sector and those with the skills to 'do business' in the informal one. Considering these arguments, it is easy to reinterpret them in terms of Fortin *et al*.'s (2000) observations around network effects: there are transactional costs in the purchasing process in each sector that generate some kind of implicit barriers to entry, with a final effect on demand similar to the costs of access modelled by these authors. As long as these costs increase the perceived (or final) price of the good, previous experience in the informal or formal sector (maybe as a worker) will reduce such costs in absolute terms, generating a source of 'linkage' effect between working and purchasing.

Other explanations of the emergence of linkages around informal transactions are based on the sociological tradition of embedded social networks and interpersonal relations (Granovetter, 1985). Using this conceptual framework, Portes (2000) explains the existence of informal market relationships in the developing world, where strong confidence links constructed among participants (producers, sellers and consumers) are the only resource against fraud in a context where no other enforcement mechanism exists. The confidence in this case, Portes continues, is generated both in terms of shared identities and feelings and in the expected social sanctions that each participant will receive (like the exclusion of the group) if this confidence is betrayed. Pisani (2013b) makes a similar interpretation. In his article, he quotes the study done by Short (1996) that goes one step further, arguing that, in some cases, informal transactions driven by embedded kinship turn into 'familiar obligations' and it becomes difficult (or even impossible) to replace them by formal ones.

However, not every social structure will support these informal relationships, since the social ties need to be strong enough to make fluent market transactions and protect participants from government repression. This is probably why, within this literature, community linkages are those most commonly identified as supporting long-lasting informal markets, where informal goods are normally produced by specific groups in a country (migrants with a similar ethnic background, for example) and are therefore better sold to members of that group. The case studies reported in Portes (2000), as well as those undertaken by Sanchis (1982), Capecchi (1989), McGee *et al*. (1989), Espinoza (1992), Guarnizo (1992) and Bailon and Nicoli (2009) support this idea. The authors all show, with some contrasts in their arguments, that the first clients of a 'new migrant' informal producer are the members of his or her ethnic group previously established in the city. These 'old migrants' are normally engaged in the informal sector in one way or another (probably through work) and eventually help new migrants to migrate, build a house and establish a (probably informal) business in the same neighbourhood.

Other determinants of informal consumption have been explored by the social psychologist literature, where the market misbehaviour of consumers has been researched in terms of their preferences for illicit goods (counterfeit, stolen or pirated). Although the conclusions are not necessarily applicable to informal expenditure (where goods are not necessarily illegal), some of them could be useful in taking into account the attitudinal factors behind demand relations. A overview of these studies is provided by Peace *et al*. (2003), Eisend and Schuchert-Güller (2006) and Pisani (2013a). It is worth focusing on two of them: stigma and perceived risk. In the first case, stigma emerge from the moral concerns that an individual may have when he or she buys goods 'off the books', as long as they are not legally recognized and potentially harmful for society as a whole. In the second case, perceived risk will depend on the enforcement ability of the government but, more importantly, on the attitudes towards risk of each person. Although enforcement ability in developing countries is low (Pisani, 2013a) and penalties are normally reduced or non-existant for consumers (Fortin *et al*. 2000; Pisani, 2013a), attitudes toward the risk itself will be different between consumers. Even small fines or superficial deterrence could have significant impacts on some consumers, while others will be less concerned about the consequences of their actions, even in the presence of strong institutional frameworks.

Finally, a surprising issue revealed during the literature review was that, although substitution between formal and informal goods is an intuitive point of departure for most of the theoretical models that study the emergence of the informal sector in developing countries,[7] explicit

[7] Recent examples are Restrepo-Echevarría (2011) and Charlot *et al*. (2013).

hypothesis-testing for this issue was not found. The only references found that directly include price variables are marketing studies of the social psychology literature for on demand misbehaviour (see Eisend and Schuchert-Güler, 2006 for a review). Most of these applications reveal intuitive negative own-price and positive cross-price effects on the intention to purchase counterfeit and pirated goods, but these studies are normally restricted to small samples of university students specially selected for the analysis of attitudes toward illegal consumption. Further estimations using country samples are done by Chiarini and Marzano (2006) and Gardes and Starzec (2009) for Italy (using a macro model) and Poland (using demand specifications), respectively. However, given the data limitations, the authors test quantity rather than price effects between formal and informal consumption. Interestingly enough, the first application found quasi-complementarities between both sectors and, in the second, quasi-substitution was limited only to a small portion of the informal goods: health, education and cultural expenditure (around a mere 11% of household expenditure).[8]

II.5. Concluding remarks

For the purposes of this research, a working definition of informal and formal transactions is needed. Considering the review presented here, informality will be better understood at the level of relationships maintained by agents in a particular market. When conventions and regulations are not observed, informality arises, but its emergence in a particular market relationship does not imply its existence in another. For example, a firm could be registered and pay taxes (a formal relationship with the government) but hire workers without a contract (an informal relationship with the workers). At the same time, the worker could hold a contract (a formal relationship with the firm), but buy goods in the informal sector (as an informal consumer). Thus, the key point of analysing informality is to identify whether or not a market relationship lacks institutional enforcement, given that this is mainly a multidimensional phenomenon that could be driven by both the free choice of agents or exclusion.

Based on this general definition, some of the stylized facts around the emergence of informality in both the labour and the goods markets have been revised in this chapter. As stated previously, when compared to the supply side of the shadow economy, the academic literature has paid only limited attention to informal consumption (the demand side). Theoretical models and empirical applications are scarce and most of the stylized facts have been constructed based on qualitative

[8] Other references encountered during the review was the paper by Fomba and Mvolo (2010), who estimate own-price and cross-price elasticities for formal and informal goods in order to implement the parameters in their simulation model of tax evasion. However, the elasticities published (though not explained in depth) seem to be for two different systems (goods inside formal markets and goods inside informal ones), so it was not possible to assess their results in terms of the hypothesis-testing proposed here.

evidence and observational case studies. Although this literature provides some important concepts on informal consumption, systematic hypothesis-testing is practically non-existant. Therefore, instead of theoretical or empirical regularities, the literature review offers a wide range of open debates where proper demand theory has rarely been used. This is the main gap identified in this thesis, and justifies the research agenda already outlined in the previous chapter.

III. A SIMPLE MODEL FOR INFORMAL CONSUMPTION

III.1. Introduction

The aim of this chapter is to present a simple theoretical model of informal consumption in highly informal developing countries like Peru. This model is based on the original ideas of Fortin *et al*. (2000), and includes the insights of standard demand theory discussed in Deaton and Muellbauer (1980b), Browning and Meghir (1991) and Carpentier and Guyomard (2001). Slight modifications have been introduced in order to include some of the stylized facts around informality in developing countries already discussed in Chapter 2. As already explained, in highly informal developing countries there are neither purely formal nor purely informal consumers. In fact, there are just consumers deciding on optimal combinations of price and quality according to their budget possibilities. There are two available markets for that purpose: formal and informal. As a result, optimal decisions will imply consumption baskets with products from different combinations of formal and informal sources, where the shares are basically determined by inherent preferences, income, prices and transactional costs.

The model is developed using unitary utility optimization processes in a static (short-run) partial equilibrium framework for households choosing consumption conditional on their previous labour allocation decisions. Two possible preference orderings are allowed. Under the first, in the initial stage of the decision process, the household with enough knowledge about the characteristics of the market and the supply of goods within it decides the amount of its budget that will be allocated and then which goods it will purchase on these markets. Considering this framework, the only relevant stage at which to reveal the main aspects of informal consumption (income elasticity, own- and cross-price elasticities and labour effects) will be the first.[9] This model (later called the aggregate version) is the one used by most of the empirical approximations encountered in the literature (Fortin *et al*., 2000; Gardes and Starzec, 2002, 2009; Reilly *et al*., 2006; Böhme and Thiele, 2012b).

Under the second preference ordering, the family with enough knowledge of the goods they will purchase and the markets where these goods are sold first decides on the allocation of funds between the items to be consumed and then where they will be bought (formal or informal markets). In this framework, just observing one of the stages is not enough; to derive the main aspect of informal consumption, both stages in the decision process need to be modelled and

[9] The second stage of this approximation is the decision on which goods to purchase on each market. This stage is not developed in this document, but can easily be incoporporated without changes to the main insights presented here.

estimated. This model (later called the disaggregated version) is less standard in the literature and avoids aggregation since it recognizes that income, price and labour supply effects will be different for the various goods consumed by the family. A similar version of the model was used earlier in the empirical approximation of Böhme and Thiele (2012b).

A priori, there is no one correct assumption on how to model household preferences and the literature reviewed has revealed no insights on this issue. The objective of this chapter is not, however, to solve this theoretical question, which is beyond the scope of the present research. By contrast, in this application a better way to proceed is to formalize the existence of the two simplest (and possibly most extreme) allocation procedures constructed under their own (and restrictive) simplifying assumptions in order to justify and investigate the hypothesis which is of particular interest for the application. As will be seen in the next chapters, in empirical terms the two models offer complementary information.

The chapter is organized as follows. In sections two and three, the aggregated and disaggregated versions of the model are presented and discussed. The main objective here is to formalize the conditional and unconditional marginal effects of income, prices and hours worked on informal and formal consumption in the two different budgeting processes. In section four, the Almost Ideal Demand System (AIDS) of Deaton and Muellbauer (1980a) is presented and discussed as this is the general empirical specification selected for demand estimation for both models. In section five, some concluding remarks are presented.

III.2. The aggregated version of the model

The simplest way to model informal consumption is to consider two markets[10] (formal and informal). Therefore, the simplifying assumption introduced by the model is that the household procedure of resource allocation is the decision as to where family goods and services will be bought. In general, the term c_j will refer to consumption allocated in market j ($\forall j = f, i$ with f: formal and i: informal) at prices p_j. Using this framework, three stylized facts of informal consumption are allowed: imperfect substitution among c_f and c_i given quality differences, price differentials and transactional costs.

In the first case, as discussed in Cermeño (1987) and Kesselman (1989), the quality of informal consumption will generally be lower than that of its formal counterparts. Cermeño (1987) justifies

[10] As for classification issues, the available markets are, in fact, three: formal, semi-formal and informal. However, in this discussion the semi-formal category is excluded to facilitate exposition of the arguments.

this assumption in considering that informal producers (small firms) will compensate for their lower productivity by using cheaper and lower-quality inputs, paying lower wages and allowing their products and services to deteriorate. By contrast, as mentioned earlier, Kesselman (1989) argues that, even if both sectors offer the same good (a possible situation, as reported in Dasgupta, 1992), certain attributes of formal outlets – such as a better reputation, follow-up, warranty and return policy – will justify the differences in the quality perceived by consumers. Arellano and Burgos (2010) show qualitative evidence in favour of these arguments for Peruvian cities.

In the second case, a natural consequence of quality differences is price differentials, as long as lower-quality products will need to be sold at lower prices in order to successfully compete with higher-quality ones (Kesselman, 1989). As a result it is expected that $p_f > p_i$ in the general situation. However, other sources of price differential exist. First, there will be a financial advantage to informal producers (or sellers), obtained from not paying most of the existing regulation and registration costs like licences, taxes or permissions (see Farrell, 2004; Fukuchi, 1998 for a discussion). This advantage is not compensated by other hiding or evasion costs (Fortin *et al*., 2000; Arias *et al*., 2010) paid by informal produces, since they are possibly low because of the low enforcement (and the low risk of detection) in developing countries (Pisani, 2013a). Second, the limited supply of formal outlets in highly informal countries (as reported in de Soto, 1986; Arellano and Burgos, 2010) generates different price formations (as discussed in Cermeño, 1987), with formal producers under a quasi-monopolistic structure and informal ones under competition, with higher prices in the former than in the latter. Third, the location decisions of formal producers in modern or richer neighbourhoods could generate some kind of Balassa–Samuelson effect.[11]

In the third case, as originally proposed by Fortin *et al*. (2000), the model recognizes the existence of transactional costs with differentiated effects on formal and informal consumption. These costs could reflect the marginal investments done by consumers in order to make a ‘good deal’ in the preferred market and could take different forms. For example, searching costs in a context of quality differences with higher incidence in the sector where quality standards are hidden for consumers (informal). In fact, it is assumed that, in informal markets, the probability of being cheated is perceived as relatively high and consumers must pay in order to avoid this risk. By contrast, in the formal sector, the consumer will not need to search for quality (the formal condition of the seller will serve as a sign to the consumer with no cost in order to avoid distrust), but possible complex rules to be followed there in order to perform the purchase will constitute a

[11] According to Balassa (1964) and Samuelson (1964), consumer prices in richer countries are systematically higher than in poorer countries.

transactional cost (in time and effort), as discussed in Phumpiu and Triveño (2007). Similarly, in both cases, if the supply of formal and informal goods is geographically segmented or if there is asymmetric information, the consumption possibilities offered by each market will not be completely evident for all consumers, so transactional costs constitute true access costs to each market supply.

These costs are denoted by a_j with different size depending on the market (j). Under such conditions, the complete price[12] paid for purchasing in the formal or informal market is $p_j(1+a_j)$ where $p_i(1+a_i) > p_f(1+a_f)$ or $p_i(1+a_i) < p_f(1+a_f)$ depending on the size of a_j and p_j. It is assumed that a_j will depend on household characteristics – mainly previous linkages maintained with the formal or informal sector. In this model, these previous linkages come in the form of labour supply, since this will reflect the experience or knowledge developed by the consumer in each sector. In general, $a_j = a_j(h_f, h_i)$ with $\partial a_f / \partial h_f < 0$, $\partial a_i / \partial h_i < 0$, $\partial a_f / \partial h_i > 0$ and $\partial a_i / \partial h_f > 0$. These relationships are identified as the source of the linkage effect discussed in the previous chapter.

Therefore, the problem of the consumer can be summarized in the following way. A household decides to purchase c_f and c_i (where the quality of c_f is higher than of c_i) at prices p_f and p_i (where $p_f > p_i$), facing transactional costs a_f and a_i, and has a budget of M to do so. Using the full income restriction of Becker (1981) and the conditional demand framework[13] of Pollak (1969) explained in Browning and Meghir (1991), with leisure (l) as the pre-allocated good at level l^*, it follows that the expression will take the form:

$$\text{max.} \quad W = W(c_i, c_f, l)$$

$$\text{s.t.} \quad M = sT + Y - sl = p_i[1 + a_i(h_i, h_f)]c_i + p_f[1 + a_f(h_i, h_f)]c_f \qquad \text{(III.1)}$$

[12] Gardes and Starzec (2009) also provide preliminary empirical evidence for the existence of such costs in informal consumption. The multiplicative way of representing transactional costs is similar to that used by Muellbauer (1981) in the context of durable goods (it will facilitate manipulation of terms under the log specification of the empirical model).

[13] There are several advantages to using conditional demand systems when labour supply is introduced in demand systems. As explained in Browning and Meghir (1991), among other reasons, the conditional framework helps to avoid separation between labour and consumption in a very simplistic way. Also, it helps to adequately model the demand for free goods without modelling the determination of predetermined goods explicitly or its budget constraint. Moreover, it will be correctly specified even if hours of work are not chosen optimally or are under the presence of non-linearities (corner solutions). This provides a methodological advantage since it is possible to focus on demand on goods without caring about things like labour supply or unemployment.

$l = l^*$

The function W is a good representation of preferences and has all the desirable properties of utility and is increasing in all its arguments, s are salaries, T is total time available and Y is non-labour income. Replacing the final restriction on the utility function and the full income restrictions, and considering that $l^* = T - \sum_r h_r$, for h_r $(r = f, i)$ representing formal (f) and informal (i) earning opportunities, (III.1) can be expressed in the following way[14]

max. $\quad W = W(c_i, c_f; h_f, h_i, \lambda)$

s.t. $\quad M = s_f h_f + s_i h_i + Y = p_i[1 + a_i(h_i, h_f)]c_i + p_f[1 + a_f(h_i, h_f)]c_f \qquad$ (III.2)

where additional household determinants λ have been already introduced in the model as preferences shifters. In this model, the vector λ is defined according to observations by Deaton and Muellbauer (1980b), Theil (1952–53), Houthakker (1952–53) and Cox and Wohlgenant (1986), where quality choices will depend, beyond income, on household size and composition, as well as on other characteristics such as education, age profiles, gender and so on (Lazaridis, 2003; Vassilopoulos *et al.*, 2009). However, considering that quality standards are attached to a particular market previously identified by the consumer, λ must be extended by additional controls influencing participation in the informal or formal economy. For example, issues already discussed in Chapter 2 – like location, ethnic and migration conditions and enforcement capacity – also need to be introduced in the model.

Solving (III.2) for c_f and c_i, and once optimization conditions are reached, it is possible to demonstrate the existence of a demand function like

$$c_j = c_j^D[\bar{p}_i, \bar{p}_f, M; h_f, h_i, \lambda] \qquad \text{(III.3)}$$

which is the conditional demand function of Pollak (1969), with all its desirable properties over the $[c_f, c_i]$ dimensional space: homogeneity of degree zero with respect to prices and income

[14] Note that salaries could be defined as $s_f = (1 - t + b)\bar{s}_f$ and $s_i = (1 - r)\bar{s}_i$ with t for marginal tax rates, b for social benefits, r for any payment made by informal workers to avoid detection and $\bar{s}_r$ as gross salaries. In empirical terms, only t and b are identified; r is not. Thus, for the rest of the discussion it is preferable to impose $r = 0$. This is possibly a gross assumption but, in a context of highly informal countries with a very low governmental enforcement capacity, r is maybe low enough to be neglected.

and diagonal negative semi-definite Slutsky matrix where $\bar{p}_i = p_i(1+a_i)$, $\bar{p}_f = p_f(1+a_f)$. From (III.3) marginal effects of prices and income could be computed and interpreted, as could those of previous linkages with the formal and the informal sector. Taking partial derivatives with respect to M, p_g and h_r; (where $g = f, i$) results obtained are:

$$\frac{\partial c_j}{\partial M} = \frac{\partial c_j^D}{\partial M} \qquad \text{(III.4)}$$

$$\frac{\partial c_j^O}{\partial p_g} = \frac{\partial c_j^D}{\partial \bar{p}_g}(1+a_g) \qquad \text{(III.5)}$$

$$\frac{\partial c_j^H}{\partial p_g} = \frac{\partial c_j^O}{\partial p_g} + c_g \frac{\partial c_j}{\partial M} \qquad \text{(III.6)}$$

$$\frac{\partial c_j}{\partial h_r} = \left[\frac{\partial c_j^D}{\partial \bar{p}_i}\frac{\partial a_i}{\partial h_r}p_i + \frac{\partial c_j^D}{\partial \bar{p}_f}\frac{\partial a_f}{\partial h_r}p_f\right] + \frac{\partial c_j^D}{\partial h_r} \qquad \text{(III.7)}$$

$$\frac{\partial c_j^T}{\partial h_r} = \left[\frac{\partial c_j^D}{\partial \bar{p}_i}\frac{\partial a_i}{\partial h_r}p_i + \frac{\partial c_j^D}{\partial \bar{p}_f}\frac{\partial a_f}{\partial h_r}p_f\right] + \frac{\partial c_j^D}{\partial h_r} + \frac{\partial c_j^D}{\partial M}s_r \qquad \text{(III.8)}$$

Expression (III.4) is the income effect, positive as long as c_j is normal and negative if it is inferior. In fact, a testable proposition of the model is to verify whether informal consumption can be considered inferior or not and whether there are noticeable differences in income responses between formal and informal consumption. Expressions (III.5) and (III.6) reveal a set of ordinary (denoted by superscript O) and Hicksian (denoted by superscript H) price effects, respectively. When $g = j$, (III.6) must always be negative and (III.5) negative as long as the good is not a Giffen one. When $g \neq j$, (III.5) and (III.6) must be positive under the maintained assumption that c_j and c_g are considered substitutes and weighted income effects are lower than cross-price effects.

Expression (III.7) reveals linkages between working and purchasing in the preferred sector, holding total expenditure (M) constant. It is composed of two effects: network effects (first term on the right in brackets) and non-separability or preferences effects (second term on the right). In the first case, the theory suggests that the network effect will be positive when $r = j$ and negative when $r \neq j$. These effects will prevail as long as formal and informal consumption are not Giffen and imperfect substitution holds between them (with weighted income effects lower than cross-price effects). Under such conditions, $\partial c_i^D / \partial \overline{p}_i < 0$, $\partial c_i^D / \partial \overline{p}_f > 0$, $\partial c_f^D / \partial \overline{p}_f < 0$, $\partial c_f^D / \partial \overline{p}_i > 0$, $\partial a_f / \partial h_f < 0$, $\partial a_i / \partial h_i < 0$, $\partial a_f / \partial h_i > 0$ and $\partial a_i / \partial h_f > 0$.

In the second case, the non-separability effect defines the quantity relation between the freely chosen good (c_j) and the pre-allocated good (h_r) when total income (M) is held constant. This implies a re-allocation procedure of the unit relaxation of the budget constraint in (III.2) where the consumption of some components of c_j will be increased and some will be reduced (Deaton, 1981). As discussed in Pollak (1969), when a positive relation is observed, h_r and c_j will be defined as quasi-complements and, when a negative relation is observed, they will be defined as quasi-substitutes. In general, it will be expected that $\partial c_j^D / \partial h_r > 0$ when $r = j$ and $\partial c_j^D / \partial h_r < 0$ when $r \neq j$.

There are several reasons for such a result. For example, unregistered transactions will be preferred by informal workers since this type of individual wants to keep hidden, and formal producers will prefer to buy formal goods only if they can enjoy the fiscal advantage of tax deduction. Similarly, if, as argued by Fortin *et al*. (2000), stigma or tax morale prevents people purchasing from informal outlets, then those who already work in the informal sector will have fewer reasons to avoid this type of outlet, while formal workers will have more. Furthermore, as noted by Arellano (2010), a well-established result for informal workers is the infrequency of their payments, which precludes any planning of their purchases (on a weekly or monthly basis) and the ability to buy in high volume. This will force them to find outlets – like informal ones – which can be highly flexible in the volumes sold. Williams (2006) and La Porta and Shleifer (2011) also comment that cash transactions facilitate informal consumption – a characteristic more notably seen in the type of payments received by informal workers. Formal workers, instead, considering their less-flexible work schedules and their more-regular incomes, will favour formal outlets where planned (and larger) purchases can be made.

Without knowledge of the cost function a_j, the two effects cannot be disentangled so, instead of testing them separately, they will be treated as a joint effect, referred to here as the linkage effect. Under these assumptions, a positive result of (III.7) when $r = j$ and negative when $r \neq j$ will be enough evidence to conclude that working and purchasing in a similar sector are linked activities. However, deviations from the maintained assumptions[15] could obscure the results. In such circumstances it is possible that (III.7) yields similar signs for $r = j$ and $r \neq j$. In those cases, the evidence in favour of the linkage hypothesis comes in the form of higher values of (III.7) when $r = j$ than when $r \neq j$, in absolute terms. This is exactly how Fortin *et al*. (2000) test their hypotheses.

Finally, (III.8) is the full hours' effect which is the same as (III.7) but without holding M constant and using the income restriction in (III.2) to derive these effects. It will reveal the effect of working one additional hour in the formal or informal sector, considering both the linkage effects discussed previously and the additional labour income produced by working more hours. This last effect is captured by the term $(\partial c_j^D / \partial M) s_r$ at the end of the right-hand expression. Therefore, it will just increase or reduce the linkages encountered by the previous expression, depending on whether consumption is normal or inferior and the positive or negative linkage derived from (III.7). In fact, (III.8) is possibly a more interesting result for public policy design as long as it reveals the total impact on informal or formal consumption of, for example, supply-side policies oriented to change the labour allocation of household members.

III.3. The disaggregated version of the model

In the previous model, households order their preferences around broad market allocation, which in turn implies the joint consumption decisions of selecting markets (formal or informal) and goods sold there. As explained previously, this second model tries to disentangle the allocation process between goods and markets. For the following discussion, we consider a utility-maximizing household which has preferences over n goods and is not indifferent to the place where these goods are bought. In fact, the household will choose from two possible markets: formal (f) and informal (i). To simplify the allocation process, it is convenient to employ a two-stage procedure: in the first, the household decides 'which' goods to consume; in the second, once optimum allocation is reached, the household decides 'where' the goods will be bought.

The term c_{kj} refer to goods of type k ($\forall k = 1,...,n$), purchased in the market j ($\forall j = f, i$). Under enough weak separability assumptions of the indirect utility function, the maximization process consists of one main utility function composed by n sub-utility programs that can be

[15] As pointed out by Fortin *et al*. (2000), working on the formal sector could provide some information on goods provided on the informal one and *vice versa* although, to be consistent with the discussion presented, the effects will be lower.

solved independently for the allocation process among each of the two markets, conditional on the total expenditure on each of the n goods. In this framework, the preferences of the household can be represented using the two-stage demand model explained in Deaton and Muellbauer (1980b). The expression takes the form

$$U = U[v_1(c_{1f}, c_{1i}), ..., v_k(c_{kf}, c_{ki}), ..., v_n(c_{nf}, c_{ni})] \qquad \text{(III.9)}$$

where U is the utility function revealing the broad group allocation process and v_k represents the sub-utility functions for places where the broad group must be purchased. Both functions (U and v_k) are good representations of preferences, have all the desirable properties of utility and are increasing in all their arguments. However, it must be recognized that an expression like (III.9) imposes important restrictions on household behaviour, mainly the substitution possibilities between markets in different groups: substitution effects between formal and informal markets for the same good are completely modelled, but this substitution for different goods in different markets can only happen through the income effect in the budget constraint.

The first stage

In the first stage of the model, the household chooses which goods to purchase, taking into account a budgeting process where the consumer is first interested in defining the amount of resources to be allocated between goods (or broad groups of goods). The assumption behind this is that preferences can be ordered in such a way that allocation is possible given only the household's knowledge of total income and group prices and with no care (at this initial stage) about the markets (formal and informal) where these goods will be purchased. As explained in Deaton and Muellbauer (1980b), this problem is based on a standard utility optimization with a budget constraint. However, the exact solution requires non-satisfactory assumptions, so it is better to solve it as a good approximation of the exact result. For this purpose, special caution is needed in the construction of quantity and price indexes since both must define a synthetic maximization procedure with perfect results, in a sense that needs to be identical as a one-step optimization with full information.

In this model, the approximation of Carpentier and Guyomard (2001) is adopted. This means that broad quantity indexes are approximated by constant price composites and broad group prices by an implicit price deflator, under the assumption that the empirical variation of the True Cost of Living (TCOL) with respect to the utility level is sufficiently small to be neglected. In this application, this assumption will hold when within-group prices are highly collinear, i.e., formal and informal prices inside each group.[16] This will be true if price formation acts as a guide for the informal economy or, according to Cassel (1984), if price increases in the formal economy spill over into the informal economy through the mechanism of relative prices. The full derivation can be reviewed in Deaton (1986), Deaton and Muellbauer (1980b) and Carpentier and Guyomard (2001). Simplifying the notation for the purposes of this research, the first-stage maximization problem is approximated by:

$$\begin{aligned} &\text{max.} \quad U = F(C_1,...,C_k,...,C_{n-1},l) \\ &\text{s.t.} \quad M = sT + Y - sl = \sum_{k=1}^{n-1} C_k P_k \qquad \text{(III.10)} \\ &\qquad\;\; l = l^* \end{aligned}$$

which is a standard maximization problem of choosing broad group quantities (C_k) subject to an income constraint and broad group price indexes (P_k), conditional on previous leisure (l) decisions. The remaining arguments are defined as before. Once (III.10) is solved for formal and informal hours of work and the set of additional demand shifters (Z) is introduced, the program takes the form:

$$\begin{aligned} &\text{max.} \quad U = F[C_1,...,C_k,...,C_{n-1};h_f,h_i,Z] \\ &\text{s.t.} \quad M = s_f h_f + s_i h_i + Y = \sum_{k=1}^{n-1} C_k P_k \qquad \text{(III.11)} \end{aligned}$$

Solving (III.11) implies a household deciding its consumption for $n-1$ free goods (C_k) conditional on the consumption level over the pre-allocated goods or, in this framework, conditional on labour supply decisions (h_r). Once this optimization program is solved, it is easy to demonstrate that final demands take the form:

$$C_k = C_k^D(P_1,...,P_k,...,P_n,M,h_f,h_i,Z) \qquad \text{(III.12)}$$

which is the conditional demand function of Pollak (1969), with all its desirable properties over the $n-1$ dimensional space. Considering (III.12), the marginal effects of prices, income and hours

[16] Note that, under the aggregate model, if broad market allocation is thought of as the first stage of the problem, similar solutions must be adopted.

of work could be computed taking partial derivatives with respect to M, P_q and h_r. These expressions take the form:

$$\frac{\partial C_k}{\partial M} = \frac{\partial C_k^D}{\partial M} \tag{III.13}$$

$$\frac{\partial C_k^O}{\partial P_q} = \frac{\partial C_k^D}{\partial P_q} \tag{III.14}$$

$$\frac{\partial C_k^H}{\partial P_q} = \frac{\partial C_k^O}{\partial P_q} + C_q \frac{\partial C_k^D}{\partial M} \tag{III.15}$$

$$\frac{\partial C_k}{\partial h_r} = \frac{\partial C_k^D}{\partial h_r} \tag{III.16}$$

Expression (III.13) is the expenditure effect, positive for normal goods and negative for inferior ones. Expressions (III.14) and (III.15) are the ordinary and Hicksian price effects, respectively, where $q = 1,...,k,...,n-1$ is a broad group with $q = k$ for own price derivatives, always negative (barring Giffen goods) or $q \neq k$ for cross-price derivatives, negative for ordinary or Hicksian substitutes and positive for ordinary or Hicksian complements. Expression (III.16) is the marginal effect of labour allocation. This expression is only the non-separability effect, but at the level of goods with a similar interpretation in terms of quasi-complements and quasi-substitutes, as in the previous case.

<u>The second stage</u>

Once the household completes the first stage and decides which goods to consume, then a second step is to decide where these goods are going to be bought: at informal or at formal markets. It is obvious that this second allocation procedure shares most of the characteristics of the aggregate model and could be formalized with just minimal changes in the notation. In principle, from the previous stage it is possible to find the budget allocation in the desired good (or broad group) which, in turn, is $C_k P_k = m_k$ for broad good k. Then m_k will be allocated within the j markets available for the consumer. From the sub-utility function k of (III.9), it follows that these preferences can be described with $v_k(c_{kf}, c_{ki})$.

Again, it is expected that $p_{kf} > p_{ki}$ and that there will be quality differences between formal and informal outlets (higher in the case of formal options). Similarly, transactional costs are allowed in the form a_{kj}, with different size depending on the market (j) under analysis and with the consumption group (k) considered (since relative incidence of transactional costs will differ according to the nature of the good purchased). Finally a set of additional household characteristics λ are also allowed. As a result, the maximization problem of the household in this second stage will be defined by the following optimization program:

$$\text{max.} \quad v_k = v_k(c_{kf}, c_{ki}; \lambda) \tag{III.17}$$

$$\text{s.t.} \quad m_k = p_{ki}[1 + a_{ki}(h_i, h_f)]c_{ki} + p_{kf}[1 + a_{kf}(h_i, h_f)]c_{kf}$$

It is important to note in (III.17) that the maximization program is no longer in a conditional demand setting. Once solved, the first stage of the allocation process, separation between labour and consumption, could be imposed in the second. Therefore, once the optimization conditions for c_{kf} and c_{ki} are reached, it is possible to demonstrate the existence of a demand function like

$$c_{kj} = c_{kj}^{D}[\overline{p}_{ki}, \overline{p}_{kf}, m_k, \lambda] \tag{III.18}$$

where $\overline{p}_{ki} = p_{ki}(1 + a_{ki})$, $\overline{p}_{kf} = p_{kf}(1 + a_{kf})$. From (III.18), again, income and price marginal effects are computed as well as those of previous linkages with the formal and the informal sector. Taking partial derivatives with respect to m_k, p_{kg} and h_r; the results obtained are:

$$\frac{\partial c_{kj}}{\partial m_k} = \frac{\partial c_{kj}^{D}}{\partial m_k} \tag{III.19}$$

$$\frac{\partial c_{kj}^{O}}{\partial p_{kg}} = \frac{\partial c_{kj}^{D}}{\partial \overline{p}_{kg}}(1 + a_{kg}) \tag{III.20}$$

$$\frac{\partial c_{kj}^{H}}{\partial p_{kg}} = \frac{\partial c_{kj}^{D}}{\partial \overline{p}_{kg}}(1 + a_{kg}) + c_{kg}\frac{\partial c_{kj}^{D}}{\partial m_k} \tag{III.21}$$

$$\frac{\partial c_{kj}}{\partial h_r} = \left[\frac{\partial c_{kj}^D}{\partial \bar{p}_{ki}} \frac{\partial a_{ki}}{\partial h_r} p_{ki} + \frac{\partial c_{kj}^D}{\partial \bar{p}_{kf}} \frac{\partial a_{kf}}{\partial h_r} p_{kf} \right] \tag{III.22}$$

which are similar expressions to those obtained for the aggregate market, but conditional on goods expenditure. Expression (III.19) is the income effect, positive as long as good k purchased in market j is normal and negative if it is inferior. Expressions (III.20) and (III.21) reveal a set of ordinary and compensated (or Hicksian) price effects, respectively. When $g = j$, (III.20) and (III.21) must be always negative (excluding Giffen goods). When $g \neq j$, (III.20) and (III.21) must be positive under the maintained assumption that c_j and market c_g are considered substitutes and weighted income effects are lower than cross-price effects.

Expression (III.22) is of particular interest and reveals linkages between working and purchasing in the preferred sector, holding expenditure m_k constant. In contrast to the aggregated model, it is composed only of the network effects as long as labour and consumption in this stage are assumed as separable. Therefore, the informational channel already discussed could be better analysed in this particular context. A negative (positive) sign in (III.22) when $r = j$ ($r \neq j$) will be sufficient evidence to conclude in favour of the network effect, but weaker evidence is again allowed when (III.22) is in a similar direction for $r = j$ and $r \neq j$, but is larger for $r = j$ than $r \neq j$, in absolute terms.

Unconditional elasticities

The second stage of the model provides an adequate framework in which to study linkages between working and purchasing in the formal or informal market as well as to derive price and income elasticities in each case. However, it should be recognized that the results are conditional on expenditures already allocated in broad groups. Therefore, the final effects provided will be incomplete for two reasons. First, from a public policy point of view, the result of resource allocation between formal and informal outlets is more relevant if it can be interpreted in terms of changes in total income or expenditure. Second, as seen in the first stage of the problem, labour allocation will also influence the demand for broad groups via non-separability and income effects. In such circumstances, unconditional effects for income, prices and linkages will differ from conditional ones, something that must be incorporated in the analysis.

For this purpose, the first and second stages will be merged, following the procedures of Deaton and Muellbauer (1980b), Deaton (1986) and Carpentier and Guyomard (2001). According to these authors, income, price or other marginal effects computed inside a broad group can be corrected, with marginal effects computed for the first stage in order to make them comparable with the aggregate model. Under the assumptions of weak separability imposed, the channel that connects the demands inside one broad group with the demands from another is the expenditure or income effect (i.e. changes in the amount of resources allocated to purchase the broad group). Therefore, unconditional versions of conditional marginal effects need to recognize correction terms constructed from implied changes in group expenditure.

Considering the condition $C_k P_k = m_k$ or $C_k^D P_k = m_k$ established earlier, and expression (III.12), it is possible to re-write the demand function (III.18) in the following way:

$$c_{kj} = c_{kj}^D[\bar{p}_{ki}, \bar{p}_{kf}, P_k C_k^D(P_1,...,P_k,...,P_n, M, h_f, h_i, Z), \lambda] \tag{III.23}$$

Therefore, if the objective is to obtain the marginal effect $\partial c_{kj} / \partial M$ (the unconditional expenditure effect, denoted by superscript U), then it will be expressed as

$$\frac{\partial c_{kj}^U}{\partial m} = \frac{\partial c_{kj}}{\partial m_k} \frac{\partial C_k}{\partial M} P_k \tag{III.24}$$

where the first multiplicand of the right-hand expression is obtained from the second-stage problem and the second from the first stage. It is assumed that all broad groups defined in the first stage of the problem are normal, so signs previously encountered in the second stage must hold.

The unconditional counterparts for expressions (III.20) and (III.21) are more complicated, since they will depend on the assumption of the TCOL imposed and the construction of the price indexes followed. As previously discussed, in this model the methodology already developed by Carpentier and Guyomard (2001) is followed. The authors use the expression $s_{ab} = sc_{ab} + (\tau_{AB})(\partial c_a / \partial m_A)(\partial c_b / \partial m_B)$ with $\tau_{AB} = P_A P_B (\partial H_A / \partial P_B)$ where s_{ab} is the unconditional Slustky substitution term between goods a and b in two broad groups A and B respectively ($a \in A$ and $b \in B$); sc_{ab} is the conditional Slustky substitution term (sc_{ab} exists if $A = B$, but $sc_{ab} = 0$ if $A \neq B$); τ_{AB} is a proportionality term depending on groups A and B ($\tau_{AB} = \tau_{BA}$); c_a and c_b are the consumption levels of goods a and b; m_A and m_B are the total expenditure on broad groups A and B; P_A and P_B are the price indexes; and H_A is the compensated demand for broad group A.

Using these expressions, it is possible to derive the unconditional Hicksian price elasticities for the model presented above, using $\partial O_A / \partial P_B = \partial H_A / \partial P_B - C_B \partial O_A / \partial m$ (Slustky equation), where O stands for the ordinary demand, m for total income and C for the consumption level; it is possible to transform it in terms of ordinary price elasticities. For that purpose, consider broad groups k and q where markets j ($j = f, i$) and g exist ($g = f, i$) with $j \in k$ and $g \in q$. After some algebraic manipulation, the resultant expressions are:

$$\frac{\partial c_{kj}^{UO}}{\partial p_{qg}} = \frac{\partial c_{kj}^{O}}{\partial p_{qg}} + \left(c_{qg} \frac{\partial c_{kj}}{\partial m_k} \right) \left(\delta_{kq} - P_k \frac{\partial C_k}{\partial M} \right) + P_k P_q \left(\frac{\partial C_k^O}{\partial P_q} + C_q \frac{\partial C_k}{\partial M} \right) \frac{\partial c_{kj}}{\partial m_k} \frac{\partial c_{qg}}{\partial m_q} \qquad \text{(III.25)}$$

$$\frac{\partial c_{kj}^{UH}}{\partial p_{qg}} = \frac{\partial c_{kj}^{H}}{\partial p_{qg}} + P_k P_q \frac{\partial C_k^H}{\partial P_q} \frac{\partial c_{kj}}{\partial m_k} \frac{\partial c_{qg}}{\partial m_q} \qquad \text{(III.26)}$$

where δ_{kq} is the Kronecker delta ($\delta_{kq} = 1$ when $k = q$ and $\delta_{kq} = 0$ when $k \neq q$). It follows also that, when $k \neq q$, $\partial c_{kj}^O / \partial p_{qg} = 0$ and $\partial c_{kj}^H / \partial p_{qg} = 0$. Therefore, as shown previously, when both markets under analysis are in the same group, the conditional price elasticities encountered in (III.20) and (III.21) are adjusted, basically, by the own-price derivatives of the broad groups and the income effects: third and second term in the right-hand expression of (III.25) and second term in the right-hand expression of (III.26). Under the maintained assumption of normality in the first and second stages, Hicksian unconditional own-price effects will be the same (in sign) to the conditional ones. In the case of Hicksian cross-price effects, substitution between markets will hold in the unconditional version only for low enough adjustment terms. If inferiority of any of the markets is allowed, unconditional own-price Hicksian effects are no longer equal (in sign) to conditional results (given a high enough adjustment effect). Similarly, conditional substitution among markets encountered in the second stage will always hold. In the case of ordinary unconditional price effects, interpretation of the results is more complex since, as usual, its size and sign will depend on the size and sign of the first-stage income effects.

When the markets under analysis are in different groups, the substitution limitations imposed by the separability assumption are evident. In principle, it should be recognized that, in this case, unconditional own-price and cross-price elasticities for markets do not exist. Therefore, the price effect that will prevail is that between broad groups. For example, if groups q and k are substitutes (complements), then markets i and j of q will be substitutes (complements) of markets i and j

of k, independent of the relation of i and j inside q or k. This is an important result, since it allows complementarities between formal and informal consumption, even under the maintained assumption of substitution between markets inside a similar broad group.

In the case of hours worked, it is possible to use (III.27) to derive the unconditional version of $\partial c_{kj} / \partial h_r$. The resultant expression will take the form:

$$\frac{\partial c_{kj}^{U}}{\partial h_r} = \frac{\partial c_{kj}}{\partial h_r} + \frac{\partial c_{kj}}{\partial m_k} P_k \left(\frac{\partial C_k}{\partial h_r} \right) \tag{III.27}$$

when M is held constant. The first term of (III.27) of the right-hand expression is the conditional derivative obtained from the second-stage problem and the second term comes from the first stage. Therefore, to estimate the unconditional version of expression (III.22), it must be corrected by the first-stage non-separability effect (between hours and broad groups). If C_k and h_r are quasi-complements (substitutes) and $(\partial c_{kj} / \partial m_k) > 0$, the last term will be positive (negative) so any positive linkage encountered above will be increased (reduced). By contrast, when c_k and h_r are quasi-complements (substitutes) but $(\partial c_{kj} / \partial m_k) < 0$, the last term of (III.27) will be negative (positive) so any positive linkage encountered above will be reduced (increased). The converse applies when a negative linkage is encountered in the second stage of the problem. As a result the sign for the linkage effect revealed by the unconditional expression could be different from the predicted sign of the conditional expression if the first-stage non-separability effects are sufficiently strong.

However, as mentioned, (III.27) is computed under the assumption that total expenditure M is maintained constant. As discussed previously, a more interesting result for public policy purposes will be the total effects: the pure effect of hours plus the income effects generated by the additional hour of work. Therefore, to compute the whole effect of hours of work on consumption, besides the second-stage linkage effects and first-stage non-separability effects, an expenditure or income effect should be recognized. This expression takes the following form (where superscript UT is just used to denote the unconditional total effect of hours):

$$\frac{\partial c_{kj}^{UT}}{\partial h_r} = \frac{\partial c_{kj}}{\partial h_r} + \frac{\partial c_{kj}}{\partial m_k} P_k \left(\frac{\partial C_k}{\partial h_r} + s_r \frac{\partial C_k}{\partial M} \right) \tag{III.28}$$

Note that (III.28) is just (III.27), but with a second term in the right hand extended by the term $s_r(\partial C_k / \partial M)$ inside the brackets. Therefore, the income effect will work just by increasing (decreasing) the first-stage non-separability effects for quasi-complements (substitutes), with the same consequences and interpretation as before.

III.4. Empirical specification

The main specification considered for estimation of the demand functions discussed follows the Almost Ideal Demand System (AIDS) proposed by Deaton and Muellbauer (1980a) and discussed in Deaton and Muellbauer (1980b). As demonstrated by the authors, this kind of specification is desirable since 'it gives an arbitrary first-order approximation to any demand system it satisfies axioms of choice exactly; it aggregates perfectly over consumers without invoking parallel linear Engel curves; … and it can be used to test restrictions of homogeneity and symmetry' (Deaton and Muellbauer, 1980a: 312).

As originally proposed, the AIDS specification departs from a cost function that represents preferences of the PIGLOG class (see Muellbauer 1975, 1976) which, in turn, takes the form[17]

$$\log c(u,p) = (1-u)\log[a(p)] + (u)\log[b(p)] \tag{III.29}$$

where c is the cost, p the price level and u the utility level lying between 0 (subsistence) and 1 (bliss). Then using the flexible (and arbitrary) functional forms of Deaton and Muellbauer (1980a) for $\log[a(p)]$ and $\log[b(p)]$ the AIDS cost for good k can be expressed as

$$\log c(u,p) = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2}\sum_k \sum_q \gamma^*_{kq} \log p_k \log p_q + u\beta_0 \prod_k p_k^{\beta_k} \tag{III.30}$$

where $\alpha_0, \alpha_k, \beta_0, \beta_k, \gamma^*_{kq}$ are parameters and $\log c(u,p)$ is linearly homogenous in p given $\sum_i a_i = 1$, $\sum_q \gamma^*_{kq} = \sum_k \gamma^*_{kq} = \sum_q \beta_q = 0$. Taking partial derivatives of the cost function in order to obtain quantities for good k, multiplying both sides by $p_k / c(u,p)$ and after some algebraic manipulation used in Deaton and Muellbauer (1980a), the budget share function is obtained as:

[17] Any log function always refers to a natural logarithm.

$$w_k = \alpha_k + \sum_q \gamma_{kq} \log p_q + \beta_k \log(M / P) \qquad \text{(III.31a)}$$

with

$$\log P = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2}\sum_k \sum_q \gamma^*_{kq} \log p_k \log p_q \qquad \text{(III.31b)}$$

with $w_k = p_k c_k / M$. In order to accommodate the specification to standard consumer theory, Deaton and Muellbauer (1980a) show that restrictions to be imposed in the models are the adding-up restriction: $\sum_k \alpha_k = 1$, $\sum_k \gamma_{kq} = 0$ and $\sum_k \beta_k = 0$; the homogeneity restriction: $\sum_q \gamma_{kq} = 0$; and the Slutsky symmetry condition: $\gamma_{kq} = \gamma_{qk}$.

As proposed by the authors, when prices are closely collinear, (III.31b) could be replaced by the Stone (1953) price index: $\log P^* = \sum_k w_k \log(p_k)$, leading to the linear approximation of the AIDS. This simplifying solution is adopted here for estimation purposes. Finally, using the conditional demands version of the AIDS of Browning and Meghir (1991), but allowing conditioning goods (h_r) to affect non-conditional demands only through the intercept, the final specification of the model takes the form:[18]

$$w_\rho = \alpha_\rho + \sum_\omega \gamma_{\rho\omega} \log(p_\omega) + \beta_\rho \log\left(\frac{M}{P^*}\right) + \sum_r \tau_{\rho r}(h_r) + \psi_\rho(\lambda) + e_\rho \qquad \text{(III.32a)}$$

where

$$\log P^* = \sum_\omega w_\omega \log(p_\omega) \qquad \text{(III.32b)}$$

using a general nomenclature for any item ρ (group of goods, markets or markets inside group of goods), λ to identify additional demand shifters, as in Deaton (1997), and e_ρ as the disturbance with all desirable properties: $e_\rho \sim N(0, \sigma^2_{e_\rho})$. Provided there are some modifications

---

[18] The derivation of this expression and others used throughout the thesis are explained in Appendix 1.

in order to derive elasticities (explained in subsequent chapters), the parameters $\gamma_{\rho\omega}$ will be used to identify price effects, β_{ρ} to identify income effects and $\tau_{\rho r}$ to identify linkage effects which, in turn, as discussed in Fortin *et al*. (2000), will implicitly take into account all the labour effects discussed previously.

III.5. Concluding remarks

This chapter has presented a simple theoretical model for informal consumption based on standard demand theory. Two versions of the model considering different preferences ordering have been reviewed. First, the aggregated version, where the household with enough knowledge of the characteristics of informal and formal markets and the supply of goods inside them decides the amount of its budget that will be allocated in each. Second, the disaggregated version, where the family makes the decisions in stages: first, which goods to buy and then where they will be bought (formal or informal markets). The objective of presenting these models was to formalize the main issues investigated in this research: income effects, the linkage effects and own-price and cross-price effects.

In the first case, the testable hypothesis to investigate is whether informal consumption should be classified as a normal or an inferior good (income elasticity higher or lower than zero, respectively). Also, if normal, whether this type of expenditure is closer to consumption necessities (income elasticity lower than one) or consumption luxuries (income elasticity higher than one) should also be tested. Finally, it should be explored whether informal goods are income-inelastic, with lower elasticities, when compared to formal ones (income elastic) in order to investigate the existence of quality differentials associated to markets or conclude if the informal sector is constrained from the demand side. In the second case, the testable hypothesis is to verify that working more hours in informal labour opportunities is associated with more informal consumption baskets. Also, that working more hours at formal jobs biases household consumption towards formal markets. This second case then inspects which formal or informal type of employment (wage-earners or the self-employed) generates the higher linkage effects. In the third case, using the Slutsky matrix, it will be possible to verify whether formal and informal consumption could be classified as imperfect substitutes and to test via the own- and the cross-price elasticities which type of consumption is more price-responsive.

The empirical specification that will be used throughout the research follows an Almost Ideal Demand System (AIDS). Its desirable theoretical properties and its empirical tractability make the AIDS a reasonable choice. The basic specification has been presented here, but some modifications will be allowed in subsequent chapters in order to fit it to the available data. In fact, in view of the database used in this application, the models presented above are estimated using three complementary procedures.

The first procedure, to be presented in Chapter 5, corresponds to the estimation of the aggregate version of the model. The second procedure, to be presented in chapter 6, corresponds to the estimation of the disaggregated version of the model. The two models will offer complementary information. While the former investigates the aggregated behaviour of informal consumption, the latter will explore whether the patterns identified are similar across different goods. However, since there is no information on prices for most of the consumption categories used, Working-Leser Engel curves (Working, 1943; Leser, 1963) instead of full demands are estimated. This means that empirical specifications (III.32a)–(III.32b) are used, but under the standard procedure in cross-sectional work of price normalization. As a consequence, only the income effects and the hours' effects will be computed here. This model is closer to empirical applications encountered in the literature (Fortin *et al*., 2000; Gardes and Starzec, 2002; Reilly *et al*., 2006; Böhme and Thiele, 2012b). The third procedure, to be presented in Chapter 7, is developed to supplement previous findings, providing estimations of price effects, but only for those consumption categories where prices were available. In this application, price availability is restricted to food consumed in the house. Therefore, the Slutsky matrix to be provided is developed in the context of the disaggregated version of the model and conditional on total outlay on food consumption.

IV. THE DATA AND PRELIMINARY EVIDENCE

IV.1. Introduction

In this chapter the main data source used throughout the research is described. The objective is twofold. First, to show the assumptions made in the construction of the dataset, the main variables used and the exclusions. Second, to review the main stylized facts that a descriptive analysis of the data reveals. This chapter is divided in terms of the three crucial variables included in the research: labour supply, consumption and prices (or unit values). Final thoughts are presented as concluding remarks.

IV.2. General considerations

The main database used in the research corresponds to the Peruvian National Household Survey (ENAHO, in Spanish) conducted by the National Institute of Statistics and Informatics (INEI, in Spanish) during 2006. The total sample size is 20,577 households and 96,236 individuals. The inference level of the survey is for the whole country, the main areas (urban and rural), eight geographical regions (the northern, central and southern coasts, the northern, central and southern highlands, the jungle and the metropolitan area of Lima) and 24 departments (sub-national administrative units). The database also provides information for two lower geographical aggregations: provinces (189 in the sample out of a total of 194) and districts (880 out of a total of 1,818). The objective of the research is to provide results for the urban areas, so an initial restriction to the sample was made. The sample size of urban areas is 11,631 households made up of 49,152 individuals.[19]

There are two main reasons for focusing the analysis on urban areas. First, the definition of informality as a working condition is more clearly applicable there. In rural areas, the high presence of small farmers reduces the ability to identify formal or informal relationships without further and possibly non-satisfactory assumptions. By contrast, in urban areas the combination of big and small firms, government activities and several contractual arrangements and registration options adequately identified in the sample help to obtain greater variability in the sample and a clearer identification of the working status of the labour force. Second, the main issue of the research is to study informality from the demand side. Since the identification of informal consumption is made through the classification of places where the household obtain its products, the focus is placed on purchases. In rural areas, markets are more difficult to classify and the greater importance of own consumption reduces the adequacy of studying purchases alone.

[19] See www.inei.gob.pe for methodological issues of the surveys.

Other data sources included in the research are the National Registry of Municipalities (RENAMU, in Spanish) for the period 2004–2006, the Population Census of 2007, the Economic Census of 2008 and other public and private data sources for different years.[20] These data sources were used to construct indicators at the district and provincial levels (lower geographical aggregations). The data are obtained for 2006 (year of the survey), or the closest information available to that year, in order to accurately reflect the cross-sectional variance between districts.

IV.3. Labour supply

The population has been classified in eight mutually exclusive categories: out of the labour force, not working, formal workers in formal firms (formal/formal), informal workers in formal firms (informal/formal), informal workers in informal firms (informal/informal), formal self-employed (self/formal), informal self-employed (self/informal) and family workers.[21] The definitions of these categories for the primary activity are presented in Table VI.1.

Table IV.1: Classifications of the working status of the sample, primary activity

Working status	Definition
Out of labour force	People below 14 years old (labour module of the survey was not applied).
Not working	People in the labour force, but not employed, i.e. do not work at least one hour.
Formal/Formal	Formal wage-earners (have a signed contract) working in formal firms (are registered with the tax authority).
Informal/Formal	Informal wage-earners (do not have a signed contract) working in formal firms (are registered with the tax authority).
Informal/Informal	Informal wage-earners (do not have a signed contract) working in informal firms (are not registered with the tax authority).
Self/Formal	Self-employed registered with the tax authority.
Self/Informal	Self-employed not registered with the tax authority.
Family workers	People working without payment in family-based activities.

In this application, 'Out of the labour force' is assumed to be only for people under the age of 14. People of retirement age (over 65 years old) have been included in the labour force, since it was possible to obtain information about their work situation. It appears that around 30% of people

[20] UNDP (www.pnud.org.pe), INEI (www.inei.gob.pe), Health Ministry (www.minsa.gob.pe), Education Ministry (www.minedu.edu.pe), Transport and Communication Ministry (www.mtc.gob.pe), Economy Ministry (www.mef.gob.pe), Labour Ministry (www.mintra.gob.pe), National Police Office (www.pnp.gob.pe), Macroconsult (2010) and Instituto Cuanto (2007).

[21] When constructing the database, an additional category appeared: formal workers (with a signed contract) working in informal firms (not registered with the tax authority) in 381 cases. This situation is difficult to explain but can appear if the worker has signed a contract with the manager of the firm as an individual. However, given the informality condition of the firm, this kind of contract as a labour relationship is not completely enforcable, so the workers could find themselves in the same situation as those who not signed a contract. Therefore, these observations were added to the informal/informal category.

over 65 years old work at least one hour per week, mainly in the informal self-employment sector. Possibly their retirement situation without adequate retirement benefits is influencing their decision to engage in this sector. The not-working status includes both non-participants and unemployed people, while the employed status is assumed for all individuals who spend at least one hour per week at work.

In the classification of working status, three sets of questions were used. The first enables to classify workers as managers, independent workers, dependent workers, family workers or in domestic service. From these categories, managers and independent workers were classified as self-employed, and dependent workers and those in domestic service as wage-based; family workers remained a particular category (as long as they were not wage-earners). The second question helps to identify whether the firm where the individuals were working was registered or not. The condition of registration (formality) was assumed to be 'holding a juridical personality' in the case of wage-based workers. In the case of self-employment, as well as this question, the special module for independent workers was used. Here, is possible to identify additional registration options which are particularly available for this sector.[22] Therefore, holding one of them was considered as a formal requisite. The third question was used to classify workers according to their (formal) relation with the firm. If they had any contractual arrangement, they were identified as formal. By contrast, informality was defined the lack of contractual arrangements. A direct question about this condition helped to classify all the workers in the wage-based sector.

There were some individuals who it was not possible to classify in these categories. First, there are the missing values in the labour force module (173 cases) or in the module for independent workers (162 cases). Second, independent agricultural workers in urban areas lack most of the information enabling classification in one of the previous categories. In order to reduce the data loss in this case, special questions about the registration of the farms was used (holding a property title was the relevant category for formality). However, for some observations (443 cases) it was impossible to obtain even this information. Third, all managers and independent workers who were owners of firms employing more than ten workers (a small-size enterprise) were also categorized as unclassified (42 cases).

With these issues in mind, the final balances of the sample in terms of the described categories were computed. The results are shown in Table IV.2. 'Total working' corresponds to 23,289

[22] The alternative registration classifications were 'Owner of a business as a natural person' and 'Registered with the tax authority' (RUC for natural persons, RUS, REIR and other systems). A general description of the systems can be found at www.sunat.gob.pe.

observations. From these, at least 57% have some relation to the informal sector (working without a contract or working in an unregistered firm) and 30% can be classified as formal in all of the relations they hold. The rest are family workers (13%). Considering the high figures of informality for the Peruvian economy (around 60%, as mentioned earlier) it is not surprising that this is identified as the dominant sector. At the same time, most of the informal workers are concentrated in the self-employment sector (30% of total workers), while most of the formal ones are working in formal firms (23%).

Table IV.2: Working status in urban areas, primary activity

	Levels	% obs.
Formal/Formal	5,322	22.9
Informal/Formal	1,361	5.8
Informal/Informal	4,952	21.3
Self/Informal	6,922	29.7
Self/Formal	1,631	7.0
Family workers	3,101	13.3
Total working	**23,289**	**100.0**
Not working	12,495	49.9
Out of labour force	12,548	50.1
Total not working	**25,043**	**100.0**
Unclassified	820	1.7
Total sample	**49,152**	**100.0**

Formal: 6,978 (29.9%)

Informal: 13,236 (56.8%)

Different aggregations for the groups presented could be useful for establishing different definitions of informality (excluding family workers). In Table IV.3, the definitions of extended formality and informality correspond to those previously presented and the most aggregated definition of the sector. As shown, only 34% of the people working have no relation with the informal sector while 66% have at least one type of relation. The wage-earning sector constitutes 58% of the sample, while the self-employment sector makes up 42%. In the first case, informality can be constructed either from the worker's perspective (lacking a contract), where informality affects 54% of the wage-earning sector, or from the firm's perspective (not being registered) where informality constitutes 43%. In the second case – self-employment – informality dominates in more than 80% of the cases. Therefore, it is possible to conclude from these figures that the dominant category of the Peruvian labour market is the wage-earning sector. However, this sector does not constitute the formal side of the economy. Only around one half of it can be defined in this way; the other half is exposed in some way to informality. By contrast, in the self-employment sector informality is dominant, though it is still possible to find some formality in this case.

Table IV.3: Working status in urban areas, different aggregations, primary activity

	Levels	% obs.
Total Workers 1/	**20,188**	
Wage-earning sector	**11,635**	**57.6**
From the perspective of the worker		
Formal	5,322	45.7
Informal	6,313	54.3
From the perspective of the firm		
Formal	6,683	57.4
Informal	4,952	42.6
Self-employment sector	**8,553**	**42.4**
Formal	1,631	19.1
Informal	6,922	80.9
Extended Formality 2/	6,953	34.4
Extended Informality 3/	13,235	65.6

1/.Excluding family workers. 2/. Formal/Formal + Self/Formal. 3/. Self/Informal + Informal/Informal + Informal/Formal.

In Table IV.4, basic statistics for workers are presented. Here, hours of work are treated in weekly terms, as the survey records it. This variable has been computed to take into account the normal number of hours a worker spends in his or hers job, in order to avoid any possible division bias. Income, in annual terms, is constructed using most types of labour income received by the worker.[23] This includes: monetary income, in-kind income and own consumption. Income has also been constructed excluding taxes and including most regular benefits (as a concept of net income). Wages are constructed from these estimates and expressed in per-hour terms. Finally, firm size is expressed as the proportion of workers in firms with ten or or fewer employees.

In terms of firm size, (by construction) all individuals in the self-employment sector work alone or own firms that employ fewer than ten workers. However, differences appear in the other labour categories. In general, from the estimates, is possible to infer that, in the wage-based sector, formal firms tend to be larger. As shown, 86% of the individuals in the informal/informal category work in small firms while only 5% in this situation are in the formal/formal category. However, it is interesting to note that hiring without a contract is not necessarily a condition only of small firms. At least 62% of informal workers holding a job in the wage-based sector are working in firms with more than ten employees.

[23] The referential exchange rate for the period according to the Peruvian Central Bank is S/. 3.274 per dollar See www.bcrp.gob.pe. Values of income and expenditure are deflated at the midpoint of the survey period.

Table IV.4: Job characteristics, working status in urban areas, primary activity

	Hours (week)	Income (annual)	Wage (per hour)	Size (% <=10 w.)
Formal/Formal	47.5	12,116	5.9	4.9
	(15.872)	*(10724.100)*	*(8.425)*	
Informal/Formal	47.4	6,433	3.2	37.9
	(19.565)	*(5590.722)*	*(4.278)*	
Informal/Informal	45.9	4,857	2.5	85.8
	(21.190)	*(3781.929)*	*(2.5318)*	
Self/Informal	44.1	5,353	3.2	100.0
	(25.359)	*(6235.926)*	*(7.629)*	
Self/Formal	51.2	11,933	5.5	100.0
	(22.752)	*(17164.060)*	*(8.738)*	
Total workers	**46.2**	**7,619**	**3.9**	**67.2**
	(21.648)	*(9143.784)*	*(7.044)*	
Formal	48.3	12,073	5.8	27.2
	(17.797)	*(12534.170)*	*(8.501)*	
Informal	45.1	5,279	2.9	88.3
	(23.343)	*(5394.997)*	*(5.904)*	
t-test				
Formal vs Informal	*10.92*	*43.15*	*25.28*	
Self (Formal vs Informal)	*11.06*	*15.25*	*9.48*	
Wage (F/F vs I/F)	*0.13*	*18.90*	*11.59*	
Wage (F/F vs I/I)	*4.21*	*46.37*	*28.74*	
Wage (I/F vs I/I)	*2.46*	*9.80*	*6.13*	

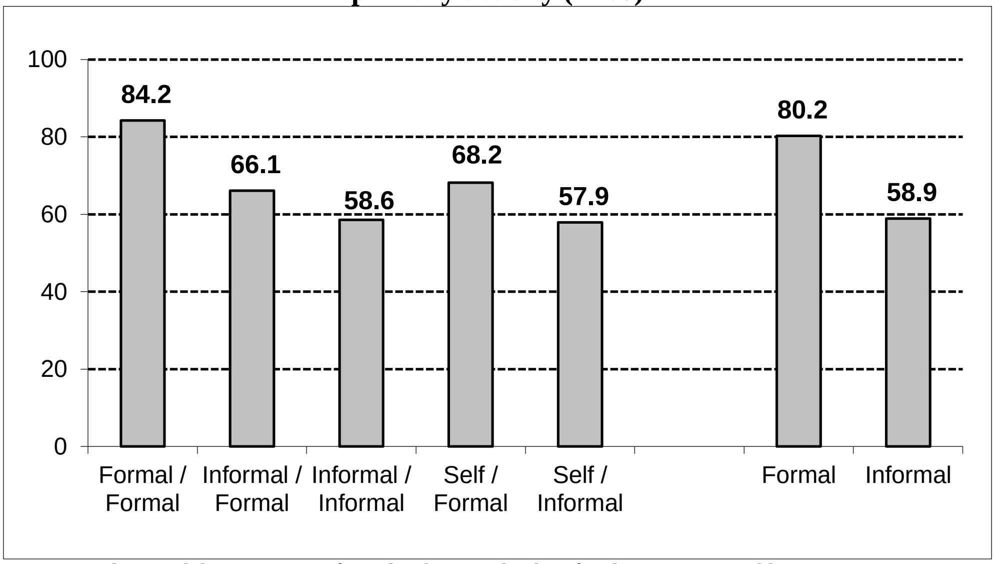
Note: Income and wages are in S/. (*Nuevos Soles*). Standard deviations in parenthesis.

Similarly, we can see that hours worked in the formal sector appear to be higher. This evidence is consistent with previous findings (see, for example, Lemieux *et al*., 1994 for Canada; Gasparini, 2001 for Argentina). The differences are more obvious in the self-employment sector. In the wage-based sector, hours of work tend to increase when the work relation is more formal, but it is interesting to note that the differences are not significant between formal and informal workers when they work in formal firms. This is possibly the result of job-related policies in this kind of firm that do not differentiate between workers with a contract or those without one. At the same time, the standard deviations of hours worked are higher for the informal sector. This result is consistent with the argument that working in the informal sector tends to be more flexible and less subject to institutional constraints (see Lemieux *et al*., 1994).

Patterns on income and wages are similar to earlier ones. Both variables are, in general, greater in the formal sector than in its informal counterparts, with the lowest figures in the case of informal workers at informal firms (similar patterns are found in Perry *et al*., 2007 for other Latin American countries). Arguably, for the aggregate sectors, incomes and wages for the formal sector are double the informal figures. However, in this case standard deviations are larger in the formal sector. This means that incomes and wages are higher, but also more variable in the formal than in the informal one. This result is similar to that shown in Lemieux *et al*., (1994).

To analyse security in incomes, workers were directly asked for the stability of their family incomes, as set out in Figure IV.1. As shown, informal labour opportunities are clearly less stable than formal ones. This is a common result considering that normally informal work is less secure. In terms of the extended definition of informality, while 80% of formal workers consider their family incomes as stable, only 59% of informal workers are in the same situation. Similar differences are encountered even when the analysis is conducted for wage-earners and the self-employed. In the first case, 84% of formal wage-earners in formal firms have stable incomes, while 59% of informal workers at informal firms consider their family incomes as stable. Similarly, although self-employment shows less stable incomes when compared with wage-earners, inside them there are still large differences. For example, 68% of formal self-employed have stable incomes in contrast to the 58% of informal self-employed.

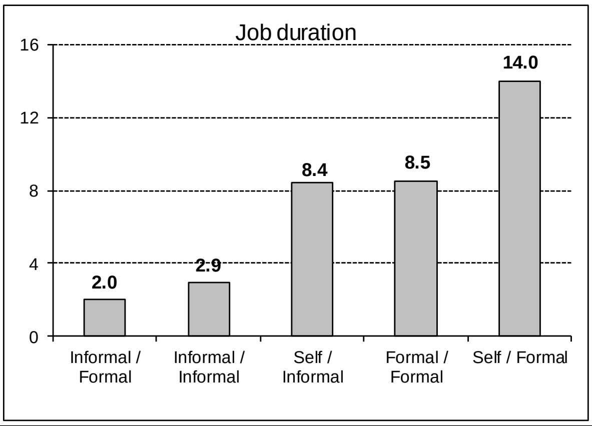
Figure IV.1: Stability of family incomes for different labour options, in urban areas, primary activity (in %)



Note: results reveal the percentage of people who consider their family incomes as stable.

In Figure IV.2, socio-economic characteristics like the age, educational level and gender patterns of the workers are tabulated. As shown, the formal self-employed are the oldest individuals; the younger ones work in the informal self-employment sector. The gender patterns indicate higher informality among women, with particularly high figures in the informal self-employment sector. Where educational attainment is concerned, formal firms tend to hire people with higher educational levels, with those hired under a proper contract showing the highest figures (>14 years). The rest of the sample exhibits lower and very comparable educational attainment: around ten years of education. At the bottom of the distribution, the informal self-employed have the fewest years of education at around 9 years.

Figure IV.2: Other characteristics of workers, in urban areas, primary activity



	Age	Educ	Male	Female
Formal/Formal	38.2	14.3	29	23
Informal/Formal	31.3	12.0	7	6
Informal/Informal	32.1	10.5	26	23
Self/Informal	42.2	9.4	29	42
Self/Formal	48.6	10.5	9	6
Total	38.4	11.2	100	100

Note: Age (in years), job duration (time spent in the last job, in years) and education (years spent in formal education). Male and female figures reveal the percentage of each group of workers under the different earning opportunities.

Under the variable 'job duration', less-permanent jobs are informal ones in the wage-based sector with a duration of around two or three years. By contrast, formal jobs in the wage-based sector tend to reveal longer durations (8.5 years on average). This evidence is consistent with previous analysis undertaken by Pasco-Font and Saavedra (2001), who show, for the beginning of the past decade, that formal workers stayed around 7 years in their jobs and informal ones around 2 years. They also found that jobs in the self-employment sector tended to be more permanent than in the wage-based sector. Similar evidence is found here with self-employed showing average durations of 8.4 yeras vs 5.4 years shown by wage earners but, again, informal opportunities tended to be less permanent than formal ones (here 8.4 vs 14 years).[24] This information clearly reveals some heterogeneity between informal earning opportunities. On the one hand, those working for others without a contract tend to have more risky labour opportunities than those individuals who have decided to run a business, even if this business is informal.

The results explored are more-or-less related to previous observations found by Fields (1990, 2007) about the good (upper-tier) and bad (free-entry) informal earning opportunities. This is the case for Peru, as analysed by Yamada (1996), with the informal self-employment sector and the informal wage-earning sector, respectively. As shown, informal workers in informal firms, although they have highly comparable educational levels to the informal self-employed, they earn less and have less-secure earning opportunities. Similarly, as shown in Table VI.5, the decision

[24] It is possible to argue that duration in formal self-employed jobs could be over-estimated, since respondants could be including their informal self-employed past in their answer. From the survey, it was not possible to track the level of over-estimation.

to work in self-employment is not necessarily related to difficulties in finding other jobs. In the case of informal self-employment, the percentage of people who began a business because of difficulties in entering the wage-based sector is around 37% and for formal self-employment just 25%. Clearly, most of the people working in this sector have decided to do so either based on a free decision of higher expected earnings or simply because they just want to do it.

Table IV.5: Reasons for engaging in the self-employment sector (in %)

	Self/Formal		Self/Informal		Self-employment	
Earn more	25.3		21.1		21.6	
Want to be independent	35.2	75%	26.0	63%	27.2	64%
Family tradition	8.6		5.5		5.9	
Other reason	6.3		9.9		9.4	
Did not find a job	24.6		37.5		35.8	
TOTAL	**100.0**		**100.0**		**100.0**	

Table IV.6: Working status in urban areas, primary activity, household head and spouse (in %)

	F / F	I / F	I / I	S / F	S / I	Total
Formal/Formal	49	5	11	5	31	100
Informal/Formal	27	13	20	2	38	100
Informal/Informal	8	5	31	5	50	100
Self/Formal	27	5	9	13	46	100
Self/Informal	15	5	22	6	52	100
Total	29	5	18	6	42	100

Note: Aggregate categories

	Formal	*Informal*	*Total*
Formal	*51*	*49*	*100*
Informal	*19*	*81*	*100*
Total	*35*	*7*	*100*

Note: head of household in the vertical and spouse in the horizontal.

In Table IV.6, the distribution of labour within the family is presented. The sample used is restricted to those households where there is both a head and a spouse, and where both are working. The sample size in this case is 1,697 observations. As shown, looking the aggregate categories, when the main earner (head of household) works in the formal or informal sector, the second earner (spouse) tends to work more than proportionally in the same sector. Similarly, when the categories are opened, similar patterns are detected for the extremes: formal workers in formal firms (F/F) and informal self-employment (S/I). These results suggest that, in the Peruvian case, network strategies (secondary earners follow primary earners' decisions) are a better description of family interactions, as defined in Lehman and Pignatti (2007). However, when other categories are analysed (I/F, I/I and S/F), coping strategies (changing the labour allocation) emerge, as defined in Galiani and Weinschelbaum (2007) – although this result is influenced by the absolute high incidence of S/I among secondary earners.

IV.4. Informal consumption

To construct formal and informal consumption variables, data on expenditure at the household level are used. A first methodological issue to consider is the data restrictions imposed and the resultant missing observations when both modules (labour supply and consumption) were merged. As a result, 818 households were excluded. At the same time, since the interest is on purchases, those observations with zero-purchasing (5 households) do not provide information so are also excluded. Therefore, 10,808 observations with complete data on consumption and labour were identified.

Table IV.7: Definitions for consumption groups

Group	Definition
Food to be consumed inside the house (Fon)	All food purchased by the household to be consumed within it, including food consumed outside the home by members less than 14 years old and excluding food for domestic animals.
Food to be consumed out of the house (Foff)	All food consumed off the premises by household members older than 14, that could be made in conjunction with other household members or not.
Clothing and personal care (CC)	All expenditure on clothing - including textiles, shoes, fashion accessories and school clothing - and personal care items (soap, shampoo, cream, baby care and hairdressing)
Health goods and services (HEA)	All expenditure on medicines and health services. Includes all kinds of insurance premiums.
Transportation and communication services (TC)	Related expenditure for public and private transportation and communication (phone, internet, mail service). Fuel used for transport is also included.
Education and culture (ED)	Expenditure for regular and non-regular schooling, cultural expenditures and different items (pencils, pens, diskettes) that can be considered closely related to this kind of consumption.
Other goods and not classified items (OT)	Includes a wide range of general household expenditure like housing (rent, taxes due on the house, fuel used for cooking and general services), maintenance (washing and cleaning, painting and related maintenance services), furniture, other services like housekeeping, garage, etc; and all other expenditures not classified elsewhere.

The survey identifies purchases for specific commodities together with the places where these were obtained. In order to perform the analysis, the commodities were divided into seven consumption groups (see Table IV.7). The groups constructed are very close to those used by

INEI.[25] In Table IV.8, a comparison of the breakdown of expenditure between the total and restricted sample is presented. As far as the different consumption strategies of households in urban areas are concerned, the main and most common strategy is purchasing. It represents 76% of total consumption and almost all the households are involved in a purchasing activity. The emphasis is relevant since the models to be estimated are based only on this kind of consumption. The reason is simple: this is the only strategy that can adequately be classified in terms of the informal condition of the place where the good was obtained. Therefore, just focusing the attention on purchases, the classification of expenditure by consumption groups is consistent with earlier estimates (see Herrera, 2003; Monge, 2007) where food consumption (inside and outside the household) explains around 45% of total expenditure and, as a mean budget share, is close to 0.5. At the same time, most of the households purchase from each of the gross groups constructed. However, zero consumption could be problematic in at least four of the seven groups: Foff, HEA and ED. Therefore, corrections to sample selection bias must be made, as proposed in subsequent chapters.

Table IV.8: Total expenditure in urban areas: strategies and groups

Total expenditure by strategy

	Total sample				Restricted sample			
	Mean	**Prop. (%)**	**Bud. Share**	**% > 0**	**Mean**	**Prop. (%)**	**Bud. Share**	**% > 0**
Purchasing	14,979	76.2	0.747	100.0	14,868	76.3	0.748	100.0
	(12236.48)		*(0.18)*		*(12172.17)*		*(0.18)*	
Own & Kind	2,668	13.6	0.134	91.3	2,608	13.4	0.133	91.1
	(4053.87)		*(0.12)*		*(3971.92)*		*(0.12)*	
Private transfer	1,215	6.2	0.074	78.3	1,216	6.2	0.073	78.3
	(2809.41)		*(0.13)*		*(2860.16)*		*(0.13)*	
Public transfer	675	3.4	0.037	73.5	681	3.5	0.037	73.5
	(1648.57)		*(0.06)*		*(1681.4)*		*(0.06)*	
Other	123	0.6	0.008	25.0	123	0.6	0.008	24.8
	(1328.27)		*(0.03)*		*(1372.82)*		*(0.03)*	
Total	**19,660**	100.0		**100.0**	**19,496**	100.0		**100.0**
	(14918.19)				*(14824.24)*			

Total expenditure (purchasing) by group

	Total sample				Restricted sample			
	Mean	**Prop. (%)**	**Bud. Share**	**% > 0**	**Mean**	**Prop. (%)**	**Bud. Share**	**% > 0**
Fon	4,866	32.5	0.359	96.4	4,856	32.7	0.362	97.2
	(3409.71)		*(0.17)*		*(3348.43)*		*(0.17)*	
Foff	1,813	12.1	0.130	85.5	1,790	12.0	0.130	85.4
	(2134.55)		*(0.14)*		*(2109.29)*		*(0.14)*	
CC	1,206	8.1	0.081	95.5	1,221	8.2	0.083	96.3
	(1397.13)		*(0.06)*		*(1391.87)*		*(0.06)*	
HEA	900	6.0	0.054	82.6	887	6.0	0.053	82.1
	(1757.06)		*(0.08)*		*(1761.88)*		*(0.08)*	
TC	2,350	15.7	0.137	94.6	2,316	15.6	0.136	94.5
	(4391.02)		*(0.1)*		*(4449.06)*		*(0.1)*	
ED	1,341	9.0	0.067	86.6	1,304	8.8	0.066	86.7
	(2830.78)		*(0.09)*		*(2721.7)*		*(0.08)*	
OT	2,502	16.7	0.171	99.6	2,494	16.8	0.172	99.6
	(3136.04)		*(0.104)*		*(3146.76)*		*(0.1)*	
Total (purchase)	**14,979**	100.0		**100.0**	**14,868**	100.0		**100.0**
	(12236.48)				*(12172.17)*			

No. observations	**11,631**	**10,808**

Note: expenditure in annual S/. (*Nuevos Soles*). Standard deviations in parenthesis.

[25] See www.inei.gob.pe.

The available data on outlets were then used to classify purchases according to their formal/informal nature. In general, the objective was to identify, as accurately as possible, where a formal or informal transaction is the more likely to happen. However, as long as the data do not provide information on the registration of the transaction done or the institutional enforcements followed by each of the places considered, the classification was carried out based on the appropriate literature (suggestions as to how to classify markets), previous knowledge of the country under investigation and anecdotal evidence collected during the research. Collection of this last piece of evidence was via non-systematic observation of a number of outlets (street vendors, city markets, stores, shops, restaurants, pharmacies, supermarkets) located on both the most- and the least-developed districts of Lima (the capital city), and in cities such as Trujillo, Arequipa, Cajamarca, Huaraz and Juliaca. During the visits, the objective was to look at the external characteristics of the outlet, buy something (in order to record the transaction) and have a short conversation with the seller about the general management of the business.

Two strategies are followed, as shown in Table IV.9. The first considers an aggregation of three markets and the second further disaggregates these three markets into five (from more informal to more formal). As will be explained later, this is done in order to control for possible classification problems and to allow change to be made to the definition of informality as a robust check in the empirical analysis.

Table IV.9: Market classification according to outlet type

<table>
<tr><th colspan="2">Markets</th><th rowspan="2">Outlets</th></tr>
<tr><th>Category I</th><th>Category II</th></tr>
<tr><td rowspan="2">Informal</td><td>Market 1</td><td>street vendor (side-walk sellers, fairs, kiosk-sellers, vehicles), other informal.</td></tr>
<tr><td>Market 2</td><td>city markets.</td></tr>
<tr><td>Semi-formal</td><td>Market 3</td><td>stores, shops, book stores, barber shops, hardware, bazaar, other (includes unclassified markets).</td></tr>
<tr><td rowspan="2">Formal</td><td>Market 4</td><td>bakeries, restaurant, pharmacies.</td></tr>
<tr><td>Market 5</td><td>supermarkets, social centres, educational centres, health centres and other formal.</td></tr>
</table>

The identification of informal outlets follows previous work done by de Soto (1986), Dasgupta (1992), Reilly *et al*. (2006), Arellano and Burgos (2010) and Böhme and Thiele (2012b), who identify this sector as street vendors and city markets – i.e. unregistered and unlicensed establishments (see Swaminathan, 1991). In the survey, the former were identified as side-walk sellers, kiosk-sellers, fairs or in any type of vehicles; the latter was fully identified so no further assumptions were needed. In both cases, the operational characteristics described by Thomas (1992) – competitiveness, ease of entry, reliance on indigenous resources, family-owned businesses, small-scale operations and labour-intensive methods of production with low-skilled workers – were verified by observation and anecdotal evidence. In general, it was also noticed that no receipt or ticket was given when certain direct purchases were made there. Considering these last issues, it was possible to conclude that most of the activities in this sector could be adequately presumed to be unregistered. However, a distinction is made between street vendors and city markets since these latter, given their permanent location in the same place, are subject to greater inspection by local governments; in fact it was possible to verify that they belong to associations and adhere to municipal regulations (like having a licence to operate). This observation is in line with that of de Soto (1986), who argues that commerce in city markets can be considered less informal than that carried out through street vendors. This is recognized under the five-market definition.

Identifying formal outlets was a more difficult task due to the lack of concrete information about the registration of the transactions made in the rest of outlets identified. One option (commonly used in the literature) considers all other transactions as formal and focuses on informal consumption. However, one objective of the research is to also produce information for the formal sector for comparative purposes, so at least an approximation was needed; however, to avoid any misclassification of outlets, construction of both a formal and a semi-formal option was preferred. In the first category, the definition of modern channels of commercialization like supermarkets, as defined in Arellano and Burgos (2010), was used. Institutional providers of goods and services – such as health, educational and social centres – were classified here (in a similar manner to Fomba and Mvolo, 2010; Böhme and Thiele, 2012b). For classification purposes, these were always identified as formal, although note that, in the last three cases, it is possible to encounter informal providers (small and clandestine); however, they are presumably a small part of the market.[26]

Other formal outlets identified were pharmacies, restaurants and bakeries (as in Fomba and Mvolo, 2010), though they are considered less formal than the others once the five- market definition is used. The main reason is that the incidence of informality in these cases could be greater. For example, for pharmacies, Calderon (2010) shows that the split of the market between big pharmacies (i.e. as part of national or international chains) and small ones (individual pharmacies) is 60% and 40%, respectively. Although not all the individual pharmacies can be

[26] Using information from ENAHO, it was observed that, of the total observations carried out in the previous month, only 16% of those in urban zones were done in private centres; again, only a fraction of these are presumed to be clandestine so, for simplicity, private attentions were also classified as formal. In the case of educational centres, ENAHO data reveal that, of those who attended such centres last year, the total share of alumni in private centres was 22%, with only a fraction attending unregulated ones. Again, for simplicity, it was assumed that all purchases made here were formal.

considered as informal, it is likely that the fraction will be greater than for other outlets. A similar structure could be assumed in the case of restaurants, but with an incidence of informality that could be even greater.[27]

The semi-formal category is composed of those outlets not classified previously. The most significant of these were shops and stores. Their definition in the sample was too generic to enable adequate classification, so it was preferable to include them in a similarly generic semi-formal category. Anecdotal evidence reveals that these outlets covered a wide range of formats: small and medium-sized firms and department stores, either in independent locations or in commercial centres. These latter were not necessarily all modern ones (where all stores are formal), by contrast it were identified other organizational structures where a high degree of informality was identified. In fact, modern retail explains only a small proportion of purchasing. Using data provided by Regalado *et al*. (2009), we can see that their share is around 10% of total sales. Therefore, it is possible to infer that, in stores and shops, the incidence of informality could be greater.

The other informal and other formal categories were constructed because some items are not normally commercialized in these typical outlets, but the goods themselves give a good idea of formal or informal consumption. For example, in the case of house expenditure, payments to public security (*serenazgo*) and housing taxes were classified as formal, while general services for the house – normally paid in cash without a receipt (private security, domestic services and others) – were classified as informal. Expenditure on electricity was classified as formal while water was differentiated by the type of connection (if it was a public connection then it was classified as formal). In the case of transport and communication, payments on private parking (normally made in cash without a receipt) were classified as informal and public and private communication services (such as phone calls) as formal. Public transport was directly answered in terms of formality, so was included in the corresponding category. Insurance premiums were assumed to be formal and private transport to the school (normally paid in cash without receipts) informal. Other items were not classified (expenditure on fuel for cooking, maintenance of the building, rent, house improvements and internet services in public points) and were included in the semi-formal category.

[27] Anectodal evidence collected during the research reveals that individual pharmacies show most of the characteristics described in Thomas (1992); however, in general, proper receipts were usually given for transactions and it was possible to check that the pharmacies follow most of the regulations of the sector (like possession of a municipal licence). Similar patterns were detected in the case of small restaurants, but transactions without a receipt were more common, mainly outside Lima.

Table IV.10: Total expenditure (purchasing) in urban areas: markets

Three-market approach				
	Mean	**Prop. (%)**	**Bud. Share**	**% > 0**
Informal	5,496	37.0	0.390	99.4
	(4221.03)		*(0.18)*	
Semi-formal	4,213	28.3	0.318	99.8
	(3930.)		*(0.19)*	
Formal	5,158	34.7	0.292	99.1
	(7734.17)		*(0.18)*	
Total (purchase)	**14,868**	100.0		
	(12172.17)			

Five-market approach				
	Mean	**Prop. (%)**	**Bud. Share**	**% > 0**
Market 1	2,398	16.1	0.174	98.4
	(2323.02)		*(0.13)*	
Market 2	3,098	20.8	0.216	90.4
	(3014.57)		*(0.17)*	
Market 3	4,213	28.3	0.318	99.8
	(3930.)		*(0.19)*	
Market 4	1,584	10.7	0.101	88.8
	(2115.82)		*(0.11)*	
Market 5	3,575	24.0	0.191	98.0
	(6692.99)		*(0.15)*	
Total (purchase)	**14,868**	100.0		
	(12172.17)			

Overlapping of the consumer base		
…the consumer purchase from (%):		
3-market definition	Only one market	0.2
	Two markets	1.4
	The three markets	98.4
5-market definition	Only one market	0.2
	Two markets	0.6
	Three markets	3.0
	Four markets	16.1
	The five markets	80.1
No. observations	**10,808**	

Note: expenditure in annual S/. (*Nuevos Soles*). Standard deviations in parenthesis.

In Table IV.10 the results of this classification are presented. Using the restricted sample and focusing solely on the three-market approach (top panel), informal consumption explains 37% of total purchases, and formal consumption around 35%. These figures are lower when compared to the size of the shadow economy – around 60% of GDP in 2006 (Schneider and Buehn, 2007) – which is a reasonable result considering that only a fraction of informal household consumption is considered in the classification. In fact, if the informal consumption constructed here is extended to include own consumption, private transfers and other sources and if we further assume that the semi-formal category is distributed equally between formal and informal markets, then the distribution of total consumption is 60% for informal transactions and 40% for formal ones – a result highly comparable to GDP figures. In terms of the mean budget shares, informal purchases represent 0.4, which is higher than the 0.3 identified for formal ones. Moreover, both markets show very high and similar rates of participation (close to 100%), indicating the strong overlapping consumer base between the formal and informal sectors, as identified by Böhme and Thiele (2012b) for African countries. This overlapping is more evident in the results presented in the lower half of Table IV.10 where as shown almost everyone in the sample purchase from the three markets.

In terms of the five-market approach (the middle section of Table IV.10), it follows from the data that the main purchasing places are Market 3 and Market 5, representing 28% and 24%, respectively, of total expenditure. However, we should note that both categories include a very different array of places, so it does not necessarily mean that the specific places included in these categories are the main sources of goods for Peruvian households. In fact, the main locus of consumption is the city market (Market 2) with 21% of the consumption purchased there. Therefore, this outlet constitutes the main channel of engagement with the informal sector. Another interesting case occurs in Market 1 (street vendors), with more than 98% of participation and 16% of total expenditure; considering the low scale of consumption in this market, these ratios enable us to infer the relatively high importance of this sector. Again, the participation rate and the information presented at the bottom of Table IV.10 help to demonstrate the strong overlap, even if the definition of informal markets has changed.

The results of informal consumption can be compared with international evidence. In this application, for the Peruvian economy the participation rate in informal outlets is close to 100% and budget shares are closer to 40%. Fortin *et al*. (2000) calculate a participation rate of 15% and a budget share of 1.4% in Canada. Gardes and Starzec (2009), in Poland, show a similar participation rate, but do not present results for the budget share. Reilly *et al*. (2006) show a 84% participation rate and a budget share of 28.5% for Serbia, whereas Böhme and Thiele (2012b) showed no participation rates, but revealed budget shares of 78% on average for West African countries such as Benin, Burkina Faso, Mali, Niger, Senegal and Togo. In principle, the differences can be explained by the size of the informal economy in each country. In Canada, the size of the shadow economy is calculated at around 12% of GDP (Schneider and Buehn, 2007)[28] or between 2% and 5% (Fortin *et al*., 2000). In Poland, the size is almost 26% (Schneider and Buehn, 2007), in Serbia around 33% (Krstic, 1998) for ex-Yugoslavia, and in African countries around 42% on average (Schneider and Buehn, 2007).[29] Therefore, Peruvian consumption figures are higher than those for Canada, Poland and Serbia, but below those for Africa. It is also interesting to note that, as we saw for Peru, household consumption shares are below GDP shares in three of the four studies reviewed here.

[28] The figures of Schneider and Buehn (2007) are the closest to the date of the surveys of the studies referred to.
[29] Calculations exclude Senegal since no results are reported for this country.

However, methodological differences also deserve some commentary. First, Böhme and Thiele (2012b) will be always higher than other estimates as long as the authors include own consumption as part of informal expenditure, something that is avoided by Fortin *et al*. (2000), Gardes and Starzec (2009), Reilly *et al*. (2006) and here, as long as the focus of attention is only market purchases. Second, as already stated, the results of Fortin *et al*. (2000) and Gardes and Starzec (2009) are based on subjective responses to unregulated transactions, so they are not necessarily comparable to the market classifications shown here, in Böhme and Thiele (2012b) and Reilly *et al*. (2006). In principle, an individual should have good knowledge of the registration of transactions made, but under-reporting could emerge if a stigma prevents them from revealing the total amount of their informal purchases. This second problem is solved via market classification, but incomplete information on registration will generate the downward bias of informal estimates, too.

Therefore, for the purposes of this research, we must recognize that some measurement error in the dependant variables remains. As discussed in Wooldridge (2002), this is not necessarily a problem as long as the measurement error is uncorrelated with the independent variables. However, if this is not the case (and we should presume that it is not), the new error structure will cause a simultaneity bias. As a consequence, IV strategies of estimation must be used in the empirical chapters.

Table IV.11: Composition of informal and formal baskets (three-market definition)

	Informal		Semi-formal		Formal		Total (purchase)	
		%		%		%		%
Fon	3,063	55.7	1,348	32.0	444	8.6	**4,856**	32.7
Foff	338	6.2	518	12.3	933	18.1	**1,790**	12.0
CC	483	8.8	637	15.1	100	1.9	**1,221**	8.2
HEA	57	1.0	68	1.6	761	14.8	**887**	6.0
TC	891	16.2	148	3.5	1,276	24.7	**2,316**	15.6
ED	132	2.4	201	4.8	971	18.8	**1,304**	8.8
OT	530	9.6	1,292	30.7	672	13.0	**2,494**	16.8
Total (purchase)	**5,496**	**100.0**	**4,213**	**100.0**	**5,158**	**100.0**	**14,868**	**100.0**

Note: expenditure in annual S/. (*Nuevos Soles*).

In Tables IV.11 and IV.12, the distribution of consumption groups are presented for both the three-market and the five-market definitions. As expected and discussed in Fortin *et al*. (2000) and Böhme and Thiele (2012b), there are some goods which are more commonly obtained from informal sources and others from formal ones. This information must be incorporated in demand models, since preferences of purchasing from one or other market will also reveal decisions on purchasing particular goods only sold there. In Table IV.13 a deeper analysis of the consumption groups in terms of the formal, semi-formal and informal markets is presented. The table shows the results for the sample of non-zero expenditure on the group analysed, therefore they are interpretable conditional on the decision to purchase a particular good (i.e., given that the household has decided to consume it). This exercise is done only using the three-market definition, given the lack of information for some group/market combinations when the five-market definition is used. This is also why the disaggregated models are run only on the three-market approach.

Table IV.12: Composition of informal and formal baskets (five-market definition)

	Market 1		Market 2		Market 3		Market 4		Market 5		Total (purchase)	
		%		%		%		%		%		%
Fon	448	18.7	2,615	84.4	1,348	32.0	232	14.6	212	5.9	4,856	32.7
Foff	338	14.1	-	-	518	12.3	933	58.9	-	-	1,790	12.0
CC	254	10.6	229	7.4	637	15.1	45	2.8	55	1.6	1,221	8.2
HEA	57	2.4	-	-	68	1.6	373	23.5	389	10.9	887	6.0
TC	891	37.2	-	-	148	3.5	-	-	1,276	35.7	2,316	15.6
ED	112	4.7	20	0.6	201	4.8	-	-	971	27.2	1,304	8.8
OT	296	12.4	234	7.6	1,292	30.7	1	0.1	671	18.8	2,494	16.8
Total (purchase)	**2,398**	**100.0**	**3,098**	**100.0**	**4,213**	**100.0**	**1,584**	**100.0**	**3,575**	**100.0**	**14,868**	**100.0**

Note: expenditure in annual S/. (*Nuevos Soles*).

Table IV.13: Purchasing markets by expenditure groups

	Mean (>=0)	**Mean** (>0)	**B. Share** (>=0)	**B. Share** (>0)	**% > 0**	**Prop. (%) (>=0)**	**Sample**
Fon (>0)	**4,998**					**100.0**	**10,501**
Informal	3,153	3,264	0.600	0.621	96.6	63.1	
Semi-formal	1,388	1,493	0.329	0.354	92.9	27.8	
Formal	457	965	0.071	0.150	47.3	9.1	
Foff (>0)	**2,097**					**100.0**	**9,226**
Informal	396	695	0.264	0.463	57.0	18.9	
Semi-formal	607	1,909	0.212	0.666	31.8	29.0	
Formal	1,093	1,427	0.524	0.684	76.6	52.1	
CC (>0)	**1,267**					**100.0**	**10,412**
Informal	502	623	0.416	0.517	80.5	39.6	
Semi-formal	661	706	0.510	0.545	93.7	52.2	
Formal	104	350	0.073	0.246	29.8	8.2	
HEA (>0)	**1,080**					**100.0**	**8,878**
Informal	70	373	0.065	0.347	18.7	6.5	
Semi-formal	83	218	0.154	0.406	38.1	7.7	
Formal	927	1,036	0.781	0.873	89.4	85.8	
TC (>0)	**2,449**					**100.0**	**10,218**
Informal	943	1,060	0.467	0.525	89.0	38.5	
Semi-formal	157	251	0.095	0.151	62.5	6.4	
Formal	1,350	1,628	0.438	0.528	82.9	55.1	
ED (>0)	**1,504**					**100.0**	**9,374**
Informal	152	209	0.288	0.394	73.0	10.1	
Semi-formal	232	338	0.241	0.351	68.7	15.4	
Formal	1,119	1,307	0.471	0.550	85.6	74.5	
OT (>0)	**2,503**					**100.0**	**10,770**
Informal	532	624	0.199	0.233	85.2	21.3	
Semi-formal	1,296	1,314	0.495	0.502	98.7	51.8	
Formal	675	730	0.306	0.331	92.4	27.0	

Note: expenditure in annual S/. (*Nuevos Soles*). (>=0) means that both zero and non-zero results are considered; (>0) means that only non-zero results are considered.

Using this information we can comment on the different channels via which the families engage in the formal and informal markets. It is clear that the most noticeable informal consumption group is food in the household (Fon). Given that this sector constitutes the main consumption group in the Peruvian economy, it will also be the main channel for engaging in informal transactions. This is not a surprising result, considering that the main purchasing place for Peruvian households is the city market, where food is the main commercialized good. These figures are also consistent with INEI (1992), which shows (for 1986) that the main category of products sold by informal outlets is food: 60% of informal vendors are dedicated to its sale. These figures have probably not changed over time. The second-biggest channel is clothing and personal care (CC) – again, an unsurprising result considering the importance of the sale of such goods (mainly personal care products) in city markets.

By contrast the most formal goods are health and education. This is probably associated with the nature of these consumption groups where, as explained, health services are provided mostly by public or private (formal) hospitals and education by public or private (formal) schools. Therefore, only related goods such as medicines, books or stationery have relevant informal markets, where it is also difficult to verify a dominant influence, although it is true that the participation rates of education in the informal market are also associated with the leisure and cultural expenditure included in this group. Finally, a third channel to formal markets is food eaten outside the household, as explained by the importance of restaurants in this category. In the case of transport and communication (TC), it seems that they are evenly distributed between formal and informal markets. Although informal transportation dominates public transportation services in Peru (see INEI, 1992; Ghersi, 1997; Bielich, 2010), formal private transportation and communication services seems to compensate the figures for the group. Interestingly enough, the gross results are highly comparable to those of Böhme and Thiele (2012b), with informal channels of commercialization more important in the case of food, beverages and clothing, less important in health and education, and more or less evenly distributed in the case of transportation and communication.

A final issue to consider, given its relevance in subsequent chapters, is the relation observed between working and purchasing from informal markets. A first look at the data is presented in Table IV. 14, where the mean budget share of formal and informal consumption is presented. The value has been calculated for the different sub-groups of the population according to the sector of labour allocation (those already presented in the last section). The analysis takes into account the hours a household allocates to each particular earning opportunity. The results correspond to cases where the household allocates zero hours (Hj=0), where it allocates positive hours (Hj>0) and where the sector under analysis is the only earning opportunity in the household (Hj>0 / Hi=0).

Table IV.14: Formal and informal consumption shares by earning opportunities

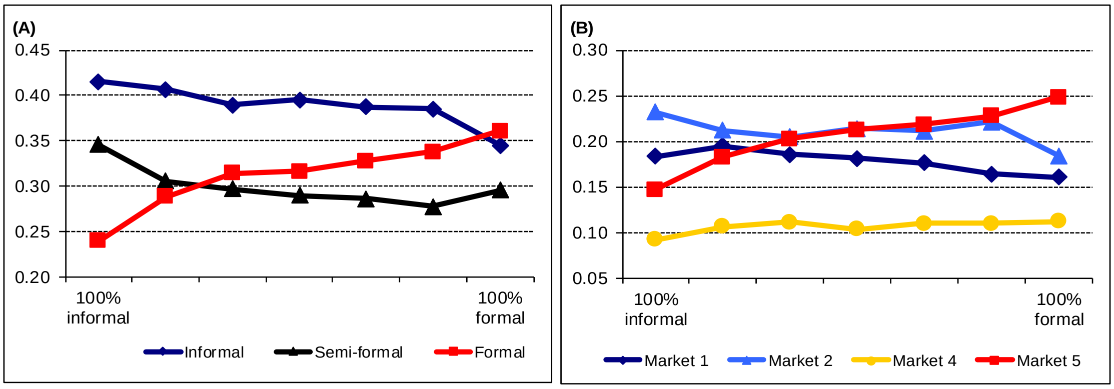
	Informal consumption			Formal consumption		
	Hj=0	Hj>0	Hj>0 / Hi=0	Hj=0	Hj>0	Hj>0 / Hi=0
Formal/Formal	0.401	0.368	0.341	0.262	0.347	0.372
Informal/Formal	0.389	0.398	0.375	0.292	0.288	0.286
Informal/Informal	0.385	0.401	0.383	0.312	0.248	0.220
Self/Formal	0.394	0.362	0.328	0.286	0.328	0.348
Self/Informal	0.363	0.420	0.429	0.317	0.264	0.252
Formal	**0.408**	**0.368**	**0.339** (A)	**0.253**	**0.339**	**0.371** (C)
Informal	**0.350**	**0.408**	**0.412** (B)	**0.354**	**0.264**	**0.240** (D)

t- tests for Hj>0 / Hi=0	(A) vs. (B)	-15.16		(A) vs. (C)	-5.66
	(C) vs. (D)	26.45		(B) vs. (D)	-42.75

Note: Formal and informal consumption is defined using the three-market categorization. Hj defines the earning opportunity under analysis and Hi the others in the table. Aggregate formal and informal jobs follow the extended definition of Table IV.3. T-tests are calculated for unknown and heterogeneous population variances for samples without replacement.

Focusing on the overall formal and informal earning opportunities, it follows that those households that have a connection with the informal sector via labour supply also allocate a higher budget share to consumption in this sector. However, at the same time we can conclude that even formal (informal) workers have at least some informal (formal) consumption, evidence which is in line with the high overlapping of the consumer base presented earlier. For example, families with only informal workers allocate 41.2% of their budget on informal goods and 24.0% on formal ones, while the figures for families with only formal workers are 33.9% and 37.1%, respectively. For more-specific job opportunities, similar patterns emerge. Those with higher shares on informal consumption are families with workers only in the informal self-employment sector (42.9%) followed by workers without a contract in unregistered firms (38.3%). Also, those with higher formal consumption are families with only formal workers at formal firms (37.2%) followed by the formal self-employed (34.8%).

Figure IV.3: Formal and informal consumption shares by hours worked in each sector



Note: Aggregate formal and informal jobs follow the extended definition. The vertical axis stands for consumption budget shares on each market and the horizontal axis for the proportion of hours of work that the family allocates in each working sector.

Similar evidence is presented in Figure IV.3. Here, the horizontal axis shows the percentage of working time expended by the household on informal or formal activities; the vertical axis indicates the consumption shares for formal, semi-formal and informal goods. Panel (A) shows the results for the three-market and Panel (B) for the five-market definition. In the five-market approximation, Market 3 is excluded, since it is similar to the semi-formal market already presented. Therefore, the calculations given here suggest that linkages between working and purchasing in the informal or formal sector could be important. However, additional information can be gleaned from these figures. First, informal consumption is more stable across earning opportunities compared to formal expenditure. Second, the behaviour of semi-formal shares (in this uni-dimensional analysis) is more comparable to informal figures. Third, when the markets are opened using a five-category approach, we can see more clearly the higher response of the extremes (Market 1 and Market 5) than the intermediates (Market 2 and Market 4).

IV.5. Food prices (unit values) estimation

The available data on food prices (unit values) will help to implement an in-depth analysis for this particular category in order to reveal the substitution relationships between formal and informal markets. Although the analysis is not carried out for overall formal and informal consumption or for other categories, the relatively high importance of food in Peruvian purchases could be at least a good approximation of the phenomenon. Also, as long as food consumption is more or less homogeneous across households, focusing only on food will also facilitate interpretation of the results.

The details of the structure of food consumption have been already commented on in a previous section, where we saw that this consumption accounts for around 30% of total purchases of Peruvian urban households. Also, it is biased in favour of informal markets, which explains more than 60% of total purchases, with formal ones below 10% and rates of participation of 97% and 47% in informal and formal markets, respectively. Although these differences are large, the joint consumption (households purchasing both formal and informal markets) is about 46% of total food consumers, which is considered adequate for price analysis. However, the main difficulty is that, although homogenization across food is good, it is not perfect. The composition of the food basket in terms of food groups is presented in Table IV.15. Therefore, since the composition of the formal and informal baskets is not exactly the same, some influence of goods within the consumption groups is still allowed and needs to be recognized in the interpretation of the results in the corresponding chapter.

Table IV.15: Composition of the formal and informal consumption baskets

	Informal		Semi-formal		Formal		Total (Fon>0)	
		%		%		%		%
Bread and Cereals	442	14.0	395	28.4	137	30.0	**974**	**19.5**
Meat and Fish	965	30.6	199	14.3	65	14.3	**1,229**	**24.6**
Milk, Eggs, Fats	385	12.2	321	23.1	67	14.6	**772**	**15.4**
Fruits	364	11.5	33	2.4	14	3.1	**411**	**8.2**
Vegetables	658	20.9	97	7.0	19	4.2	**774**	**15.5**
Other	339	10.8	344	24.8	154	33.7	**837**	**16.8**
Total (Fon>0)	**3,153**	**100.0**	**1,388**	**100.0**	**457**	**100.0**	**4,998**	**100.0**

Note: expenditure in annual S/. (*Nuevos Soles*).

To avoid any bias in the results, a correct identification and estimation procedure of the price vectors for formal and informal markets must be followed. In this section, the methodology used to compute prices as well as the robust checks performed are presented. In order to maximize the amount of information used in price computation, this analysis is restricted only to the three-market definition. In this context, the survey distinguishes between expenditure on and quantities of 49 food products (see below), for which it was possible to obtain the implicit unitary values.

These unitary values are the best available price data for the study but, as emphasized by Deaton (1997), they may be subject to important measurement error and quality effects that can bias the results. To solve this, there are techniques that partially control for this bias during estimation. One is proposed by Deaton (1988, 1990, 1997) and is based on a stripped-down model of demand and unit values. Although intuitive, when tested the procedure did not perform well. Systematic tests of Deaton's procedure are done by Niimi (2007), showing that the ability of the technique to reduce bias is very limited. According to Niimi, when the computed elasticities are compared to estimates using market prices, the differences are still high. We cannot, therefore, conclude that the procedure provides a significant reduction in bias in its demand estimation. In fact, during her empirical application, Niimi (2007) shows a better performance when community mean unit values are used, which motivates her to recommend this alternative as a second-best option in the absence of adequate market price data. Niimi's recommendation is particularly attractive in this context, where the stripped-down model with censoring implies the estimation of several parameters, something that could complicate the econometric application. Therefore, the kind of work to be done here is to compute mean unit values at the district level (lower geographical disaggregation) and use them as prices.[30]

The procedure entails the computation of the household unit values for each of the products in each of the markets where these products can be bought.[31] Then, the district means for each of the

[30] In this research, the terms 'unit values' and 'prices' are used as synonyms although they are not.
[31] For the imputation and homogenization of price data between different quantity units, INEI methodologies were followed (www.inei.gob.pe).

product/market combinations is estimated and the combinations are placed into six broad food subgroups (see Table IV.16) so similar sub-group/market combinations can be computed. Finally, the aggregate district unit values for each market are estimated (see Table IV.17). This procedure, done in steps (later called the 4-step procedure), will help to maximize the identification of market prices as long as within a district at least one product in one market is purchased reducing loss of information when averages are taken. The aggregation of products into sub-groups follows the methodologies explained in INEI (2001) and INEI (2008), and each price vector is computed using geometrical means with weighting factors defined at the district level. Niimi (2007) shows that this weighting strategy outperforms reduction of the commented bias associated with unit values, so this will be the base price vector considered in this estimation. However, Niimi also comments that the results could be sensitive to the choice of weights, so other alternatives are explored with weighting factors at the national and regional (departmental) levels. These will also be used during our estimation, but only for the purposes of comparison.

Table IV.16: Food-on consumption sub-groups

Sub-group	Products
Cereals	Bread, cakes, rice, maize and derivatives, wheat and derivatives, quinoa and derivatives, noodles.
Meat and Fish	Red meat, chicken and other white meat, sub-products of meat, sub products of chicken, sausages, mliver, tripe, fresh fish, canned fish, shellfish and other sea products.
Milk, Eggs and Fats	Milk, eggs, cheese, margarine, butter, other milk products, oil and fats.
Fruits	Lemon and lime, tangerine, orange and papaya, banana, other fruits.
Vegetables	Potatoes, hot pepper, grains and derivatives, onion, tomato, carrot and pumpkin, corn, sweet potato, yucca and melloco, other vegetables.
Others	Alcoholic drinks, non-alcoholic and soft drinks, water and juices, sugar, salt, spices, coffee, tea, cocoa, herbs and coca, sweets other food, prepared food, unclassified food.

Table IV.17: Food-on market classification according to type of outlet for food consumption

Market	Outlet
Informal	street vendors, vehicles, fairs, city markets.
Semi - formal	shops and others.
Formal	Bakeries, restaurants and supermarkets.

There are 147 individual prices to be computed (49 products in 3 markets). A first issue is to identify the number of data points available for estimation, given that there must be at least one household which reports a unitary value for each of these 147 products over the whole sample. If not, some products will need to be dropped from the sample. The total number of households available for estimation (those which report at least one price over all urban households) is 11,212 and the responses per product is in the range 16–8,559, with the product having the least number of household responses being 16 and the highest 8,559.

With these data, then, the second issue to solve is to identify the possible outliers and replace them. Following Niimi (2007), outliers were identified as those that are more than five standard deviations away from their means. The analysis is done at the household level and 858 data points out of 313,267 are classified as outliers (less than 1%). Given their lower incidence, the treatment of outliers will not be a source of bias in this application. However, their replacement was considered necessary. Thus, the mean values of the immediately superior geographical aggregation are used, until the problem is solved. Therefore, since the data used are at the household level, the first aggregation was districts, followed by provinces, departments, zones, areas and, finally, the whole country.

Once outliers have been replaced, estimation of the mean unitary values at the district level is performed (the values for households in the same district are averaged). However, several missing values appear. This happens because, in some districts, certain products may not be consumed or – in terms of the identification of markets – some simply do not exist. From the 52,479 data points to be identified (147 prices in 357 districts), there are 'real' observations for 27,758 (53%) of them. With these data, the 4-step aggregation procedure commented earlier is performed, with the objective of reducing the amount of imputation needed when market prices are identified. In fact, once the final stage is reached (aggregated markets), only 55 out of 1,071 data points (5%) were found to be missing, most of them (53 data points) corresponding to formal markets. This result was to be expected given the lower participation rates displayed previously for formal food consumption. As long as a demand system cannot be estimated with missing prices, imputation will be needed. The strategy is the same as for the replacement of outliers: the mean values of the immediately superior geographical aggregation are used, until the problem is solved. Finally, with the complete vector of district prices, geographical identification codes are used to merge this database with the household sample.

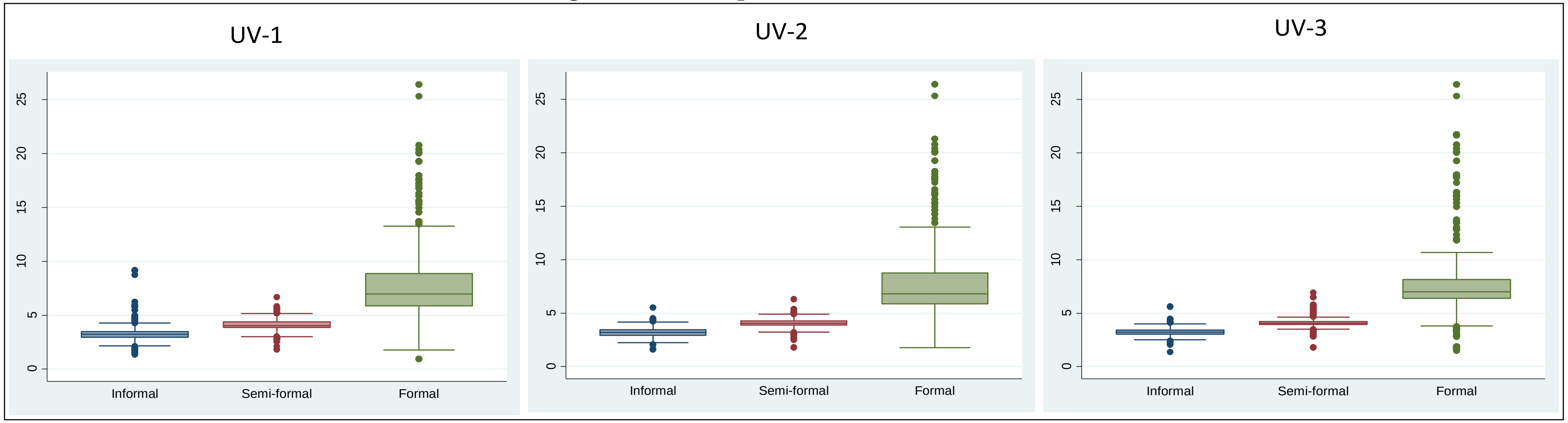
Using this final database, the results for the price vectors constructed are presented in Table IV.18 and Figure IV.4. The results correspond to unit values once outliers have been replaced, and both the results before and after the imputation of missing values are displayed. Note that, in the case of food consumption groups, only the versions pre-imputation are shown, since the replacement of missing values was performed only at the market level.

Table IV.18: Results by unit value vector

			Fon: Food to be consumed inside the house (post-imputation)	Fon: Food to be consumed inside the house (pre-imputation)	Bread and cereals (pre-imputation)	Meat and fish (pre-imputation)	Milk, eggs and fats (pre-imputation)	Fruits (pre-imputation)	Vegetables (pre-imputation)	Others (pre-imputation)
UV-1	District weights	Informal	**3.246**	3.246	2.422	6.034	4.917	1.439	1.525	5.888
			(0.589562)	*(0.589562)*	*(0.475283)*	*(0.732015)*	*(1.208971)*	*(0.270463)*	*(0.268034)*	*(2.221092)*
		Semi-formal	**4.101**	4.101	3.158	7.687	5.299	1.608	1.786	4.612
			(0.460803)	*(0.460803)*	*(0.710708)*	*(1.642458)*	*(0.461668)*	*(0.645529)*	*(0.456398)*	*(1.016956)*
		Formal	**7.613**	7.592	4.287	10.321	6.272	3.208	2.343	12.579
			(2.953409)	*(2.912742)*	*(1.672881)*	*(2.142203)*	*(2.520382)*	*(2.750535)*	*(0.833985)*	*(3.655869)*
UV-2	Departmental weights	Informal	**3.222**	3.222	2.409	6.028	4.900	1.438	1.522	5.912
			(0.403718)	*(0.403718)*	*(0.449303)*	*(0.636204)*	*(1.211180)*	*(0.273070)*	*(0.267367)*	*(1.977229)*
		Semi-formal	**4.079**	4.079	3.127	7.487	5.269	1.601	1.754	4.553
			(0.346229)	*(0.346229)*	*(0.632594)*	*(1.312761)*	*(0.429951)*	*(0.618624)*	*(0.394486)*	*(0.807169)*
		Formal	**7.526**	7.500	4.223	10.238	6.190	3.208	2.368	12.569
			(2.872085)	*(2.829222)*	*(1.528170)*	*(2.045902)*	*(2.529059)*	*(2.746278)*	*(0.872204)*	*(3.612467)*
UV-3	National weights	Informal	**3.248**	3.248	2.423	6.067	4.692	1.430	1.511	6.036
			(0.324242)	*(0.324242)*	*(0.351441)*	*(0.749053)*	*(1.171600)*	*(0.239368)*	*(0.214434)*	*(1.746712)*
		Semi-formal	**4.093**	4.093	3.092	7.310	5.240	1.555	1.759	4.534
			(0.321504)	*(0.321504)*	*(0.486138)*	*(1.204125)*	*(0.403798)*	*(0.552361)*	*(0.379970)*	*(0.521527)*
		Formal	**7.342**	7.322	4.189	10.294	6.182	3.209	2.341	12.893
			(2.214346)	*(2.162436)*	*(1.488874)*	*(2.082579)*	*(2.451883)*	*(2.747338)*	*(0.846651)*	*(3.537252)*

Note: Results in S/. per kg, correspond to the mean values for the 10,501 observations identified for Fon consumption. Standard Deviations in parenthesis.

Figure IV.4: Box-plots of unit value vectors



The figures show the expected results: informal prices are lower than formal ones. At the same time, there are no noticeable differences in the estimates either according to the different methodologies or when imputation is or is not taken into account. On average, informal prices are around 57% lower than formal ones and show a higher dispersion. At the same time, semi-formal markets seem to be an intermediate category between these two extremes. These results are also verified when prices are computed for each of the food sub-groups considered in the estimation except for the ‘others’ category, where the great variety of goods included could possibly be affecting point values. For the disaggregated results, the highest gap between formal and informal markets seems to be on fruit (around 55%) and the lowest gap on milk, eggs and fats (around 22%).

Without appropiate market prices, it is difficult to assess the robustness of the calculation made here. However, a crude way to verify accuracy is to use market prices for food groups and to replicate the unit value computation at that level. For that purpose, data on market prices were obtained from ENAPROM (Encuesta Nacional de Propósitos Múltiples) for 1994 and constructed using INEI methodologies (see INEI, 2001, 2008). The results were translated into 2006 terms using departmental inflation rates and then merged into the ENAHO dataset. Items contained in each data source vary, with more-detailed data in ENAPROM than in ENAHO. In fact, ENAPROM includes information on 941 food items compared to ENAHO’s 186, but both sources can be arranged in terms of the 49 food products discussed earlier, showing significant overlap in terms of the food sub-sub-groups.

Although this exercise demonstrates that the market prices computed here are comparable with unit values, there are at least three limitations to the data and subsequent exercises. First, information on market prices does not distinguish between markets (formal or informal), so comparisons made here are only applicable to sub-group and overall food prices. Second, the data are only available at the departmental level (i.e. there are only 25 unique values for each market price computed). Thus, in order to compare unit values with market prices, mean departmental unit values were calculated, with the result that the quantity of data used to perform the tests is dramatically reduced, with the natural implications that this will have on statistical precision. Third, the basis for the calculation of market prices is a dataset for 1994 (more than ten years’ difference with the sample used for estimation purposes), so changes in tastes or consumption patterns coiuld account for some ‘measurement errors’ in market prices. For all these reasons, it is preferable to interpret the results of the robust checks not as conclusive, but as good references with the best available data.

Mean comparison and correlation could be performed using these data. Departmental mean comparisons are shown in Table IV.19. As expected, unitary values are not market prices so there are differences detectable at the mean value of each estimate. However, it is possible to verify that the calculation of unit values done here shows results close to market prices. The average mean value difference is around 7% when food prices are computed and around 4% when mean differences across consumption groups are simply averaged. Similar gaps computed in Niimi (2007) were around 18% on average (using comparable methods to those used here in the computation of prices). At the same time, in contrast to the results presented by Niimi, in this application most of the differences are not statistically significant at the conventional level; however, considering the small sample used here, it would be preferable to consider this last result only as referential.

Table IV.19: Unit values vs market prices (mean values and mean differences)

	Mean value							Mean difference (%)		
	UV-1		UV-2		UV-3		MP	UV-1	UV-2	UV-3
Fon: Food to be consumed inside the h	3.575	***	3.581	***	3.623	***	**3.861**	-7.4	-7.3	-6.2
	(0.403)		*(0.392)*		*(0.267)*		***(0.499)***			
Bread and Cereals	2.781		2.784		2.820		**2.710**	2.6	2.7	4.1
	(0.305)		*(0.318)*		*(0.264)*		***(0.469)***			
Meat and Fish	6.254	***	6.266	***	6.299	***	**7.356**	-15.0	-14.8	-14.4
	(0.521)		*(0.501)*		*(0.599)*		***(0.999)***			
Milk, Eggs, Fats	5.023		5.003		4.865		**4.977**	0.9	0.5	-2.2
	(0.418)		*(0.439)*		*(0.524)*		***(1.024)***			
Fruits	1.426	**	1.427	**	1.428	**	**1.607**	-11.2	-11.2	-11.1
	(0.265)		*(0.256)*		*(0.216)*		***(0.397)***			
Vegetables	1.518		1.515		1.510		**1.535**	-1.1	-1.3	-1.6
	(0.276)		*(0.271)*		*(0.204)*		***(0.293)***			
Other	5.775		5.811		6.018		**5.955**	-3.0	-2.4	1.1
	(0.669)		*(0.673)*		*(0.438)*		***(0.854)***			

Note: MP= market price, standard deviations in parenthesis, mean difference t-test performed: *** p<0.01, ** p<0.05, * p<0.1. Point value differs from results presented in Table IV.17 since, here, mean departmental results are displayed and in the previous table the household mean was used.

Table IV.20: Correlation coefficients (unit values vs market prices)

	UV-1	UV-2	UV-3
Fon: Food to be consumed inside the house	**0.591**	**0.577**	**0.563**
Bread and Cereals	0.674	0.662	0.680
Meat and Fish	0.463	0.406	0.415
Milk, Eggs, Fats	0.624	0.637	0.686
Fruits	0.413	0.408	0.285
Vegetables	0.604	0.615	0.631
Other	-0.040	-0.022	-0.063

Similarly, when the correlation coefficients are calculated (Table IV.20), we can see the positive association between unit values and market prices in all cases except the ‘Other’ category, where the association is negative. This is not a surprising result given the differences in the composition of the ‘Other’ category in each sample. In general, correlations seem to be very similar across methodologies, although far from perfect. On average, given that the value of the aggregate food consumption correlation is around 0.58 and using the average of consumption groups, this value is around 0.45. These correlation coefficients are highly comparable to those of Niimi (2007), who conducted a similar exercise – in fact, she computes on average a coefficient of 0.48.

IV.6. Concluding remarks

In this chapter the main data issues have been discussed and some descriptive statistics on informality – both in the labour and the goods’ markets – presented and discussed. In terms of labour supply, for example, the informal sector proves to provide the main earnings opportunity, basically explained by the relatively high importance of the self-employment segment. Also, most of the stylized facts related to labour and the observable characteristics of workers are evident: informal workers tend to earn less than their formal counterparts; they are employed by small firms and have more flexible work schedules. At the same time, formal workers shows higher levels of human capital (in terms of education), enjoy more permanent jobs and more stable incomes. However, it is not completely true that the informal sector, as a whole, is a residual activity. In contrast, at least for the self-employed, it is found that, for the majority of workers, this is an option freely undertaken. Finally, when family interactions are included in the analysis, it would appear that household networking arrangements are in force in the Peruvian case: secondary earners tend to work in the sector that primary earners have already chosen.

In terms of the goods’ market, the information revised here shows a relatively high importance of informality in consumption in Peruvian households. At least 40% of household expenditure is allocated there, with formal consumption accounting for around 30%. Interestingly enough, the data also revealed that almost all Peruvian families located in urban zones participate in both markets to some extent. The main channels of engagement identified in the informal sector are city markets (Market 2) while supermarkets (Market 5) are the main channels for formal consumption. Finally, the data also reveal that informal and formal purchasing baskets are different. According to the purchasing groups constructed here, ‘Food to be consumed in the house’ and ‘Clothing and Personal Care’ are mainly bought in the informal sector and ‘Health goods and services’ and ‘Education and Culture’ products mainly in the formal market.

In terms of prices, the analysis has been restricted to food consumption, where it was possible to identify unit values. The methodology for price construction uses the district means of unit values of products bought in the different markets (formal or informal), aggregated using weighted geometrical means. The weights are defined at the district, departmental and national levels. The base estimation correspond to weights defined at the district level, following Niimi (2007), who demonstrates that these types of unit value help to reduce quality and measurement bias. The estimation reveals that informal goods are, in general, cheaper than formal ones, with average gaps of around 57%. Unit values constructed here also show a higher correspondence with average market prices at the sub-group level and overall food consumption. However, although the differences are small (relative to other studies), the unit values are still not market prices; this must be taken into consideration when interpretating the results.

V. ESTIMATION OF THE AGGREGATED MODEL

V.1. Introduction

In this chapter, the first empirical application of the theoretical model is presented. As discussed earlier, this application corresponds to the estimation of the aggregated model for informal consumption already presented in Chapter 3. The general specification followed is the Almost Ideal Demand System (AIDS) developed by Deaton and Muellbauer (1980a); however, given the data restrictions already explained in Chapter 4, prices in the current chapter are normalized to the unity. Therefore, the final specifications are more Working-Leser Engel curves. This estimation is closer to those encountered in the literature, where the conditional demand framework is exploited in order to identify the total expenditure elasticities and test the existence of linkages between working and purchasing from the informal (and formal) sectors.

Two empirical exercises will guide the discussion. The first is based on the assumption of exogenous expenditure and labour supply. Although simplistic, the flexibility of the econometric procedures helps to provide an in-depth exploration of the results and is considered to be an adequate initial reading of the information. Different exercises are proposed in this context. First, in terms of the specification of the demand equations and in order to provide a better exploration of the linkage hypothesis here, different types of labour opportunity are included (as explained in Chapter 4, five earning outcomes are included). Secondly, the results for different definitions of informality (three- and five-market approaches) will be explored. Thirdly, calculations are undertaken for different sub-population groups: individuals at different points of the income distribution and living in cities with different levels of development.

The second exercise explicitly deals with the potential endogeneity of total expenditure and hours of work. However, as long as the number of potentially problematic variables (at least 15) is too high, for identification reasons the dimension of the problem is restricted only a amsll number of endogenous covariates (only three). At the same time, the empirical approximation is less ambitious than the previous one in terms of the exercises presented. For example, the demand specification using five earning opportunities is replaced by two opportunities (overall formal and informal) and household expenditure uses a unique definition of informal consumption wn no results for different sub-groups of the population is presented. This is more or less the type of demand equation encountered in the literature.

The chapter is organized as follows. In Section 2, the specification is explained, the mathematical expression of the elasticities is presented and the econometric issues around the estimation procedures are discussed. In Section 3, the regression results for each of the empirical exercises are presented and commented on. In Section 4, the computed elasticities and the policy implications of the results are discussed. Finally, in Section 5, the main findings of the application are summarized in the form of concluding remarks.

V.2. Empirical specification and econometric issues

Using (III.32a)-(III.32b) and once prices are normalized to unity, the Working-Leser Engel curve employed for estimation purposes take the form:

$$w_j = \alpha_j^* + \beta_j \log(M) + \sum_r \tau_{jr}(h_r) + \psi_j(\lambda) + e_j \tag{V.1}$$

Following Banks *et al*. (1997), the major contribution of Muellbauer (1976), Deaton and Muellbauer (1980a) and Jorgenson *et al*. (1982) was to place this kind of specification within consumer theory. The most desirable property for its use is the satisfaction of the adding-up property. This means that, given the estimation for n goods (or markets), the parameters show $\sum_{j=1}^{n}\alpha_j = 1$, $\sum_{j=1}^{n}\beta_j = 0$, $\sum_{j=1}^{n}\tau_{jr} = 0$, $\sum_{j=1}^{n}\psi_{j\lambda} = 0$. This makes Working-Leser functions extremely useful in empirical work. The superiority of the specification is discussed in Leser (1963) and Deaton and Muellbauer (1980b).

As mentioned in Chapter 3, in (V.1) the parameters of main interest are β_j and τ_{jr}. The first, as in Reilly *et al*. (2006), will capture the effects of total income on the budget shares for informal and formal consumption and the second, as in Fortin *et al*. (2000), will implicitly take into account all the effects associated with h_r (network and non-separability or preferences effects). As mentioned, without knowledge of the cost function (a_j) and once prices are set to 1, individual effects cannot be distinguished, so τ_{jr} will be more generally referred to as the linkage effect.

All the terms are defined as in Chapter 3 but with two modifications, given the discussion in Chapter 4. First, j refers to the three different markets constructed: informal, semi-formal and formal. Second, in terms of hours of work (h_r), r now refers to five earning opportunities: formal workers in the wage-earning sector employed in formal firms (Hrs_ff), informal workers in the wage-earning sector employed in formal firms (Hrs_if), informal workers in the wage-earning sector employed in informal firms (Hrs_ii), the formal self-employed (Hrs_sf) and the informal self-employed (Hrs_si). In order to model all the labour supply opportunities that are available for the family, hours of work devoted by family workers and secondary job opportunities are also included, but particular interest is only placed on the first five. Hours of work are expressed in weekly terms. Participation dummies for each earning opportunity are also included. As discussed in Browning and Meghir (1991), the inclusion of these variables is relevant as long as they help to control for possible misspecification in the demand equations. These variables take into account the participation costs in each of the earnings opportunities and will pick up the non-linear effects of hours of work on consumption. This makes less clear the interpretation of the participation dummies, but facilitates the interpretation of the continuous part of labour supply.

The definition of income used is the natural logarithm of annual *per capita* expenditure considering only purchased consumption, as defined in Chapter 3, where the discussion revealed that this is the main consumption strategy of the household (more than 75% of total consumption) and the only one that can be disaggregated between formal and informal expenditure. Although the definition used helps the model to accommodate adding-up restrictions and the multiple-step budgeting process, it also implies a separation between purchasing and other consumption strategies (own consumption, private and public donations etc.).

Other variables included as part of the vector λ are the standard demand shifters (demographics) selected from similar studies (Fortin *et al*., 2000; Gardes and Starzec, 2002; Reilly *et al*., 2006): logarithm of family size, dummy variables for the presence of different members (spouse, son or daughter, others), proportion of men, characteristics of the head of the household (age and age-squared, gender and years of education) and a dummy for Lima (capital city). These variables are maintained across all estimations since they are considered the vector of characteristics used to model quality choices.

Additional regressors are included to control for possible exogenous variables both influencing decisions on purchasing and on working in the same sector. They are grouped according to the theoretical discussion presented in Chapters 2 and 3. The first group is location, introduced on the assumption that firms and markets are located depending on the observable characteristics of cities or neighbourhoods. The variables included are the logarithm of the district's population and its square, the population density of the district and three dummy variables: if the family house is not in a slum, if the family lives in a different departmental capital to Lima and if the family lives in a department on the borders of the country (except those in the jungle).The second group is related to research by Portes (2000), who argues in favour of cultural background or social ties as important factors in the emergence of informality in developing countries across some specific groups. In this application, certain dummies are included to identify these effects: if the head of the household or the spouse is indigenous, if the head of the household or the spouse is a migrant, and if at least one of the family members belongs to a group such as workers' unions, social clubs, associations of producers, etc. The third group is related to the enforcement capacity of the government in order to prevent informality. Several proxies have been included – for example, the number of police stations in the district per thousand inhabitants, the number of workers in local government per thousand residents, the logarithm of the local government's budget expressed in *per capita* terms and a dummy variable if the local government has a slums development plan.

As a result, the estimations are carried out with four different specifications: (1) the base model with only demographics, (2) the base model plus location variables, (3) the base model with cultural background, and (4) the base model with enforcement.[32] Given the linear nature of the Working-Leser Engel curve, the coefficients that accompany each of the regressors will have the standard interpretation of marginal effects on the budget shares for informal and formal markets. However, in order to derive the elasticities on total consumption, some manipulation is needed. In this application, directly applying the definition of elasticity to (V.1) it is possible to demonstrate that the resultant expressions for the variables of particular interest are:

$$\eta_j = \frac{\partial c_j}{\partial M}\frac{M}{c_j} = \frac{\beta_j}{\hat{w}_j} + 1 \tag{V.2}$$

$$\mu_{jr} = \frac{\partial c_j}{\partial h_r}\frac{h_r}{c_j} = \tau_{jr}\left(\frac{\hat{h}_r}{\hat{w}_j}\right) \tag{V.3}$$

$$\mu_{jr}^T = \frac{\partial c_{kj}^{UT}}{\partial h_r}\frac{h_r}{c_{kj}} = \mu_{jr} + \frac{\hat{h}_r \hat{s}_r}{\hat{M}}\eta_j \tag{V.4}$$

All expressions are completely identifiable once $\hat{w}_j, \hat{h}_r, \hat{M}, \hat{s}_r$ are set at sample means. In this case (V.2) is the expenditure elasticity, (V.3) is the elasticity with respect to changes in hours worked and (V.4) is the same expression but allowing movements in total income – i.e., this is the total effect for public policy simulations. However, note that since, empirically, hours of work are expressed in weekly terms and expenditure in annual terms, $\hat{h}_r \hat{s}_r$ in (V.4) is multiplied by 48.

The econometric application to be specified deals with several issues to be carried out during estimation. First, the demand system described earlier must be estimated following a simultaneous

[32] The definition of all variables is presented in Appendix 2, with their basic tabulations (sample means and standard deviations).

equation model if there are cross-equation correlations. In such cases, in order to avoid efficiency losses, the natural choice is the Seemingly Unrelated Regression model (SUR), but modified in order to cater for intra-equation heteroscedasticity.[33] This is done using Gould *et al*.'s (2006) maximum likelihood SUR procedure with robust standard errors, like those of White (1980) and Huber (1967). This method is referred to as the GLS–SUR.

Second, as discussed in Deaton (1986), the expenditure system uses the AIDS add-up perfectly, generating a singular variance–covariance matrix of the error term, where the usual generalized least-square estimator is not defined. The solution proposed by Deaton (1986) is to drop one of the categories in the system. In this application, Deaton's recommendation is followed, dropping the semi-formal category and focusing attention on formal and informal consumption.

Third, there are several potential endogenous variables in the application that could lead to inconsistent estimates of the SUR–GLS models. The most evident that are of particular interest for the research are expenditure and labour supply. The main reason for suspecting the endogeneity of M is that it is the sum of individual expenditure and therefore, if individual consumption is assumed to be endogenous in the model, the sum will be simultaneously determined. The bias introduced by ignoring simultaneity in the context of the demand analysis has been explored in detail by Deaton (1986). According to him, the bias will be small if the equations fit well and zero under the rational random behaviour model (Theil, 1971, 1976; Deaton, 1975). Although these arguments probably justify most of the applications encountered in the literature, under the assumption of exogenous total expenditure (see Deaton, 1997 for some examples), in the present application it is preferable to at least deal approximately with simultaneity issues, given the possible measurement error in the dependant variable, as already mentioned. Also, this more conservative way of proceding is undertaken in order to control for possible sources of scepticism in the results. As noted in Hoddinott and Haddad (1995), large purchases will affect both sides of the equation in cross-sectional frameworks and aggregation across goods (or markets) could generate categories with large shares of total expenditure. This problem may be more pervasive when a small number of categories are included as dependent variables, as here.

The main reason for suspecting endogeneity on h_r is that, in the standard consumption model, leisure and consumption decisions are undertaken at the same time and at the same stage in the

[33] This was properly tested using the Breusch–Pagan/Cook–Weisberg heteroscedasticity test. Chi-squared values were between 15 and 50 for the estimations done (higher than the critical 3.84). Cross-equations were also tested using Breusch–Pagan. Chi-squared in this case were between 1,600 and 1,800 (above the critical 3.84).

decision process. Therefore, as long as leisure is just an additional ‘good’ in the demand setting, there is a problem of reverse causality of hours influencing consumption and consumption influencing labour decisions. However, in the context of a conditional demand framework, as developed in the previous chapter, this is not necessarily the case. In equation (III.2), leisure is a predetermined good (i.e., quantity exogenous for a consumption decision among goods or markets). This issue is discussed in detail in Pollak (1969, 1971), Deaton (1981) and Deaton and Muellbauer (1980b), where fixed commitments, institutionally determined hours of work or long-term contractual arrangements are used to justify labour decisions, as pre-allocated at specific levels when consumption decisions are taken. Arguments like this justify the assumption of exogenous labour supply in demand analysis (as in Hayashi, 1995; Reilly *et al*., 2006), but again seems to be a risky option given that an inflexible labour supply may be reasonable for some, but not necessarily all, individuals in the sample. For that reason, it is preferable to deal explicitly with potential endogeneity that arises from the inclusion of a choice variable on the right-hand side of the regression equation.

Therefore, there are at least 15 potential endogenous variables: expenditure, 7 variables for hours of work and 7 participation dummies. However, leading with a high number of endogenous variables was not a practical choice. For that reason, in order to deal with endogeneity, a reduction in the dimension of the problem and some modifications to the original model are proposed. In the first instance, variables that are not of particular interest in the application – like hours of work as family workers and moonlighting – are left aside, as are their participation dummies. To formalize this decision, the original estimation sample is restricted to only those households with no family workers and to members having only one job a. In the second case, the five earning opportunities of interest are grouped into formal and informal hours of work (with their respective participation dummies) following the extended definition of informality already discussed in Chapter 4. In third place, as long as they are not of particular interest, participation dummies are assumed to be truly predetermined in the short run (between hours and participation, to place the assumption on the latter is probably less controversial) so the potential endogeneity of these variables is not treated explicitly. The validity of this last assumption is tested using the C-statistic for exogeneity of included instruments (explained in Hayashi, 2000; Baum *et al*., 2003).

As a result, there are three potentially endogenous variables: total expenditure, formal hours of work and informal hours of work. The procedure proposed handles each problematic variable separately. Therefore, two exercises are proposed: an IV estimation for total expenditure under the assumption of exogenous labour supply and an IV estimation for labour supply under the assumption of exogenous total expenditure.[34] The exclusion restrictions imposed in the model are justified by both theoretical and statistical reasons. For example, to identify total expenditure, inter-temporal separability is assumed (as in Browning and Meghir, 1991) so proxies of permanent income are included as long as they will affect current expenditure, but not preferences. The variables included were the value of total assets in the household and house characteristics (also used in Böhme and Thiele, 2012b in their composed index). Fortin *et al*. (2000) also includes property rents and extraordinary incomes (like lottery winnings, inheritances).

To identify hours of work, salaries in both sectors are included (as suggested by Browning and Meghir, 1991), using the district means and following an imputation procedure similar to unit values (see Chapter 4). Also, as in Fortin *et al*. (2000), a variable that identifies rationed workers (working less time than preferred) is included as long as it will influence labour allocation in informal sectors. Finally, in order to obtain enough over-identification restrictions, a set of exogenous shocks experienced by the household in the last year were also included. The shocks identified in the sample were: a household member losing his or her job, the family business going bankrupt, a family member dying, a member being critically ill or having an accident, the head of the household abandoning the family, the family experiencing a fire in the house, the family being burgled or suffering from some other crime against it, the family experiencing a natural disaster and other shocks not classified previously.

However, it should be recognized that some of the exclusion restrictions may be controversial in the empirical context of this application. For that reason, rigorous statistical procedures were followed in order justify the validity of the econometric exercises presented. Therefore, the orthogonality for excluded instruments was adequately tested using the Hansen–Sargan J statistic (see Hayashi, 2000) and, for relevance, using both individual F-tests (Cameron and Trivedi, 2005) and multiple-endogenous-regressors tests proposed by Kleibergen and Paap (2006) with critical values derived from Stock and Yogo (2005). When a valid set of instruments is found, then exogeneity for potentially endogenous regressors is tested using the C-test (as explained in Hayashi, 2000; Baum *et al*., 2003).

V.3. The regression results

<u>First empirical exercise: estimation under the assumption of exogeneity</u>

[34] Some exercises were developed to deal with 5 and 3 endogenous variables. In these cases, although orthogonality was achieved under most of the specifications used, the set of instruments that maximizes relevance tests defines an extremely weak instrument context so the decision was to exclude these exercises. These results are not presented in order to conserve space.

In Table V.1, the results for the SUR–GLS estimator are displayed for the variables of main interest under the four specifications mentioned previously (full regression results for these and other estimates in the chapter are in Appendix 3). The total sample size used once missing values for control variables were dropped is 8,007 observations. Using the simplest specification (first column), the results show, in general, adequate goodness-of-fit measures with significant Wald statistics for both overall and group variable estimates. However, considering the value of the R2 reported, it follows that there is still some noise in the estimation. This seems to be particularly problematic in the case of informal consumption. This conclusion is similar to that of Reilly *et al*. (2006) and is used to justify the inclusion of additional covariates in the models (Columns 2 to 4). The new regressions again show adequate goodness-of-fit measures with significant Wald statistics for both overall and group variable estimates. Also, the goodness-of-fit measures reported has increased and the information criteria used (AIC and BIC) help to conclude that regressions perform better. However, as in the previous case, R2 values are still low. Again, the problem seems to be more pervasive in the informal specification so, although the global adjustment of the regression has improved, a good deal of work is still needed on the determinants of informal consumption.

The point estimates of the variables tell an interesting story about consumption allocation and, in most cases, with intuitive appeal according to the theoretical model. Also, it is worth noting that signs for the core variables of the application do not change across the specifications used and are very similar in the estimated magnitudes displayed. In the case of the logarithm of family expenditure, it is well determined and shows a negative correlation with informal budget shares and a positive correlation with its formal counterpart. Therefore it is possible to argue that income effects are differentiated between formal and informal channels of commercialization. The estimates suggest that we can expect families to actually change the composition of their consumption baskets given changes in their income: more informal as income decreases (possibly because of protection strategies) and more formal as income increases (possibly because of their higher preference of better-quality goods). This is an expected result. For example, as stated by Cermeño (1987) and Pisani (2013a), from an economic point of view, poor people will be closer to informal markets and rich people closer to their formal counterparts, given the higher prices to be paid in the latter. At the same time, according to Böhme and Thiele (2012b), quality differentials could be behind a result like this; they suggest that families will substitute informal consumption for more formal options as they become richer in order to increase the overall quality level of their consumption basket. Similarly, from a sociological perspective, Portes (2000) helps to explain this result if informality is seen as the way of life encountered by those marginalized from the legal or regular economy. It could be argued that these excluded groups are the poor revealing the contraints of the informal sector from the demand side (see Böhme and Thiele, 2012b). Arellano and Burgos (2010) extend this last argument, arguing that scarce formal supply to poor neighbourhoods will determine a positive correlation between informal consumption and poverty, even if poor households actually prefer to purchase from formal options.

The value of the expenditure coefficient shows that, holding everything else constant, a 10% increase in annual income (or expenditure) will lead to a reduction of around 0.3 of one percentage point in the budget share of informal expenditures and an increase of about 0.8 of one percentage point in the budget share of formal ones. However, given Deaton and Muellbauer's (1980a) discussion, these results are not revealing the inferiority of informal market purchases. As will be seen later, the negative effects of income on budget shares are not related to negative elasticity estimates. In fact, estimated elasticities for informal markets are lower than the unity, but still positive. These results suggest that informal consumption can be classified as normal necessities and formal ones as normal luxuries for the Peruvian economy. In consequence, inferiority is rejected on average and *ceteris paribus*.

The results of 'hours of work' are in line with the linkage hypothesis discussed earlier: working more hours at informal earning opportunities is correlated with higher informal consumption shares and lower shares in its formal counterpart. The converse is also applicable: working more hours in formal job opportunities tends to bias household consumption in favour of formal shares. As explained in Chapter 3, the intuition behind the results is that previous experience in a sector (as a worker) helps the consumer to reduce transactional costs or reveals a preference of the worker to stay there as a consumer. Although neither effect can be disentangled, the positive (negative) correlations encountered and tested between formal and informal hours of work with formal (informal) and informal (formal) consumption, respectively, are considered enough evidence to conclude in favour of the linkage hypothesis. Considering the different specifications used, this result seems to be robust to the inclusion of the additional variables possibly correlated with working and purchasing activities.

However, in contrast to previous applications (see, for example, Fortin *et al*., 2000; Gardes and Starzec, 2002; Reilly *et al*., 2006), an important result outlined in this chapter is that not all informal or formal earning opportunities have the same effect on consumption. The hypothesis testing presented at the bottom of Table V.I shows that the five-labour specification is superior to other alternatives. In short, the differences encountered between informal and formal options and the self-employment or wage-earning options are statistically significant and robust to changes in the controls included in the model. Therefore, it is possible to conclude that the various types of labour supply affect consumption differently.

Table V.1: Results for the SUR–GLS regression (different specifications)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expenditure)	-0.02679***	0.08424***	-0.03744***	0.08121***	-0.02526***	0.08400***	-0.02597***	0.08291***
	(0.00413)	*(0.00384)*	*(0.00425)*	*(0.00386)*	*(0.00415)*	*(0.00389)*	*(0.00412)*	*(0.00382)*
Hrs_ff	-0.00027***	0.00022***	-0.00027***	0.00022***	-0.00027***	0.00022**	-0.00028***	0.00022***
	(0.00009)	*(0.00008)*	*(0.00009)*	*(0.00008)*	*(0.00009)*	*(0.00008)*	*(0.00009)*	*(0.00008)*
Hrs_if	0.00017	0.00011	0.00013	0.00013	0.00017	0.00010	0.00016	0.00012
	(0.00019)	*(0.00015)*	*(0.00019)*	*(0.00015)*	*(0.00020)*	*(0.00015)*	*(0.00019)*	*(0.00015)*
Hrs_ii	-0.00012	-0.00032***	-0.00013	-0.00028***	-0.00013	-0.00032***	-0.00012	-0.00032***
	(0.00008)	*(0.00006)*	*(0.00008)*	*(0.00006)*	*(0.00008)*	*(0.00006)*	*(0.00008)*	*(0.00006)*
Hrs_sf	0.00006	0.00054***	-0.00007	0.00056***	0.00002	0.00054***	0.00009	0.00053***
	(0.00017)	*(0.00018)*	*(0.00017)*	*(0.00018)*	*(0.00017)*	*(0.00018)*	*(0.00017)*	*(0.00018)*
Hrs_si	0.00029***	0.00006	0.00023***	0.00007	0.00027***	0.00006	0.00028***	0.00006
	(0.00007)	*(0.00006)*	*(0.00007)*	*(0.00006)*	*(0.00007)*	*(0.00006)*	*(0.00007)*	*(0.00006)*
D(Hrs_ff)	0.01063	-0.00110	0.01395**	-0.00065	0.00893	-0.00057	0.01174*	-0.00249
	(0.00684)	*(0.00629)*	*(0.00672)*	*(0.00622)*	*(0.00683)*	*(0.00630)*	*(0.00683)*	*(0.00626)*
D(Hrs_if)	-0.01265	-0.01452	-0.01211	-0.01483	-0.01214	-0.01433	-0.01315	-0.01468
	(0.01158)	*(0.00930)*	*(0.01143)*	*(0.00923)*	*(0.01167)*	*(0.00931)*	*(0.01159)*	*(0.00931)*
D(Hrs_ii)	0.00449	-0.00970*	0.00401	-0.01051**	0.00379	-0.00987*	0.00419	-0.00914*
	(0.00650)	*(0.00520)*	*(0.00643)*	*(0.00516)*	*(0.00650)*	*(0.00521)*	*(0.00647)*	*(0.00519)*
D(Hrs_sf)	-0.02336**	-0.00404	-0.00658	-0.00657	-0.02294**	-0.00378	-0.02578**	-0.00320
	(0.01138)	*(0.01084)*	*(0.01096)*	*(0.01074)*	*(0.01140)*	*(0.01085)*	*(0.01137)*	*(0.01080)*
D(Hrs_si)	0.01158**	-0.00846*	0.01020*	-0.00773*	0.01163**	-0.00858*	0.01066*	-0.00733
	(0.00563)	*(0.00472)*	*(0.00555)*	*(0.00469)*	*(0.00561)*	*(0.00471)*	*(0.00561)*	*(0.00471)*
Observations	8,007		8,007		8,007		8,007	
Goodness of fit								
R-squared	0.0947	0.2822	0.1265	0.2918	0.1002	0.2824	0.0996	0.2871
Pseudo LogL	7926.212		8154.336		7955.881		7973.908	
AIC	-15742.42		-16174.67		-15789.76		-15821.82	
BIC	-15358.08		-15706.47		-15363.49		-15381.57	
Wald (overall) 1/	3640.04***		4255.44***		3698.89***		3748.43***	
Chi2 (participation) 2/	24.57***		22.08**		23.48***		25.01***	
Chi2 (hours) 3/	103.15***		88.44***		100.59***		104.56***	
Chi2 (demographics)	521.28***		425.04***		513.04***		517.30***	
Chi2 (location)			846.28***					
Chi2 (background)					57.31***			
Chi2 (enforcement)							69.14***	
Income effects								
Chi2 (informal vs. formal) 4/	266.23***		294.98***		253.09***		258.60***	
Hours effects 5/								
Chi2 (informal vs formal) 4/	111.85***		78.36***		112.29***		107.60***	
Chi2 (pooled, joint hours) 6/	301.97***		263.91***		303.73***		297.39***	
Intra-equation	104.60***	161.27***	68.57***	149.00***	105.57***	162.41***	103.07***	156.97***
Chi2 (pooled, formal and informal) 7/	192.77***		167.59***		192.57***		194.33***	
Intra-equation	60.39***	56.44***	45.16***	53.67***	58.84***	56.45***	61.06***	58.31***
Chi2 (pooled, self and wage) 8/	206.01***		183.00***		208.36***		199.43***	
Intra-equation	64.19***	145.50***	36.83***	134.03***	66.03***	146.63***	62.48***	140.04***

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1; 1/ All variables included; 2/ Includes D(Hrs_ff),D(Hrs_if), D(Hrs_ii), D(Hrs_sf), D(Hrs_si); 3/ Includes Hrs_ff, Hrs_if, Hrs_ii, Hrs_sf, Hrs_si ; 4/ Test of parameter differences between the formal and informal equation; 5/ Includes participation dummies; 6/ Tests the specification of five labour options against a unique hours supply vector that sums the five of them ; 7/ Tests the specification of five labour options against two labour options (formal and informal). The definition of informality is the extended version used in Chapter 4; 8/ Tests the specification of five labour options against two labour options (self-employment and wage-earners).

In terms of informal consumption, the strongest evidence for a positive linkage is found for the informal self-employed (Hrs_si). This is highly intuitive considering that most of the arguments around the network or preference hypotheses seem to be more applicable to those informal consumers who are also informal producers or sellers. Informational advantages (in terms of quality and quantity of goods supplied in the informal sector) could be higher for them. Also, they will be more interested in remaining hidden. Similarly, if, as stated by Yamada (1996) and discussed in Chapter 4, this sector is a freely chosen one, there will be less of a stigma in buying informal goods, so preferences to remain as an informal consumer could be also strong. Interestingly enough, the positive linkage between informal consumption and working in the informal self-employment sector is not necessarily reflected in a negative linkage with formal consumption (the coefficients are not statistically different to zero), suggesting that these consumers adjust their semi-formal consumption.

The negative linkage with formal consumption is found for informal workers at informal firms (Hrs_ii). In this case positive linkages are probably less applicable (correlations with informal consumption are not detected) all the while that their condition as wage-earners reduces information flows to them. However, at the same time, they will show greater difficulties in buying in the formal sector. The result is understandable considering that incomes of these workers are the most inestable among others options (see Chapter 4). Also, as described by Yamada (1996), this part of the informal sector in Peru constitutes the segmented proportion with marginal jobs. Then it would also be natural to conclude that this sector would mainly absorb marginalized individuals (Portes, 2000) or that it constitutes a segment of developing societies where discriminatory practices act as additional barriers to modern transactional channels (Arellano and Burgos, 2010).

By contrast, formal wage-earners (Hrs_ff) reveal a higher aversion to purchasing from informal outlets possibly because of a fear of stigmatization or of moral concerns, a result not detected for other formal labour options like the self-employed (Hrs_sf). This is a reasonable result since it can be argued that these concerns are less probable for the formal self-employed, given that this sector is closer to informality. Anecdotal evidence suggests that most of the formal self-employed workers have become formal after a period in informality and that there is great mobility across these two sectors. Although there are no studies that properly prove this hypothesis in Peru, some descriptive results encountered in Diaz (2014) suggest similar behaviour. However, positive linkages are found for both formal labour options. When tested, the marginal effects between them are not significantly different[35], but anecdotal evidence suggests that the reasons for the linkage could, in fact, be different. For example, beyond any informational advantage or preference effect affecting both formal labour options, small-size formal producers or sellers will have additional reasons to bias household consumption through formal markets: they could use family expenditure (in the formal sector) to reduce the amount of taxes paid (via tax credit mechanisms), an option not available to other formal workers.

In terms of the participation dummies, the results are less clear, but most are still in line with the linkage hypothesis. The two robust results detected across the specifications used are the positive effect of engaging in the informal self-employment sector – D(Hrs_si) – over informal consumption and the negative effect of engaging in the informal wage-earning sector – D(Hrs_ii) – over formal consumption. Similarly, negative effects of D(Hrs_sf) and D(Hrs_si) over informal and formal consumption, respectively, are encountered in three out of four specifications. However, a positive effect of D(Hrs_ff) on informal consumption is also found. This is a counter-intuitive result possibly generated by non-linearities at the zero-hour level, and which is difficult to explain with the theory to hand. In any case, this result is found in only two out of four specifications so does not necessarily obscure previous findings.

The other variables (as presented in Appendix 3) show the expected signs. For example, in the case of demographic characteristics, the presence of spouses and the higher proportion of women tend to bias the budget allocation in favour of informal markets. This is an interesting finding, especially if spouses and women are mainly responsible for purchasing activities and have developed trust in the sellers. Then, if informal consumption needs investments in time in order to search for quality, this family structure (with member with free time) will reasonably predict better access to informal markets. Anecdotal evidence suggests that, in Peruvian families, purchasing activities (mainly food) tend to be done by non-working women[36], so they sometimes enjoy a better knowledge of the available markets and their characteristics and, in most cases, have built up relationships of trust with some traditional (informal) sellers (in most cases also women).

In the case of the education of the head of household, more-educated households purchase a higher proportion of their goods from formal markets and a lower proportion from informal ones. This is possibly related to the more sophisticated tastes associated with education or to consumers

[35] Chi-squared Hrs_ff vs. Hrs_sf: (1) 2.64; (2) 3.00*; (3) 2.27; (4) 2.57

[36] It is preferable to interpret these results as the effect of non-working women in the household in order to reconcile this evidence woth that presented in Chapter 8. There, proper bargaining models are estimated between husbands and wifes. Using that specification results reveal that increasing control over money of wifes in fact reduces informal consumption.

who are better prepared for complex transactions in the formal sector. By contrast, the effect of the age of the household head seems to have a non-linear effect on informal consumption, increasing it for young heads and decreasing it for older ones. The inflexion point identified for informal consumption is around 47 years old. In the case of formal consumption, a non-linear relation is also identified, but the inflexion point is found to be around 97 years, possibly suggesting that the relation could be defined as linear in this case.

The additional variables also reveal an interesting story. For example, in terms of the proxies of location, the size of the district has an inverted-U effect over the informal budget share. A result like this is consistent with the Harris–Todaro approach, which describes an increase in informality in the early stages of urbanization in developing countries until the point where cities reach maturity. Also, lower formal shares are associated with living in slums and higher informal shares with life in the borders of the country. This last result is in line with the black commerce (contraband) that occurs in these places, where police control of smuggled goods is practically non-existant (see SUNAT, 2011). Interestingly enough, the dummy that identifies Lima is possibly capturing the highest heterogeneity in the channels of commercialization in the main city, since both formal and informal markets tend to be larger compared to the semi-formal option. However, the point estimates could be interpreted as showing that, even in this most developed city, informality dominates.

In terms of cultural background, indigenous or migrant families generally show higher proportions of informal expenditure in their purchasing baskets – an intuitive result. Also, in terms of proxies for enforcement capacity, it seems adequate to argue that higher enforcement efforts by local government result in lower informal and higher formal shares. For example, more police stations per thousand individuals increases formal consumption shares and reduces its informal counterparts (at least in one of the two specifications). Similar results are driven by the number of municipal workers. In the case of local budgets, no significant results are encountered. This is possibly explained because only a small proportion of these budgets is used for control activities. Finally, in the case of the particular public policies adopted (slum development plans), the effects are not clear since they appear to have a negative and statistically similar effect on formal and informal shares.

In order to explore with more detail the core results, the expenditure and hours effects are estimated for different points in the expenditure distribution. For this purpose, five regression models are estimated for the sub-populations in each quintile of the *per-capita* expenditure distribution. In order to perform cross-quintile parameter testing, between-model covariances are calculated using Weesie's (1999) procedure. If previous results are sensitive to the absolute value of resources enjoyed by the family (and, thus, their monetary poverty), then this exercise should reveal it. The results are presented in Table V.2, where the estimations are done only for specification (1) and the main variables of interest are displayed. Other results show similar patterns and are omitted for the sake of brevity. It is interesting to note that marginal effects are sensitive to the position of the household on income distribution.

Table V.2: Total expenditure and hours effects for different points in the income distribution

	Quintile 1		Quintile 2		Quintile 3		Quintile 4		Quintile 5	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expenditure)	0.04337***	0.00736	-0.00206	0.08551***	0.00728	0.15719***	-0.07818**	0.10435***	-0.10473***	0.10236***
	(0.01359)	*(0.01348)*	*(0.04019)*	*(0.03100)*	*(0.04296)*	*(0.03855)*	*(0.03413)*	*(0.03196)*	*(0.00956)*	*(0.01196)*
Hrs_ff	-0.00031	0.00014	-0.00046*	0.00051*	0.00008	-0.00010	-0.00028*	0.00012	-0.00015	0.00012
	(0.00036)	*(0.00028)*	*(0.00027)*	*(0.00027)*	*(0.00019)*	*(0.00017)*	*(0.00017)*	*(0.00016)*	*(0.00014)*	*(0.00015)*
Hrs_if	0.00019	0.00016	0.00039	0.00043	-0.00011	0.00008	0.00027	-0.00026	-0.00003	0.00050
	(0.00040)	*(0.00026)*	*(0.00051)*	*(0.00043)*	*(0.00033)*	*(0.00023)*	*(0.00047)*	*(0.00036)*	*(0.00043)*	*(0.00036)*
Hrs_ii	-0.00029	-0.00011	-0.00035**	-0.00012	-0.00015	-0.00022*	-0.00005	-0.00059***	0.00016	-0.00089***
	(0.00018)	*(0.00011)*	*(0.00017)*	*(0.00011)*	*(0.00016)*	*(0.00013)*	*(0.00018)*	*(0.00017)*	*(0.00028)*	*(0.00032)*
Hrs_sf	0.00013	0.00158***	0.00083*	0.00025	-0.00030	0.00055	-0.00043	0.00099***	-0.00004	-0.00005
	(0.00063)	*(0.00054)*	*(0.00049)*	*(0.00042)*	*(0.00033)*	*(0.00034)*	*(0.00034)*	*(0.00036)*	*(0.00027)*	*(0.00038)*
Hrs_si	0.00024*	0.00032***	-0.00001	0.00033***	0.00029**	-0.00003	0.00033*	-0.00025	0.00063***	-0.00027
	(0.00014)	*(0.00011)*	*(0.00014)*	*(0.00011)*	*(0.00013)*	*(0.00011)*	*(0.00017)*	*(0.00016)*	*(0.00023)*	*(0.00022)*
Observations	1,602		1,601		1,602		1,601		1,601	
Goodness of fit										
Pseudo LogL	1423.676		1614.402		1811.801		1772.687		1871.271	

Note: Robust standard errors in parentheses with the adjustment for between vce using Weesie (1999), *** p<0.01, ** p<0.05, * p<0.1.

In the case of total expenditure, estimates for informal consumption show that the effect tends to turn from positive to negative as long as we move from the left (poorest households) to the right (richest ones). In fact, there are three identifiable ranges. The first one is defined for poor households (in Q1) where increasing total expenditure actually increases informal budget shares. The second is defined for middle-income families (in Q2 and Q3) where informal consumption shares seem to be practically invariant to total expenditure. The third range, for the richest individuals of the sample (Q4 and Q5), reveals negative marginal effects. Although, point values for Q5 are slightly higher (in absolute terms) than those for Q4, when formally tested they are not statistically different[37] at conventional values. The estimates for formal consumption seem to be more stable across quintiles. Only for those households at the bottom of the distribution (Q1) is the marginal effect statistically equal to zero. For all other consumers, the expected positive effect of expenditure on formal consumption shares encountered previously is replicated. Although the point-value estimates seem to define an inverted-U relationship (with higher results in the

[37] Chi-squared L(expenditure) Q4 vs. Q5: 0.56.

middle), when formally tested, the computed coefficients across Q2 and Q5 are not statistically different[38] at conventional values.

This evidence is probably associated with the greater preferences of poor people for informal markets in contrast to those located at the top of the distribution range. As explained in Wan (1996) and Böhme and Thiele (2012b), when demand models are estimated for households at different points in the income distribution, some necessities are found to be luxuries for extremely poor households as the main concern of these families is subsistence consumption and therefore quantity instead of quality. Similarly, following the arguments of Arellano and Burgos (2010), we can see that, if the poor are typically limited in their participation in the formal economy (because, for example, of scarce formal supply in their neighbourhood or of liquidity constraints), it is more or less obvious that, for some in the lower ranges of the income distribution, any additional "*Nuevo Sol*" in the pockets of consumers will tend to increase more than proportionaly their allocation of resources to informal markets.

Differences are also found in the case of hours of work. The results are very informative in terms of the validity of the linkage hypothesis in each quintile of the distribution. In fact, validity seems to be clearer for the upper range, defined as between Q3 to Q5, than for the lower, defined as between Q1 and Q2. In the former, most of the effects previously identified are replicated. For example, in the case of Q3, the detectable effect of informal consumption is its positive correlation with informal self-employment and the detectable effect in terms of the formal equation is the negative correlation with the informal wage-earning sector. For Q5, these effects are also found; for Q4, both effects are accompanied by the negative effect of Hrs_ff on informal shares and the positive correlation between Hrs_sf and formal consumption. In the latter, some counter-intuitive correlations displayed (mainly in Q2) at least cast doubts on the validity of the hypothesis.

The data may support conflation of the first two quintiles, so the results presented here do not deviate from our hypothesis. However, at the lower end of the distribution, weaker evidence in favour of the linkage hypothesis could also be an expected result. If the effect of labour supply is related to some kind of strategic behaviour by consumers, these attitudes are probably better developed by individuals with some minimum living standards. For the rest (the poorest elements of the population), as long as most of their decisions are dominated by subsistence behaviour, less space for strategies is allowed, so income effects will prevail and may well be enough to explain their consumption allocation. This argument is similar to those of Williams (2002), Williams and

[38] Chi-squared L(expenditure): Q2 vs. Q3: 2.10; Q2 vs. Q4: 0.18; Q2 vs. Q5: 0.26; Q3 vs. Q4: 1.11; Q3 vs. Q5: 1.85; Q4 vs. Q5: 0.00.

Paddock (2003) and Williams and Windebank (2005) and Williams (2008) who, using attitudinal data in both English and European cities, found that economic constraints are the main explanation for the participation of the poor in informal markets, with any other reason being generally less important.

Beyond the total resources enjoyed by the family, it is also interesting to demonstrate the validity of the hypothesis considering cities at different stages of development, as long as the formal and informal markets can emerge in different ways. One way to approximate this issue is by using the classification provided in Macroconsult (2010). According to this study, the urban structure of the country can be classified in terms of metropolitan areas (the biggest ones, better developed in economic terms and can be considered as commercial anchors for the whole country), major cities (cities with secondary importance with influence at the macro-regional level with primary articulation to the metropolitan area), intermediate cities (cities of third importance with influence at the regional level and with primary articulation to the major cities), minor cities (cities at the bottom of the urban structure with commercial influence at the local level and with primary articulation to intermediate cities) and of surroundings (small cities with influence only on the rural area around them and primary articulation with minor cities).

Table V.3: Total expenditure and hours effects for different urban structures

	Metropolitan areas		Intermediate cities		Minor cities	
	Informal	Formal	Informal	Formal	Informal	Formal
L(expenditure)	-0.07134***	0.11111***	-0.02711***	0.08614***	-0.00290	0.06479***
	(0.00826)	*(0.00677)*	*(0.00706)*	*(0.00575)*	*(0.00625)*	*(0.00646)*
Hrs_ff	-0.00029*	0.00021	-0.00013	0.00010	-0.00038**	0.00034**
	(0.00015)	*(0.00015)*	*(0.00015)*	*(0.00014)*	*(0.00015)*	*(0.00016)*
Hrs_if	0.00006	0.00025	-0.00005	0.00013	0.00055	-0.00018
	(0.00027)	*(0.00022)*	*(0.00034)*	*(0.00030)*	*(0.00038)*	*(0.00025)*
Hrs_ii	-0.00017	-0.00039***	-0.00015	-0.00022*	-0.00023	-0.00027**
	(0.00013)	*(0.00010)*	*(0.00015)*	*(0.00012)*	*(0.00016)*	*(0.00011)*
Hrs_sf	-0.00035	0.00063	-0.00007	0.00032	0.00000	0.00076***
	(0.00041)	*(0.00043)*	*(0.00025)*	*(0.00028)*	*(0.00028)*	*(0.00027)*
Hrs_si	0.00035***	-0.00008	0.00024**	0.00012	0.00022*	0.00013
	(0.00013)	*(0.00011)*	*(0.00011)*	*(0.00010)*	*(0.00012)*	*(0.00009)*
Observations	2,140		2,718		3,149	
Goodness of fit						
Pseudo LogL	2376.565		2880.827		2875.03	

Note: Robust standard errors in parentheses with the adjustment for between vce using Weesie (1999), *** p<0.01, ** p<0.05, * p<0.1.

To implement this classification in the model, the categories were aggregated between metropolitan areas (as defined in the study), intermediate cities (including major and intermediate cities) and minor cities (including minor cities and their surroundings). The classification could be thought of as areas at different stages of the urbanization process, with the first group representing the biggest and most modern cities and the last group the smallest and least-modern ones.[39] If the results differ according to the urban structure, the following exercise should reveal this. The results are presented in Table V.3 only for specification (1) to conserve space.

The first result to note refers to the expenditure effect, which is greater, in absolute terms, for both formal and informal shares, as long as the cities are larger. In addition, the gap between coefficients in metropolitan areas is almost double those in intermediate cities and almost three times those in minor cities. Similar differences are found when intermediate and minor cities are compared. The differences are statistically significant[40] at conventional levels, so it is possible to verify that the effects of income on the re-composition of the purchasing basket will be stronger for those families living in the biggest urban areas of the country. The explanation for is probably the better access to and size of the formal sector in bigger cities, the greater presence of wealthier families, the wider quality gap between formal and informal goods or the major preference for quality in areas in latter stages of the urbanization process.

When hours of work are analysed, it is worth noting that the linkage hypothesis is equally valid for the different urban areas considered. In fact, the robust results encountered are, again, the positive linkage of the informal self-employed with informal consumption and the negative linkage of the informal wage-earning sector with its formal counterparts. These results are statistically similar across the different types of city.[41] Other interesting results are the negative effect of Hrs_ff on informal consumption in metropolitan areas and minor cities and the positive effect of both Hrs_ff and Hrs_sf on formal consumption in minor cities; again all are in line with the linkage hypothesis. In general, the first results suggest that moral concerns or the aversion of formal wage-earners to purchasing from informal markets are equally applicable to bigger and smaller cities, but not to intermediate ones. The second result suggests that the positive linkages of formal workers to formal consumption are better determined in smaller cities with the lowest level of development.

[39] The identification made is at provincial and district levels. The metropolitan areas identified were Lima, Arequipa, Trujillo and Chiclayo. The major cities were Piura, Chimbote, Iquitos, Cusco and Huancayo. The intermediate cities were Tacna, Ica, Sullana, Chincha Alta, Huaraz, Pucallpa, Cajamarca, Ayacucho, Huanuco, Tarapoto, Juliaca and Puno and the minor cities Tumbes, Talara, Pisco, Moquegua, Cerro de Pasco, Puerto Maldonado, Tingo María, Moyobamba, Huancavelica, Aguaytía, Chachapoyas, Tocache and Abancay. The surroundings were all other cities. They were identified in the sample using the geographic codification. To confirm Macroconsult (2010) arguments around development level of the areas considered, the Human Development Index (HDI) for 2007 was calculated for the areas under consideration (as long as this information was available at the provincial and district levels). The results are (where bigger index numbers reveal higher development levels): i) metropolitan areas: 0.677; ii) intermediate cities: 0.622; iii) minor cities: 0.582.

[40] Chi-squared informal equation: metropolitan vs major (16.56***), metropolitan vs minor (43.63***), major vs minor (6.59**). Chi-squared formal equation: metropolitan vs major (7.91***), metropolitan vs minor (24.49***), major vs minor (6.09**).

[41] Chi-squared Hrs_si: metropolitan vs major (0.38), metropolitan vs minor (0.56), major vs minor (0.03). Chi-squared Hrs_ii: metropolitan vs major (1.22), metropolitan vs minor (0.74), major vs minor (0.09).

In Table V.4 the income effect is calculated for each of the working groups identified in the sample considering specification (1). However, as long as the sample used in this application is at the household level and hours of work are the sum of the labour supply of the family, the working groups are not mutually exclusive. In this context, interaction terms in the original sample will be difficult to interpret. To overcome this problem, the sample was restricted in order to take into account only mutually exclusive groups. This means that, for this exercise, households with members working in different sectors were excluded and only households where labour is allocated in one of the five earning opportunities were included. The size of this new sample is 4,216 observations (almost 50% of the original sample). Previous estimates were replicated in this new context, given the interactions between the participation dummies and the logarithm of expenditure.

Table V.4: Total expenditure marginal effects by labour options

	D(Hrs_ff)	D(Hrs_if)	D(Hrs_ii)	D(Hrs_sf)	D(Hrs_si)
Informal	-0.07184***	-0.02302	-0.00597	-0.03463***	-0.02026***
	(0.00776)	*(0.02045)*	*(0.01184)*	*(0.01154)*	*(0.00802)*
Formal	0.11446***	0.12219***	0.04642***	0.07721***	0.07233***
	(0.00748)	*(0.01749)*	*(0.01463)*	*(0.00906)*	*(0.00653)*
Observations	4,216				
Goodness of fit					
Pseudo LogL	3875.4751				

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

The negative income effects on informal consumption shares and positive on their formal counterparts are maintained in almost all the labour groups. This means that the effect of total expenditure on consumption allocation holds, independent of the labour supply allocation of the family. However, some differences deserve comment. First, in the case of the informal equation, it seems that marginal effects on those households with formal workers at formal firms show a stronger negative effect when compared with households composed of individuals taking other labour options. Specifically, the results show that, for the remainder of the wage-earning sector (informal and semi-formal options), the marginal effects are not well determined and for the self-employment sector (both formal and informal) – although a negative result is encountered – in both cases they are significantly weaker[42] when compared with formal wage-earners. Second, in the case of the formal equation, the positive effect of income on formal shares is encountered across all the labour options. However, as in the previous case, the marginal effects for the most

[42] Chi-squared results for equality on parameters (informal equation). Hrs_ff vs. Hrs_sf: 7.93***. Hrs_ff vs. Hrs_si: 26.34***

modern sector (Hrs_ff) are generaly statistically stronger[43] when compared to all other labour options.

These results can be interpreted in line with the linkage hypothesis if formal wage-earners show a greater preference for formal and a lesser preference for informal consumption. Then, when compared to other labour options, any increase in income will have higher reallocation effects in favour of formal budgets. This result is accounted for by the wider gap between expenditure parameters on the formal and the informal equations. At the other extreme are those households – composed of informal wage-earners – where the reallocation of resources, given the changes in their living conditions, has the lowest effect (narrowest gap). If, as stated previously, this is the segmented working sector where it is difficult to purchase from formal markets, then their reallocation ability will be reasonably lower.[44] Different is the case for the self-employment sector. Here, it seems that, for both formal and informal segments of this sector, the reallocation of resources driven by expenditure changes has a similar effect. [45]

Although these results seem intuitive, as mentioned earlier, they could be driven by the arbitrary definition of informal consumption used. To control for this issue, the estimation is replicated for formal and informal markets, but defined using the five-market approach. As in the discussion presented in Chapter 4, only Market 1 is considered the informal sector and Market 5 its formal counterpart. This more restrictive notion of informality helps to estimate the model for the 'very' formal and 'very' informal markets. The regressions are presented in Table V.5 – for the main variables of interest only, in order to save space. As shown, most of the conclusions already discussed around the core variables appear to be robust in this new exercise.[46]

In the case of total expenditure, under the four specifications used, this regressor again has a negative and significant effect on informal consumption shares and a positive and significant effect on its formal counterpart. The point estimates show that a 10% increase in total annual expenditure will decrease informal shares by about 0.2 percentage points and will cause formal shares to grow by around 0.6 percentage points. These results seem to be slightly lower (in absolute terms) when compared with the previous definition of informality in both the formal and

[43] Chi-squared results for equality on parameters (formal equation). Hrs_ff vs Hrs_if: 0.17. Hrs_ff vs Hrs_ii: 20.43***. Hrs_ff vs Hrs_sf: 11.37***. Hrs_ff vs Hrs_si: 22.40***.
[44] Chi-squared results for equality on parameters (formal equation): Hrs_ii vs Hrs_if: 11.86***, Hrs_ii vs Hrs_sf: 3.51*, Hrs_ii vs Hrs_si: 3.07*.
[45] Chi-squared results for equality on parameters. Informal equation: Hrs_sf vs Hrs_si: 1.17; Formal equation: Hrs_sf vs Hrs_si: 0.22
[46] However, it should be noted in Appendix 3 that the direction of some of the effects for the additional variables included and commented on above do not hold. This is a striking result that deserves further investigation.

the informal equations. Therefore, households are less sensitive to income changes when more-restrictive definitions of informality are considered in estimations. As will be seen later, these results still reject inferiority.

Table V.5: Results for SUR–GLS regression (different specifications), five-market definition

	(1)		(2)		(3)		(4)	
	Informal	**Formal**	**Informal**	**Formal**	**Informal**	**Formal**	**Informal**	**Formal**
L(expenditure)	-0.01895***	0.06207***	-0.01912***	0.05936***	-0.01550***	0.06221***	-0.01952***	0.06075***
	(0.00308)	*(0.00347)*	*(0.00308)*	*(0.00344)*	*(0.00304)*	*(0.00352)*	*(0.00306)*	*(0.00344)*
Hrs_ff	0.00004	-0.00008	0.00001	-0.00009	0.00004	-0.00008	0.00004	-0.00008
	(0.00006)	*(0.00007)*	*(0.00006)*	*(0.00007)*	*(0.00006)*	*(0.00007)*	*(0.00006)*	*(0.00007)*
Hrs_if	0.00051***	-0.00024*	0.00049***	-0.00022*	0.00052***	-0.00024*	0.00053***	-0.00023*
	(0.00014)	*(0.00013)*	*(0.00013)*	*(0.00013)*	*(0.00014)*	*(0.00013)*	*(0.00013)*	*(0.00013)*
Hrs_ii	0.00023***	-0.00041***	0.00019***	-0.00038***	0.00022***	-0.00041***	0.00024***	-0.00041***
	(0.00006)	*(0.00005)*	*(0.00006)*	*(0.00005)*	*(0.00006)*	*(0.00005)*	*(0.00006)*	*(0.00005)*
Hrs_sf	0.00009	0.00040***	0.00010	0.00043***	0.00007	0.00041***	0.00008	0.00041***
	(0.00013)	*(0.00014)*	*(0.00013)*	*(0.00014)*	*(0.00013)*	*(0.00014)*	*(0.00013)*	*(0.00014)*
Hrs_si	0.00049***	-0.00015***	0.00044***	-0.00014***	0.00048***	-0.00014***	0.00048***	-0.00015***
	(0.00006)	*(0.00005)*	*(0.00006)*	*(0.00005)*	*(0.00006)*	*(0.00005)*	*(0.00006)*	*(0.00005)*
Observations	8,007		8,007		8,007		8,007	
Goodness of fit								
Pseudo LogL	11051.017		11286.781		11123.226		11195.208	
AIC	-21992.03		-22439.56		-22124.45		-22264.42	
BIC	-21607.69		-21971.36		-21698.18		-21824.17	
Wald (overall)	2578.835***		3221.053***		2697.863***		2801.823***	
Chi2 (hours)	167.22***		144.23***		164.03***		167.17***	
Income effects								
Chi2 (informal vs formal)	251.33***		236.16***		230.21***		247.93***	
Hours effects								
Chi2 (informal vs formal)	286.26***		243.56***		276.70***		290.24***	
Chi2 (pooled, joint hours)	222.59***		205.47***		214.25***		208.32***	
Intra-equation	65.57***	145.96***	56.97***	133.87***	62.52***	143.32***	60.98***	139.47***
Chi2 (pooled, formal and informal)	98.84***		86.91***		97.42***		92.78***	
Intra-equation	44.54***	43.70***	34.62***	41.14***	43.17***	43.72***	39.66**	45.24***
Chi2 (pooled, self and wage)	183.12***		168.49***		175.07***		173.58***	
Intra-equation	48.89***	129.81***	41.13***	119.18***	45.90***	127.30***	47.69***	122.82***

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1; see Table V.1 for definition of tests.

For hours of work, the results are also in line with the linkage hypothesis: working more hours in the informal (formal) sector is associated with ‘more informal (formal)’ purchasing baskets. This result holds under the four specifications used in this application. However, some changes should be noted. First, the parameters associated with informal labour options in both the formal and the informal equations are stronger under this second definition, while the parameters associated with formal labour options are weaker. Therefore, using a stricter notion of informal and formal markets reveals higher linkages (positive and negative) for informal workers to (informal and formal) consumption.

Secondly, in line with previous results under the five-market definition, the effects of Hrs_ff (formal wage-earners in formal firms) on consumption vanish. In fact, the only detectable effect in this case is the positive linkage that working in the formal self-employment sector has on formal consumption. This result has interesting implications in that it reveals the differences already discussed on the reasons for the linkages between the two formal sectors. Beyond the informational advantages, the formal self-employed could have additional reasons for increasing their households' formal consumption as they would thus obtain a fiscal advantage. This result is probably behind the reason for the robustness of the positive linkage of this sector when compared with other formal options.

Thirdly, additional effects appear: the positive correlation of both Hrs_ii and Hrs_if with informal shares and the negative ones of both Hrs_si and Hrs_if with formal shares. However, under this second definition, the positive linkage of Hrs_si with informal shares is still higher than the effect computed for Hrs_ii, while the negative effect of Hrs_ii on formal shares is still greater than that computed for Hrs_si.[47] Therefore, as detected in the case of the three-market definition is valid to conclude that positive effects of informal labour supply on informal consumption mainly comes through the self-employment sector, while negative linkages through the wage-earning counterpart.

In sum, changing the definition of informal markets does not alter the main conclusions concerning the positive and negative linkage effects. However, it does provide additional information on the robustness of some particular results. Three effects are robust to changes in the definition of informal/formal consumption: the positive linkage of working more hours in the informal self-employment sector on informal consumption, the positive linkage of formal self-employment on formal consumption, and the negative linkage between working in the informal wage-earning sector and formal consumption.

<u>Second empirical exercise: taking endogeneity into account</u>

The previous empirical exercise was executed under the assumption of exogenous expenditure and labour supply. As discussed previously, this could be considered a strong assumption as long

[47] Chi-squared results for Hrs_ii vs Hrs_si (informal equation): (1) 11.10***; (2) 10.24***; (3) 11.11***; (4) 8.75***. Chi-squared results for Hrs_ii vs Hrs_si (formal equation): (1) 15.94***; (2) 12.90***; (3) 15.95***; (4) 15.16***. Chi-squared results for equality of parameters Hrs_if vs Hrs_ii (informal equation): (1) 3.72*; (2) 4.39**; (3) 4.09**; (4) 4.05**. Chi-squared results for equality of parameters Hrs_if vs Hrs_ii (formal equation): (1) 1.42; (2) 1.26; (3) 1.49; (4) 1.65. Chi-squared results for equality of parameters Hrs_if vs Hrs_si (informal equation): (1) 0.02; (2) 0.13; (3) 0.07; (4) 0.16. Chi-squared results for equality of parameters Hrs_if vs Hrs_si (formal equation): (1) 0.49; (2) 0.31; (3) 0.44; (4) 0.31.

as there are valid reasons to believe that both sets of variables not only influence consumption, but could also be influenced by family preferences. This simultaneity could lead to endogeneity bias in the results and obscure previous conclusions around the hypothesis testing. Therefore, procedures that deal explicitly with this problem must be developed in order to again compute the results of main interest of the research: expenditure effects and the linkage effect. In this sub-section, a standard IV procedure is proposed in order to deal with the problem. However, as stated previously, although several experiments with the data were explored, the proposal to deal with several potential endogenous variables was discarded. Reducing the dimension of the problem via the imposition of additional assumptions and data restrictions was preferred. In fact, only three variables will be studied explicitly: total expenditure and overall formal and informal hours of work. Before beginning the application, it would be useful to firstly revise the independent OLS results in the new estimation context proposed.[48]

In Table V.6 these estimates are presented for the main variables of interest and under the same specifications used in the previous empirical exercise. It is important to note that the sample size used in this application is lower than previous ones, given the data restrictions already explained as well as the additional missing values on the instruments to be implemented in this application. However, beyond the possible consequences on the statistical precision of the estimates, given this data reduction, the total number of 4,874 observations (above the 50% of the original sample) is still considered adequate for estimation purposes. The table has two parts: in the top part, the hours of work are aggregated into formal and informal categories (considering the extended definition of informality explained in Chapter 4) and, in the bottom half, the results for the five labour options discussed in the previous exercise are replicated. In general, overall and individual goodness-of-fit measures are comparable to previous ones, with similar interpretation in terms of the low R2 values, mainly on the informal equations.

The information provided by the estimates is still in line with previous conclusions. Increasing total expenditure generates a reduction in informal shares and increases formal counterparts. From the point estimates, as will be seen later, it is again possible to infer that both formal and informal consumption are classified as normal (necessity and luxury, respectively) so inferiority is still rejected. For the estimates of hours of work, the evidence is also in favour of the linkage hypothesis: working more hours in the formal sector tends to bias consumption towards ‘more formal’ purchasing baskets and informal hours of work towards ‘more informal’ ones. However, it should be noted that, in the aggregate version of hours of work, the positive effects of informal

[48] The estimator for comparison purposes is an equation by equation OLS as long as this is the way that the IV application proposed is done. Considering Greene’s (2003) arguments, the estimators of independent OLS estimates are numerically similar to SUR when similar covariates are used in estimation.

hours of work are not well-determined at conventional values of significance. This is possibly because the new definition of informal labour supply is not completely capturing the effects of the informal self-employment sector, which is again consistently significant across all the specifications used. However, a possible downward bias on the estimation results (as will be concluded later) cannot be disregarded.

Table V.6: OLS estimates of total expenditure and hours of work (different specifications and definitions of labour supply)

	(1)		**(2)**		**(3)**		**(4)**	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expenditure)	-0.02398***	0.08261***	-0.03223***	0.08006***	-0.02213***	0.08166***	-0.02292***	0.08063***
	(0.00548)	(0.00526)	(0.00561)	(0.00521)	(0.00547)	(0.00530)	(0.00547)	(0.00524)
Hrs_formal	-0.00022**	0.00017*	-0.00024**	0.00018*	-0.00023**	0.00017*	-0.00024**	0.00018*
	(0.00011)	(0.00010)	(0.00010)	(0.00010)	(0.00011)	(0.00010)	(0.00011)	(0.00010)
Hrs_informal	0.00004	-0.00015***	0.00001	-0.00012**	0.00002	-0.00014***	0.00004	-0.00014***
	(0.00007)	(0.00005)	(0.00007)	(0.00005)	(0.00007)	(0.00005)	(0.00007)	(0.00005)
Observations	4,874		4,874		4,874		4,874	
Goodness of fit								
R-squared	0.103	0.283	0.131	0.293	0.110	0.284	0.110	0.290
Pseudo LogL	1612.358	2355.668	1691.885	2387.230	1633.263	2359.278	1631.223	2378.663
F (overall)	38.24***	106.11***	36.28***	85.31***	34.92***	90.85***	34.68***	89.05***
F (hours)	2.56*	5.85***	2.72*	4.63***	2.43*	5.52***	2.86*	5.90***
L(expenditure)	-0.02408***	0.08342***	-0.03199***	0.08093***	-0.02231***	0.08250***	-0.02302***	0.08144***
	(0.00549)	(0.00531)	(0.00563)	(0.00526)	(0.00549)	(0.00534)	(0.00548)	(0.00528)
Hrs_ff	-0.00030***	0.00023**	-0.00031***	0.00022**	-0.00030**	0.00022**	-0.00033***	0.00024**
	(0.00012)	(0.00010)	(0.00012)	(0.00010)	(0.00012)	(0.00010)	(0.00012)	(0.00010)
Hrs_if	0.00004	0.00031	0.00002	0.00032	0.00005	0.00030	0.00003	0.00033*
	(0.00024)	(0.00020)	(0.00024)	(0.00020)	(0.00025)	(0.00020)	(0.00024)	(0.00020)
Hrs_ii	-0.00008	-0.00036***	-0.00009	-0.00032***	-0.00010	-0.00036***	-0.00008	-0.00036***
	(0.00010)	(0.00008)	(0.00011)	(0.00008)	(0.00010)	(0.00008)	(0.00010)	(0.00008)
Hrs_sf	0.00007	0.00031	0.00000	0.00035	0.00004	0.00032	0.00009	0.00029
	(0.00030)	(0.00032)	(0.00030)	(0.00033)	(0.00031)	(0.00032)	(0.00030)	(0.00032)
Hrs_si	0.00027**	-0.00003	0.00022**	-0.00001	0.00025**	-0.00003	0.00027**	-0.00003
	(0.00011)	(0.00009)	(0.00011)	(0.00009)	(0.00011)	(0.00009)	(0.00011)	(0.00009)
Observations	4,874		4,874		4,874		4,874	
Goodness of fit								
R-squared	0.107	0.287	0.135	0.296	0.115	0.288	0.114	0.294
Pseudo LogL	1624.738	2367.680	1701.972	2398.588	1644.940	2371.271	1644.318	2392.309
F (overall)	28.87***	77.83***	29.10***	68.21***	27.51***	69.71***	27.82***	69.45***
F (hours)	2.85**	6.53***	2.48**	5.32***	2.66**	6.28***	3.04***	6.55***

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Turning our attention to the exercises proposed in this sub-section, the objective now is to re-estimate the parameters of total expenditure and overall formal and informal hours of work under standard IV procedures. Two exercises are developed for this purpose. The first deals with the endogeneity of expenditure solely under the assumption of exogenous labour supply. The second deals with hours of work, assuming that total expenditure is exogenous. In each case, the sub-set of the available instruments already described in previous sections was used. However, the inclusion of instruments in the final specification of each of the particular exercises developed here was undertaken basically using statistical procedures: maximization of the global F-statistic in order to pass relevance tests and minimization of the J-Sargan statistic in order to pass orthogonality tests.[49] These objectives were fulfilled with different levels of success in each case.

In the first exercise, the instruments were the logarithm of the value of the assets owned by the household and five measures of exogenous shock: if a member loses his/her job, the bankruptcy of the family business, the abandonment of the family by the household head, a family experience of robbery or other crimes against it and the family's experience of a natural disaster. In the second exercise the instruments were the logarithm of the value of the assets owned by the household and its square, the logarithm of wage rates in the formal and formal sector, a measure of non-labour income, a dummy that identifies the inadequate physical condition of the house and one measure of risk: if the family has experienced robbery or other crime against it.

In Table V.7, the diagnostic tests for the first IV exercise are presented. As explained above, the interest is in handling the endogeneity of total expenditure under the assumption of exogenous labour supply of the family. As shown, both relevance and orthogonality conditions are met. Similarly, when the C-test is done for suspicious included exogenous variables, the procedure reveals that the assumption of exogeneity of labour supply will not alter the results. Finally, the C-test for the potential endogenous variable reveals that total expenditure is correlated with the error terms, at least in the formal equation, so the IV application is justified. In Appendix 3 the first-stage regression results are displayed. The additional instruments display the expected signs. For example, the logarithm of the value of total assets has a strong and positive effect on expenditure and most of the shocks, when significant, show a negative correlation with this variable. These results are stable and robust to the inclusion of additional exogenous variables in the main equation.

Table V.7: Diagnosis tests for IV application under one endogenous variable (expenditure with exogenous labour supply), different specifications

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
Underidentification Test	685.668***		675.361***		675.922***		682.408***	
F-value (1stage)	181.227***		176.460***		177.465***		178.933***	
Hansen J test	6.972	4.584	3.908	3.077	6.622	4.617	7.424	4.662
C (participation and hours)	6.563	4.525	3.646	3.072	6.345	4.585	6.716	4.618
C (expenditure)	2.151	30.668***	2.528	30.915***	1.525	30.051***	1.704	29.913***

Note: *** p<0.01, ** p<0.05, * p<0.1. Tests are performed under a 2SLS specification with small sample correction and robust standard errors. The underidentification test is the Kleibergen-Paap rk LM statistic, F-value is computed as the Kleibergen-Paap rk Wald F statistic and compared with the critical value of 19.28 (5% of maximal IV relative bias) and 11.12 (10% of maximal IV relative bias) computed in Stock-Yogo (2005) as the minimum tolerable levels, the Hansen J test for overidentifying restrictions is computed under the null that instruments are valid and exclusion restrictions adequate, C-test for controversial included instruments tests the exogeneity of the variables implemented under the null of instrument validity, C-test for endogenous regressors tests exogeneity under the null that suspect variables can be treated as exogenous.

[49] Using just a sub-set and not the long list of instruments is also in order to avoid the finite sample problems as commented on in Wooldridge (2002).

In Table V.8, the 2SLS results are presented only for the variable of main interest. The goodness-of-fit measure still reveals adequate global adjustment of the models. The estimations are run under correction for small sample size and with robust standard errors considering possible loss of precision. The estimates confirm previous conclusions around the marginal effects of this particular regressor: total expenditure reveals a negative effect on informal consumption shares and a positive effect on formal ones. Therefore, it is still possible to confirm the differences in the effect of total consumption on resource allocation, with families increasing their participation in informal consumption as long as their total expenditure decreases and reducing it for income expansion. This result, already encountered in previous exercises under the assumption of exogeneity, holds when this assumption is dropped. Also, the point estimates presented here reveal that coefficients are slightly larger (in absolute terms) under the IV estimation when compared with OLS or SUR estimates. A 10% increase in annual expenditure will reduce informal shares by between 0.3 and 0.5 of one percentage point and will increase formal ones by around 1.3 percentage points. Therefore, the evidence suggests a downward bias of estimates when the exogeneity of total expenditure is not explicitly handled but, as will be shown later, it is also still possible to conclude that informal consumption can be classified as a normal necessity and formal consumption as a normal luxury. Inferiority is again rejected.

Table V.8: Regression results for the IV model under one endogenous variable (expenditure with exogenous labour supply), different specifications

	(1)		**(2)**		**(3)**		**(4)**	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expenditure)	-0.03660***	0.12819***	-0.04655***	0.12788***	-0.03289***	0.12799***	-0.03406***	0.12567***
	(0.00968)	(0.00860)	(0.01007)	(0.00899)	(0.00985)	(0.00880)	(0.00967)	(0.00859)
Observations	4,874	4,874	4,874	4,874	4,874	4,874	4,874	4,874
LogL	1608.341	2285.586	1686.809	2313.037	1630.370	2288.043	1628.093	2310.029
F-value (overall)	37.20***	110.16***	34.58***	86.67***	34.19***	93.67***	33.82***	91.35***

Note: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parenthesis. Regression with small sample corrections.

In Table V.9 the diagnostic tests for the second IV exercise are presented – i.e. the estimations that explicitly handle the endogeneity of hours of work under the assumption of exogenous total expenditure (and participation dummies). The first-stage results are again displayed in Appendix 3, where the logarithm of assets displays a non-linear relationship with both formal and informal labour supply; in the first case, first decreasing and then increasing; and in the second case first increasing and then decreasing. However, in both cases, inflexion points occur below percentile 30 (27 for formal hours of work and 18 for informal ones), so it is possible to infer that the positive relationship dominates in the case of formal hours of work and the negative relation in the informal one. This is verified when the sign of the marginal effect is inspected at the mean value of assets. The result possibly reveals the positive impact of long-term wealth on the capacity to engage in formal labour options. A similar result is encountered for the dummy variable that identifies inadequate household conditions on formal hours of work. In the case of own wages, they reveal a negative correlation with hours of work so, for the sample used here, it seems that income effects tend to dominate over substitution effects. However, in the case of informal hours of work, a negative cross effect is also displayed, something that could argue in favour of desincentives (incentives) to work informally given higher (lower) formal wages. Finally, the measure for non-labour income displays the standard result reducing hours of work, mainly in the informal sector. These results are stable and robust to the inclusion of additional exogenous variables in the main equation.

Table V.9: Diagnosis tests for IV application under two endogenous variables (hours of work with exogenous expenditure and participation dummies), different specifications

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
Underidentification Test	52.738***		47.029***		50.738***		53.435***	
F-value (1stage, Hrs_formal)	7.71***		8.22***		8.22***		7.32***	
F-value (1stage, Hrs_informal)	6.32***		6.92***		6.40***		6.25***	
F-value (1 stage, joint)	6.400		6.225		6.362		6.291	
Hansen J test	5.695	7.051	3.505	6.842	5.789	6.742	6.985	6.286
C (participation and expenditure)	3.856	5.185	2.603	5.39	4.072	4.718	5.557	4.397
C (hours)	42.116***	63.311***	42.452***	58.655***	41.435***	64.400***	38.718***	58.700***

Note: *** p<0.01, ** p<0.05, * p<0.1. Tests are performed under a 2SLS specification with small sample correction and robust standard errors. Individual F-statistics must be compared with the rule of thumb of 10 (Cameron and Trivedi, 2005), Joint F statistics are compared with the critical value of 16.88 (5% of maximal IV relative bias) and 9.92 (10% of maximal IV relative bias) computed in Stock-Yogo (2005) as the minimum tolerable levels. See footnote in Table V.7 for other details in the tests.

Although, these correlations seem intuitive, from the diagnosis presented here it is possible to conclude that instruments are orthogonal (non-significant Hansen test), but they are weak (F-statistic below critical values). Therefore, considering the discussion in Cameron and Trivedi (2005), it is expected that the results will suffer from finite sample biases. This means that, although estimators could be asymptotically consistent, for finite samples they will be biased in the direction of the (inconsistent) OLS estimates. Also, the loss of precision normally attributable to IV estimators (Wooldridge, 2002) could be even greater in the presence of weak instruments (Cameron and Trivedi, 2005). To partially control for these problems, additional estimators to the standard 2SLS are computed and compared. As discussed in Stock and Yogo (2005), Cameron and Trivedi (2005) and Flores-Lagunes (2007), some of the IV procedures that are more robust to the presence of weak instruments are the Limited Information Maximum Likelihood (LIML) and the Generalized Method of Moments (GMM). These two additional estimators and the 2SLS are estimated with small sample corrections and standard errors robust to heteroskedasticity.

Table V.10: Regression results for the IV model under two endogenous variables (hours of work with exogenous expenditure and participation dummies), different specifications

(1)	2SLS		LIML		GMM	
	Informal	Formal	Informal	Formal	Informal	Formal
Hrs_formal	-0.00519***	0.00307*	-0.00584***	0.00357	-0.00518***	0.00322*
	(0.00165)	(0.00174)	(0.00198)	(0.00225)	(0.00164)	(0.00173)
Hrs_informal	0.00334***	-0.00547***	0.00389***	-0.00654***	0.00342***	-0.00520***
	(0.00100)	(0.00106)	(0.00124)	(0.00142)	(0.00099)	(0.00105)
Observations	4,874	4,874	4,874	4,874	4,874	4,874
LogL	107.852	-268.526	-234.179	-881.185	74.277	-137.136
F-value (overall)	22.25***	46.04***	19.38***	37.44***	22.28***	46.90***
(2)	2SLS		LIML		GMM	
	Informal	Formal	Informal	Formal	Informal	Formal
Hrs_formal	-0.00269*	0.00278*	-0.00294*	0.00335	-0.00282*	0.00286*
	(0.00158)	(0.00167)	(0.00176)	(0.00213)	(0.00157)	(0.00166)
Hrs_informal	0.00443***	-0.00540***	0.00483***	-0.00638***	0.00434***	-0.00516***
	(0.00095)	(0.00103)	(0.00108)	(0.00135)	(0.00094)	(0.00102)
Observations	4,874	4,874	4,874	4,874	4,874	4,874
LogL	-11.446	-193.969	-234.354	-778.704	21.735	-72.473
F-value (overall)	20.24***	35.42***	18.54***	28.95***	20.26***	35.99***
(3)	2SLS		LIML		GMM	
	Informal	Formal	Informal	Formal	Informal	Formal
Hrs_formal	-0.00518***	0.00315*	-0.00583***	0.00367*	-0.00517***	0.00326*
	(0.00162)	(0.00172)	(0.00193)	(0.00219)	(0.00161)	(0.00171)
Hrs_informal	0.00330***	-0.00550***	0.00385***	-0.00651***	0.00338***	-0.00525***
	(0.00099)	(0.00106)	(0.00123)	(0.00140)	(0.00098)	(0.00105)
Observations	4,874	4,874	4,874	4,874	4,874	4,874
LogL	129.851	-290.423	-215.719	-877.996	99.654	-166.906
F-value (overall)	20.32***	38.54***	17.66***	31.53***	20.29***	39.24***
(4)	2SLS		LIML		GMM	
	Informal	Formal	Informal	Formal	Informal	Formal
Hrs_formal	-0.00590***	0.00391**	-0.00677***	0.00446**	-0.00584***	0.00395**
	(0.00165)	(0.00167)	(0.00204)	(0.00210)	(0.00165)	(0.00166)
Hrs_informal	0.00255***	-0.00461***	0.00304**	-0.00541***	0.00270***	-0.00447***
	(0.00098)	(0.00098)	(0.00127)	(0.00127)	(0.00098)	(0.00097)
Observations	4,874	4,874	4,874	4,874	4,874	4,874
LogL	272.774	109.448	-82.385	-393.124	236.391	179.196
F-value (overall)	21.18***	41.56***	18.30***	35.12***	21.07***	42.72***

Note: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parenthesis. Regression with small sample corrections. The numbers on the left correspond to the specification used. See Table V.1 for additional variables included in each specification.

In Table V.10 the regression results are presented for the different estimation methods and the specifications used across the application.[50] As shown, in general a positive marginal effect of informal hours of work on informal budget shares is verified, with negatives ones in the case of

[50] Diagnosis tests were re-calculated under each of these new estimators, without changes to those already presented in Table V.9.

its formal counterparts. Similarly, we find that working more hours in the formal sector has a negative impact on informal consumption shares and a positive impact on the formal ones. These results are robust to the specification used and the different IV methods implemented so it could be considered strong evidence in favour of the linkage hypothesis. Only in two estimated equations is it found that the parameter associated with formal hours of work on formal consumption is not well determined: specification (1) and (2) under the LIML method. However, they do not deviate from the maintained conclusion of the application.

The point value estimates reveal that hours' effects are above OLS regressions results in absolute terms. In fact, on average and *ceteris paribus*, one additional hour of work during a representative week in the informal sector will increase (reduce) informal (formal) shares by between 0.3 and 0.5 (0.5 and 0.7) of one percentage. By contrast, one additional weekly hour of work will increase (reduce) formal (informal) shares by between 0.3 and 0.5 (0.3 and 0.7) of one percentage. Therefore, as in the case of total expenditure, we can confirm that the evidence in favour of the linkage hypothesis is stronger when IV methods are used.

Table V.11: Hypothesis testing under weak instruments

(1)	2SLS		LIML		GMM	
	Informal	Formal	Informal	Formal	Informal	Formal
K (Ho: equal to zero)	45.97***	78.01***	45.97***	78.01***	45.97***	78.01***
K (Ho: equal to parameter)	0.30	0.85	0.30	0.09	0.25	0.83
(2)	2SLS		LIML		GMM	
	Informal	Formal	Informal	Formal	Informal	Formal
K (Ho: equal to zero)	47.62***	74.43***	47.62***	74.43***	47.62***	74.43***
K (Ho: equal to parameter)	0.09	0.43	0.01	0.04	0.12	0.74
(3)	2SLS		LIML		GMM	
	Informal	Formal	Informal	Formal	Informal	Formal
K (Ho: equal to zero)	45.31***	78.83***	45.31***	78.83***	45.31***	78.83***
K (Ho: equal to parameter)	0.33	0.41	0.04	0.05	0.25	0.73
(4)	2SLS		LIML		GMM	
	Informal	Formal	Informal	Formal	Informal	Formal
K (Ho: equal to zero)	41.39***	67.38***	41.39***	67.38***	41.39***	67.38***
K (Ho: equal to parameter)	0.46	0.33	0.06	0.05	0.33	0.51

Note: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parenthesis. Regression with small sample corrections. The tests correspond to each of the estimations presented in Table V.10.

An additional problem with weak instruments, as discussed by Nelson and Starz (1990), is that the normal approximation of the t-statistic under IV performs poorly if it is not based on strong instruments (see Mikusheva and Poi, 2006). Therefore, hypothesis testing will be distorted. For this reason, even in the case of the additional estimators (like LIML or GMM) which will be robust to finite sample biases, it is problematic to rely on t-tests. To control for this issue, alternative procedures must be developed. A survey of methods available are discussed in Andrews and Stock (2005) and Stock *et al*. (2002). In this application, the Magnusson (2010) minimum-distance (MD) versions of the weak identification robust tests introduced by Kleibergen (2002) and Moreira (2003) are used. These are implemented using the K-statistics programmed in Finlay *et al*. (2013) for the joint hypothesis that both parameters are equal to zero and both take the exact values estimated previously. This statistic is based on a Lagrange Multiplier (score) valid in the presence of potentially weak instruments. The results are presented in Table V.11 under the two null hypotheses. As shown, rejection in the first case and non-rejection in the second one gives adequate confidence in previous estimates and helps to conclude that they are well determined and significant at conventional levels.

A final check is to use the Conditional Likelihood Ratio test (CLR) developed in Moreira (2003). The motivation for using this additional procedure is based on the observation of Finlay and Magnuson (2009), who states that K-statistics suffer from a spurious decline of power in some regions of the parameter space. By contrast, as discussed in Andrews *et al*. (2007), the CLR is nearly optimal in a class of invariant similar tests (Mikusheva and Poi, 2006). However, it has two drawbacks. First, it assumes homoscedasticity (which is not the case for the LM test presented previously). Second, as programmed by Mikusheva and Poi (2006), it is only possible to implement for one endogenous regressor. For this reason, in this exercise the formulation of the demand equation is changed, replacing hours of work in the formal and informal sector by the difference between them. Therefore, the exercise implies the assumption that the parameters accompanying formal and informal hours of work in the original formulation of the demand equations are equal in absolute terms.[51]

Table V.12: Diagnosis tests for IV application under one endogenous variable (difference between hours of work with exogenous expenditure and participation dummies)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
Underidentification Test	43.642***		42.387***		43.152***		43.966***	
F-value (1stage)	6.940		6.742		6.861		7.018	
Hansen J test	3.89	5.376	3.700	4.844	4.086	5.349	3.137	6.237
C (expenditure and participation)	2.121	3.767	1.768	4.005	2.405	3.724	2.008	4.300
C (hours difference)	43.138***	55.216***	38.000***	49.788***	42.643***	55.644***	39.649***	52.408***

Note: *** p<0.01, ** p<0.05, * p<0.1. Tests are performed under a 2SLS specification with small sample correction and robust standard errors. F statistics are compared with the critical value of 19.28 (5% of maximal IV relative bias) and 11.12 (10% of maximal IV relative bias) computed in Stock-Yogo (2005) as the minimum tolerable levels. See footnote in Table No. V.7 for other details of the tests.

To render them comparable with previous estimates, the instruments considered in this exercise are similar to those used in the case of two endogenous variables (formal and informal hours of

[51] This assumption was tested using Wald and K tests in the previous exercise and it was found that equality of the parameters in absolute terms exists for some ranges of possible values.

work). The only difference is that the logarithms of the wage rates of formal and informal earning opportunities now enter as the difference between them. The diagnosis tests are presented in Table V.12. As shown, similar conclusions to those already obtained still hold: orthogonality of the instruments is achieved, but relevance tests reveal that they are only weakly correlated to the endogenous regressor.

Table V.13: Regression results for the C-IV model under one endogenous variable (difference in hours of work with exogenous expenditure and participation dummies)

	(1)		**(2)**		**(3)**		**(4)**	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
Hrs_formal-Hrs_informal	-0.00447***	0.00521***	-0.00427***	0.00494***	-0.00447***	0.00525***	-0.00415***	0.00503***
	(0.00098)	(0.00102)	(0.00097)	(0.00100)	(0.00099)	(0.00104)	(0.00093)	(0.00101)
Observations	4,874	4,874	4,874	4,874	4,874	4,874	4,874	4,874
Wald (overall)	276.63***	643.60***	368.83***	712.37***	295.75***	638.65***	316.49***	690.22***
CLR Test	45.4011***	81.4034***	41.0309***	71.5148***	45.0235***	81.7393***	40.6478***	75.6476***

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parenthesis. Regression with small sample corrections. See Table V.1 for additional variables included in each specification.

The estimations results are presented in Table V.13. The results are still in line with the linkage hypothesis, given the way that the new variable has been defined. As shown, increasing weekly hours of work in the formal or reducing them in the informal sector decreases informal budget shares and increases their formal counterparts. Point value estimates reveal that the size of the effects is around -0.4 and 0.5 of one percentage point, respectively. These coefficients are well determined and, in general, lie between previous coefficient values (in absolute terms). The CLR test for the significance of the parameter of interest is presented at the bottom of the estimations. The results of the test confirm the significance of the results presented here and provide additional evidence in favour of the linkage effect.

V.4. Computed elasticities and policy implications

In this section the regression results are used to derive the main elasticities of interest in the chapter, so they can easily be interpreted for public policy purposes. The focus is put on total expenditure and hours of work as defined in expressions (V.2) to (V.4). The computation of elasticities done here uses the average sample means for all the variables so, strictly speaking, the exercise proposed for hours of work is something like a projection over any participation regime; i.e., an approximation of unconditional averages. In Table V.14 the expenditure elasticities are displayed for the different estimation methods presented in this application: SUR–GLS results for both three- and five-market definitions and the IV estimation. In this last case, only those estimations that explicitly handle the endogeneity of total expenditure are presented. Also, the results are supplemented by the point elasticities computed for the sub-populations in order to assess differentiated effects. In each case, taking into account the relative stability of the marginal effects across the specifications used, the results are presented for specification (1).

Table V.14: Expenditure elasticities

Whole sample results					
	SUR-GLS (3 market)	SUR-GLS (5 market)	IV-2SLS		
Informal	0.932***	0.893***	0.951***		
	(0.01051)	*(0.01740)*	*(0.00788)*		
Formal	1.290***	1.323***	1.437***		
	(0.01319)	*(0.01804)*	*(0.02934)*		
Quintiles					
	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Informal	1.107***	0.995	1.017	0.804**	0.663***
	(0.03347)	*(0.09502)*	*(0.10118)*	*(0.08556)*	*(0.03079)*
Formal	1.038	1.364***	1.577***	1.320***	1.238***
	(0.07014)	*(0.13191)*	*(0.14160)*	*(0.09805)*	*(0.02786)*
City types					
	Metropolitan areas	Intermediate cities	Minor cities		
Informal	0.824***	0.932***	0.992		
	(0.02045)	*(0.01769)*	*(0.01649)*		
Formal	1.338***	1.308***	1.236***		
	(0.02059)	*(0.02053)*	*(0.02352)*		
Labour sector					
	D(Hrs_ff)	D(Hrs_if)	D(Hrs_ii)	D(Hrs_sf)	D(Hrs_si)
Informal	0.793***	0.938	0.985	0.902***	0.953**
	(0.02234)	*(0.05541)*	*(0.03030)*	*(0.03263)*	*(0.01869)*
Formal	1.315***	1.441***	1.217***	1.236***	1.283***
	(0.02055)	*(0.06307)*	*(0.06847)*	*(0.02775)*	*(0.02556)*

Note: Elasticities are computed considering sample means. Robust standard errors in parenthesis with Weesie (1999) corrections for between vce in the case of quintiles and cities. *** p<0.01, ** p<0.05, * p<0.1. The statistical tests are for the null hypothesis 'different from one'. IV estimation corresponds to the 2SLS for the empirical exercise that explicitly handles endogeneity in this variable. The results revealed by sub-groups use SUR–GLS estimates for the three-market definition. All estimations presented here corresponds to specification (1).

A first important result to note is that, as mentioned earlier, inferiority of informal consumption is rejected in all estimation methods, even for the most informal markets and the richest individuals in the sample. As also mentioned earlier, informal consumption with elasticities below 1 can be classified as necessities and formal consumption with elasticities above 1 as luxuries. One additional *Nuevo Sol* in the pocket of a Peruvian consumer will increase both formal and informal expenditure. However, the increase in the former will be greater than in the latter, so in terms of the purchasing basket, as long as the family becomes richer, it will be more biased in favour of formal expenditure. In terms of the point value estimates, for the whole sample a 10% increase in annual expenditure is consistent with increases of between 8.9% and 9.5% in informal consumption and between 12.8% and 14.4% in the case of its formal counterparts, depending on the estimation method used. Using the sample means of formal and informal consumption,[52] this means that any additional S/.10 that Peruvian families have will be distributed S/.3.5 at informal markets and S/.4.6 at formal ones (the rest in semi-formal markets).

This is an important result since it reveals a distinctive feature in highly informal developing countries when compared to developed ones. For example, Fortin *et al*. (2000) found zero or even negative elasticities in their application for Canada. The inferiority of informal goods may be reasonable in a developed economy with a small informal sector, where informal purchases are basically a tax-avoiding strategy. By contrast, in a developing economy with a very big informal sector (like that of Peru), households buy from informal markets not only to avoid regulation, but also to satisfy needs. In this context, it is reasonable to expect an elasticity of less than 1, but still positive income elasticities for informal consumption.

This evidence is in line with findings in other developing countries, like the estimates for Serbia by Reilly *et al*. (2006) and Böhme and Thiele's estimates (2012b) for Africa, although point values differ across them. The results presented for Peru in this application show income elasticities that range between 0.89 and 0.95. The results computed by Reilly *et al*. (2006) for Serbia are on the high side of this range, with point value estimates of 0.94. By contrast, results presented by Böhme and Thiele (2012b) for Africa range from 0.68 to 0.73 among the countries considered in the application.[53] A possible ordering emerges, with transitional European countries at the top of the distribution, Latin American countries in the middle but close to the first ones, and African countries way down at the bottom. A possible explanation for this behaviour is related to the development or income levels of the economies under analysis – clearly higher for Serbia and Peru than for Africa.[54] However, this is entirely speculative, given the limited evidence around this issue and the methodological differences between papers. An interpretation like this also contradicts the results encountered in this application on city types, where less-developed cities show stronger informal income elasticities. Therefore, as long as other explanations are still possible (maybe country-specific), more research will be needed on this issue to reach to any empirical regularity.

---

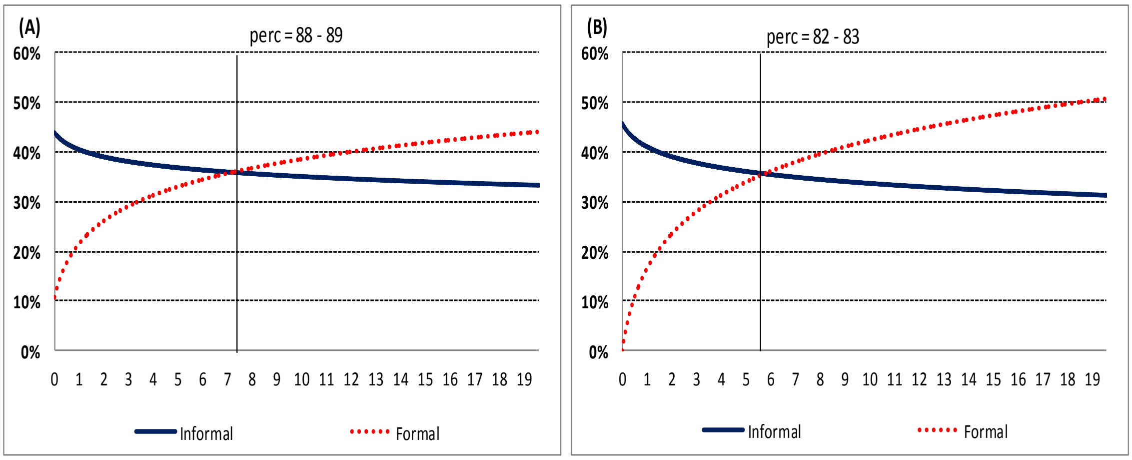
[52] Annual means for the sample of 8,007 observations (SUR–GLS model): Informal consumption: (S/. 5,725), Formal consumption (S/. 5,191), Total expenditure (S/. 15,319). Annual means for the sample of 4,874 observations (IV model): Informal consumption: (S/. 5,496), Formal consumption (S/. 5,049), Total expenditure (S/. 14,792).

[53] These results are for informal channels of commercialization in order to make them comparable with the results presented here. To obtain a unique number, weighted averages for formal goods, informal goods, imported goods and services are computed.

[54] Results from the World Bank (www.worldbank.org) show that *per capita* Gross National Income (GNI) in PPP values for the period 2010–2014 are US$ 12,480 (Serbia); US$ 11,160 (Peru), and between US$ 890 and US$ 2,210 (for the African countries considered by Böhme and Thiele, 2012b).

In any case, the results for income elasticities presented thus far are consistent with three stylized facts: the protective attitudes of families (increasing their exposure to informal markets) as incomes are reduced, quality differences as long as families prefer to change their purchasing basket composition when income increases, and positive effects of economic growth on the formalization of the country from a consumption perspective. This formalization process is explored in detail in Figure V.1, using the predicted budget shares for changes in total expenditure, holding everything else constant at their sample means. The results are shown for the most aggregated definition of informality (three-market) and considering both SUR–GLS (Panel A) and IV (Panel B) methods for comparison purposes.

Figure V.1: Predicted budget shares (formal vs informal) for different realizations of total expenditure and considering different estimation methods



Note: The vertical axis displays the budget shares and the horizontal the annual *per capita* expenditure in thousands of new *soles*. Specification (1) is used.

As shown, income growth will effectively generate more formal baskets, but this is a process with limits. Informal consumption shares will dominate even for income levels equal to those of the richest households. In fact, it seems that, for the Peruvian economy, the intersection between formal and informal shares will happen around the percentiles 82 and 89 and there is a convergence of around 30% beyond that point. Moreover from the estimations presented here, informal shares are less responsive (in absolute terms) to income if compared with formal ones. Formal shares go from 0–10% (poorest families) to around 45–50% (richest families) and informal shares from around 40–50% (poorest families) to around 30% (richest families). Therefore, the formalization process described in Cermeño (1987), where poor households are exclusively informal consumers and the richest ones exclusively formal, with middle-class consumers purchasing from both sectors, needs to be revised. In fact, the results suggest that Peruvian consumers are, in one way or another, informal and it is their formal consumption e that is accommodated depending on their income.

A second important result to note is that income elasticities are not equally distributed for all the distributional range or for the different sub-population groups in society. This conclusion is reached using the SUR–GLS estimates for the different sub-samples proposed. As shown and discussed previously, among the poorest individuals in the sample (Q1), a positive shock on expenditure will generate 'more informal' consumption baskets with informal purchases growing faster than their formal counterparts. This result changes for individuals in the upper quintiles (Q2 to Q5), but even there it is still possible to conclude that income growth will be more effective on the right of the income distribution. The results are more or less confirmed by the other groups considered during the estimation. For example, income elasticities prove to be stronger for families living in metropolitan areas than those living in intermediate or minor cities. Similarly, the gaps encountered for income elasticities are higher for those households with workers in the most modern sector (formal wage-earners) than for any other working group. In consequence, an interesting result of the application is that targeted income transferences promoted by governments in developing countries could come with the short-term cost of increasing informality or at least reducing the effectiveness of any formalization policy. Also, pro-poor growth promoted by governments in developing countries needs to be strong enough to dramatically change the living conditions of households in order to achieve positive side-effects in the formal/informal structure of consumption.

In Table V.15, the results for the hours' marginal effects (in consumption levels) and elasticities are presented. To preserve the generality of the discussion, here the focus is placed on the whole sample estimates for the different estimation methods. Again, the computed elasticities presented here correspond to specification (1) to save space and in the case of the IV method only 2SLS estimates are displayed. The results reveal that supply-side policies oriented to a reduction in the hours of work a household allocates to informal sectors will reduce the consumption shares allocated to informal markets. Just looking at the (more robust) IV estimates, point value estimates reveal that reducing by 10% the hours that a family works in informal sectors will cause an annual reduction of almost 4% in informal consumption and an increase of almost 9% in formal expenditure. In terms of the marginal effects computed, one less weekly hour allocated in the informal sector reduces annual informal expenditure by S/.49 and increases formal ones by S/.81. By contrast, a shock that increases formal working hours by 10% will cause a reallocation in favour of formal purchases of almost 3% and reduce informal ones by practically the same magnitude. In terms of the additional weekly hour, this means a reduction of S/.77 in informal and of S/.45 in formal consumption.

Table V.15: Hours-of-work elasticities (pure labour effect), different estimation methods

		Marginal effects (consumption levels)		Elasticities	
		Informal	Formal	Informal	Formal
IV (3 market)	Hrs_for	-76.82***	45.35*	-0.353***	0.278*
		(24.39716)	*(25.78811)*	*(0.11219)*	*(0.15781)*
	Hrs_inf	49.38***	-80.94***	0.397***	-0.866***
		(14.73745)	*(15.70852)*	*(0.11846)*	*(0.16802)*
SUR-GLS (3 market)	Hrs_ff	-4.10***	3.34***	-0.014***	0.016***
		(1.35955)	*(1.28653)*	*(0.00478)*	*(0.00611)*
	Hrs_if	2.65	1.61	0.002	0.002
		(2.95395)	*(2.31150)*	*(0.00272)*	*(0.00288)*
	Hrs_ii	-1.82	-4.95***	-0.006	-0.021***
		(1.27239)	*(0.95259)*	*(0.00394)*	*(0.00398)*
	Hrs_sf	0.86	8.23***	0.001	0.015***
		(2.64241)	*(2.75986)*	*(0.00351)*	*(0.00495)*
	Hrs_si	4.41***	0.89	0.021***	0.006
		(1.05516)	*(0.88307)*	*(0.00499)*	*(0.00564)*
SUR-GLS (5 market)	Hrs_ff	0.55	-1.24	0.004	-0.009
		(0.95955)	*(1.08857)*	*(0.00750)*	*(0.00783)*
	Hrs_if	7.82***	-3.72*	0.016***	-0.007*
		(2.09422)	*(2.01566)*	*(0.00429)*	*(0.00380)*
	Hrs_ii	3.50***	-6.25***	0.024***	-0.040***
		(0.90589)	*(0.77645)*	*(0.00623)*	*(0.00491)*
	Hrs_sf	1.32	6.08***	0.004	0.017***
		(1.98717)	*(2.15312)*	*(0.00587)*	*(0.00585)*
	Hrs_si	7.52***	-2.23***	0.079***	-0.022***
		(0.85194)	*(0.70598)*	*(0.00895)*	*(0.00682)*

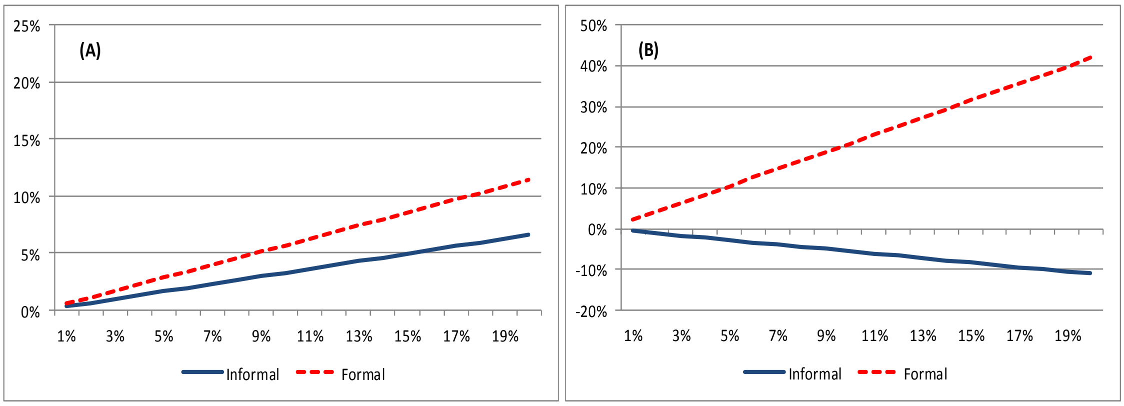
Note: Elasticities are computed considering sample means. Robust standard errors in parenthesis,*** p<0.01, ** p<0.05, * p<0.1. IV estimation corresponds to the 2SLS for the empirical exercise that explicitly handles the endogeneity in these variables. Marginal effects are computed as the parameter of the share equation multiplied by the total expenditure.

Interestingly enough, these results are in the lower limit when compared with those reported by developed countries. Using Fortin, et.al. (2000), estimates for Canada reveal elasticities between working and purchasing at informal markets that range between 0.38 and 0.73 (the authors don´t provide results for the formal sector). Barring methodological differences, the results are possibly saying that linkage effects are generally lower for developing countries. Again, this is a reasonable result if tax avoidance is the main motivation of informal consumption in the developed world, so more strategic behaviours are expected. However, in contrast to Böhme and Thiele's (2012b) findings, linkages are still detectable, which makes the results more comparable to those of Reilly *et al*. (2006), although the authors compute elasticities to make a quantitative assessment.

Moreover, in this application we can see that linkage effects are not equally applicable for all formal and informal working opportunities. Using the (possibly downwardly biased) SUR–GLS estimates for the different earning options, and considering the two definitions of informal consumption implemented, specific channels of linkage hypothesis could be suggested. The robust results encountered in the application are highlighted. They suggest that the positive linkage with informal consumption comes basically in the form of informal self-employment while the positive linkage with formal consumption takes the form of formal self-employment. Therefore, we could argue that, if supply-side policies are used to reduce (increase) informality (formality) on the consumption side, the focus must be put on the self-employment sector.

Given this set of results, expenditure and hours' elasticities can be used to compute the full hours' effects as defined in equation (V.4) and produce policy simulations. For that purpose, a supply-side formalization policy is simulated via the reduction of informal hours of work of the family and a consequent increase in the formal ones (so total labour supply remains unaltered). This is possible, for example, if we consider a reduction in labour costs for the formal sector (work taxes or labour regulations) or a better enforcement capacity for the government. This policy will produce two effects. First, a negative effect on informal consumption and a positive effect on formal consumption, given the linkage effects already modelled. Second, a positive effect on both the formal and informal sector given the income increase of the family, larger in the former than in the latter. This last result comes from the substitution of low-salary informal jobs with high-salary formal ones.

Figure V.2: Policy simulations (formalization policy on the supply-side)



Note: To save space, the results are presented only for specification (1) in each case. The vertical axis displays the expected growth of expenditure in each sector and the vertical axis the decrease in hours worked in the informal sector (policy result).

The simulations are presented in Figure V.2. In Panel (A) the results for the SUR–GLS are displayed. In Panel (B) the results for the IV method are shown. In both cases, the three-market definition is used. To read the results, in each case, the percentage decrease of hours worked in the informal sector is displayed on the horizontal axis. In the case of the SUR–GLS estimations, as long as there are five labour options then the reduction (in percentages) is assumed to be equal in the informal self–employment sector (Hrs_si) and the informal wage sector (Hrs_ii). These hours are assumed to be completely absorbed by the formal sector in the following way: the number of hours reduced in the informal self-employment sector are absorbed by the formal self-employment sector (Hrs_sf) and the number of hours reduced in the informal wage sector are absorbed by the formal wage-earning sector (Hrs_ff). In this exercise, the intermediate category of wage-earners (Hrs_if) is not considered. In the case of the IV, as long as there are only two labour options then simulation is easier. The reduction in overall informal hours (Hrs_inf) of work in this case is completely absorbed by the formal sector (Hrs_for). On the vertical axis, the result of the policy for the different reductions is displayed in terms of the percentage change in the consumption level of informal and formal markets.

As shown, a policy like this generates a formalization process on the demand-side of the economy. However, the differences between the estimation methods should be noted. For the SUR–GLS estimations, formalization occurs when formal consumption grows more than its informal counterpart. This mainly happens because, in this estimate, the income elasticities dominate the linkage effects. As stated previously, linkage effects, computed under the assumption of exogenous labour supply, are weaker when compared with estimates where this assumption is dropped. By contrast, under the IV methods, the policy seems to be much more effective, with formal consumption growing faster and informal consumption even decreasing. In this last situation, estimated linkage effects dominate expenditure elasticities. This type of result has been used in previous studies to justify the existence of the second-round effects of formalization policies on the labour market as long as they also generate formalization on the demand side. This application found evidence in favour of these second-round effects in the case of Peru; this constitutes an important policy implication of the study.

V.5. Concluding remarks

The results presented in this chapter help to draw two clear conclusions about informal consumption. First, in the case of total expenditure, increasing the total available resources of the family increases formal shares and reduces informal ones. However, inferiority of informal consumption was rejected, even for the richest individuals in the sample, and for the most informal markets. This means that both informal and formal consumption are normal: necessity in the first case and luxury in the second. This result is robust to the different specifications used, the definitions of informality implemented and under the assumption of exogenous and endogenous total expenditure. In fact, once endogeneity is explicitly taken into account in the estimation, the marginal effects proved to be stronger (in absolute terms) than under alternative methods of estimation. Using this more robust result, we can see that any 10% increase in total expenditure will cause a 9.5% increase in informal consumption and a 14.4% in its formal counterpart. The evidence is consistent with the differentiated income effects across the various channels of commercialization and quality differentials. Families are then interested in increasing their consumption from the better outlets as long as they become richer, or in replacing them with lesser-quality ones as they become poorer.

In terms of the policy analysis developed here, the results suggest that informal purchases are used to smooth consumption (as a protective strategy for families) and reveal the positive effects of economic growth on the formalization of the country from a consumption point of view. However, we should acknowledge that, in a highly informal country like Peru, this is a process that comes with limitations. From the simulations presented here, it was possible to verify that informal consumption shares were kept to around 30% even for income levels similar to those of the richest individuals in the sample.

Once heterogeneity across sub-populations was allowed for, it was also possible to verify that the effects of total expenditure were, in fact, different. In general, the re-composition of the consumption basket between formal and informal sources was greater for households located to the right of the expenditure distribution, those who live in the most developed cities and families composed of members working in the modern sector of the economy. In general, here, the gaps between the marginal effects of total expenditure on formal and informal allocation were higher for other households in the sample. Furthermore, in the case of the poorest individuals in the sample, it was possible to verify that small positive shocks to their income will cause a reallocation of resources in favour of informal channels of commercialization. Preferences and limited access to formal options, could be behind this result. The policy consequences of this piece of evidence are the short-term costs of targeted income policies in the form of the higher informal transactions promoted. Also, the evidence suggests that pro-poor growth policies need to be strong enough to change the market allocation of the most deprived.

Second, the linkage between working in and purchasing from informal markets is a strong result for a developing country like Peru. Therefore, we can argue that working more hours in the informal sector has a positive effect on informal and a negative effect on formal consumption. The converse is also applicable (although the evidence was weaker): working more hours in the formal sector has a positive effect on formal and a negative effect on informal consumption. These results were robust to the different specifications used and the definitions of informal markets and under the assumption of exogenous and endogenous total expenditure. As seen for expenditure, the marginal effects computed once endogeneity is explicitly handled during estimation prove to be higher (in absolute terms), making clearer the evidence in favour of the linkage hypothesis. The intuition behind the linkage hypothesis is that previous experience in a sector (as a worker) helps the consumer to reduce transactional costs (network effect) or reveals a preference by the worker to stay there as a consumer (preference or non-separability effect). Point value estimates of the IV application revealed that a 10% increase in weekly informal hours of work increase (reduce) annual informal (formal) consumption by 4% (9%) respectively. In contrast, a 10% increase in formal hours causes both a reduction and an increase of 3% on informal and formal consumption, respectively.

However, we should note that a strong result also encountered in the application is that not all formal or informal hours of work have the same effects on consumption. In fact, the evidence suggests that the main channels for the positive linkage of informal hours of work with informal consumption occur through the informal self-employment sector. At the same time, the positive linkage of formal hours of work with formal consumption occurs through the formal self-employment sector. These effects were accompanied by the negative linkage of the informal wage-earning sector with formal consumption.

This last set of results, overall, suggests that policy-makers are possibly underestimating the effects of their supply-side public policies against informality. For example, reducing hours of work in the informal sector and increasing them in formal jobs will have second-round effects through the expected re-composition of formal and informal demand. Similarly, it is possible to infer that the effectiveness of a policy like this could be greater if the focus of the formalization policy on the supply-side acts mainly through the self-employment segment.

VI. ESTIMATION OF THE DISAGGREGATED MODEL

VI.1. Introduction

In this chapter, the second empirical application of the theoretical model described in Chapter 3 is presented. This corresponds to the estimation of the disaggregated version of the model, which means that the household no longer allocates total expenditure directly to markets. On the contrary, they will first decide on their consumption of particular goods and then decide where these goods will be purchased. To implement this empirically, two household decisions must be estimated, the first related to the consumption of particular goods and the second related to the allocation decision across markets, conditional on goods' expenditure. In the remainder of the chapter, the first decision is referred as the first stage and the second decision as the second stage of the allocation process.

Given the data availability already described on Chapter 4, this model is implemented according to consumption groups, so some aggregation is implicitly still present in the estimation to be discussed here. Although the results could cause some theoretical difficulties concerning the assumptions about the budgeting process discussed, they are considered a good first approximation for the purposes of this research. Specifically, they provide an adequate framework within which to investigate both income and hours effects, but for particular consumption categories, in order to provide a deeper exploration of the results and identify the type of consumption where the linkage hypothesis is more reliable.

The chapter is organized as follows. In Section 2, the empirical specification and econometric issues are discussed. In Section 3, the estimation results and the computed elasticities for the two-stage Engel curves are presented. In Section 4, the main findings are summarized in the form of concluding remarks.

VI.2. Empirical specification and econometric issues

In order to estimate the disaggregated version of the model, seven consumption groups were constructed where sufficient data on purchasing markets existed to identify them. The consumption groups included are food consumed in the household (Fon), food consumed out of the household (Foff), clothes and personal care (CC), health goods and services (HEA), transport and communication (TC), education and culture (ED) and other or non-classified goods (OT). Therefore, following (III.16) and (III.22), in empirical terms the objective is to estimate a demand system composed of seven consumption groups and seven demand sub-systems made up of three markets (formal, semi-formal and informal) within each consumption group.

The empirical specification of the model follows the AIDS demand structure already explained in Chapter 3. However, expression (III.32a)-(III.32b) needs to be slightly modified in order to fit with the two-stage model to be estimated here. Using the demand equation (III.16), the normalization of prices already explained in the previous chapter and considering $A=[h_r, Z]$, the Working–Leser Engel curve to be estimated for each of the k consumption groups in the first stage is

$$w_k = \alpha_k^* + \beta_k \log(M) + \sum_r \tau_{kr} h_r + \psi_k(Z) + e_k \qquad \text{(VI.1)}$$

Based on the estimation of (VI.1), the elasticities of interest to be computed for the first stage of the allocation process can be derived. The expressions take the following form:

$$\eta_k = \frac{\partial c_k}{\partial M}\frac{M}{c_k} = \frac{\beta_k}{\hat{w}_k} + 1 \qquad \text{(VI.2)}$$

$$\mu_{kr} = \frac{\partial c_k}{\partial h_r}\frac{h_r}{c_k} = \tau_{kr}\left(\frac{\hat{h}_r}{\hat{w}_k}\right) \qquad \text{(VI.3)}$$

where (VI.2) represents the income elasticity and (VI.3) the hours' (quantity) effect elasticity for consumption group k. For the second stage, using (III.22) and a similar algebraic manipulation as that used in Chapter 5 to derive the empirical specification (V.1), but considering that m_k replaces M and $A=[h_r, \lambda]$, once prices are normalized to unity, the Working–Leser Engel curve to be estimated for each of the j markets inside each of the k consumption groups is

$$w_{kj} = \alpha_{kj}^* + \beta_{kj} \log(m_k) + \sum_r \tau_{kjr}(h_r) + \psi_{kj}(\lambda) + e_{kj} \qquad \text{(VI.4)}$$

Again, based on the estimation of (VI.4) and following a similar procedure, the elasticities of interest to be computed for the second stage of the allocation process take the form:

$$\eta_{kj}^C = \frac{\partial c_{kj}}{\partial m_k}\frac{m_k}{c_{kj}} = \frac{\beta_{kj}}{\hat{w}_{kj}} + 1 \qquad \text{(VI.5)}$$

$$\mu_{kjr}^{C} = \frac{\partial c_{kj}}{\partial h_r}\frac{h_r}{c_{kj}} = \tau_{kjr}\left(\frac{\hat{h}_r}{\hat{w}_{kj}}\right) \tag{VI.6}$$

where (VI.5) is the income elasticity and (VI.6) the hours' (quantity) effect elasticity for market j conditional on expenditure on group k (this is the reason for the superscript C). To find the unconditional elasticities, Carpentier and Guyomard's (2001) procedure is used, so the final expressions take the form:

$$\eta_{kj}^{U} = \frac{\partial c_{kj}}{\partial M}\frac{M}{c_{kj}} = \left(\frac{\beta_{kj}}{\hat{w}_{kj}}+1\right)\left(\frac{\beta_k}{\hat{w}_k}+1\right) = \eta_{kj}^{C}\eta_k \tag{VI.7}$$

$$\mu_{kjr}^{U} = \frac{\partial c_{kj}^{U}}{\partial h_r}\frac{h_r}{c_{kj}} = \tau_{kjr}\left(\frac{\hat{h}_r}{\hat{w}_{kj}}\right) + \tau_{kr}\left(\frac{\hat{h}_r}{\hat{w}_k}\right)\left(\frac{\beta_{kj}}{\hat{w}_{kj}}+1\right) = \mu_{kjr}^{C} + \mu_{kr}\eta_{kj}^{C} \tag{VI.8}$$

$$\mu_{kjr}^{UT} = \frac{\partial c_{kj}^{UT}}{\partial h_r}\frac{h_r}{c_{kj}} = \mu_{kjr}^{U} + \left(\frac{\hat{h}_r\hat{s}_r}{\hat{M}}\right)\eta_{kj}^{U} \tag{VI.9}$$

where (VI.7) and (VI.8) are unconditional expressions of (VI.5) and (VI.6), respectively identified by the superscript (U). Using both expressions it is possible to derive (VI.9), which is the total effect of hours once total expenditure (M) is allowed to change, giving the increase in labour incomes produced by the hours' effect. The unconditional elasticities are of particular interest in the application since they will help to compare the disaggregated and aggregated versions of the model and to conclude whether formalization policies have a higher or a lower influence on particular consumption groups. As in Chapter 5, all these expressions are computable using sample means. However, considering that, in empirical terms, hours are introduced in weekly terms and expenditure in annual terms, as done previously the term $\hat{h}_r\hat{s}_r$ in (VI.9) needs to be multiplied by 48.

In estimating this model, some econometric issues emerge. Most – cross-equation correlations, heteroskedasticity and simultaneity – have already been explained in Chapter 5. To handle the first two, in the previous chapter a SUR–GLS model was proposed and, for the third, IV strategies were developed. However, beyond these problems, in the particular case of the disaggregated model the relatively high importance of zero-expenditures identified in Chapter 4 for most of the consumption categories to be modelled generates good reasons to consider the emergence of censored bias. To solve this, different techniques are available: Tobit estimators (see Tobin, 1958; Amemiya, 1984) and Heckman (1979) two-step procedures. As pointed out by Niimi (2007) in a similar context, using a Tobit model will imply two restrictive assumptions (see also Lamb, 1989; Gould, 1992; Blisard and Blaylock, 1993). First, as discussed in Greene (2003), it assumes that the decision-making processes of choosing to purchase or not to purchase the good and deciding the consumption level are the same. Second, Tobit models will only take into account 'true corner solutions', i.e., those emerging only because of economic factors (if prices are too high or incomes too low). When these assumptions were tested (see Greene, 2003), evidence against the Tobit model was found, so the Heckman procedures prove to be better.[55]

However, a standard Heckman procedure for separate equations is not an adequate solution for systems with different levels of censoring on the equations within them. Therefore, more general methods using the whole sample are needed. The solution adopted here is the two-step estimation method of Shonkwiller and Yen (1999), previously used in Monge (2007). Here, as in the Heckman procedure, the first stage is a Probit model, estimated for the decision to purchase or not to purchase. A general and simplified expression for it is

$$P(w > 0) = \Phi(\pi_0 Q_1) \qquad \text{(VI.10)}$$

where Q_1 represents the different covariates assumed to influence the purchasing decision and π_0 the set of parameters associated with them. Once the estimation of the first stage is complete, the probability density function $\phi(\pi_0 Q_1)$ and the cumulative density function $\Phi(\pi_0 Q_1)$ are computed. Using these expressions, the corrected second-stage equation will take the form

$$w = \Phi(\pi_0 Q_1) G(\pi_1 Q_2) + \pi_2 \phi(\pi_0 Q_1) + u \qquad \text{(VI.11)}$$

where $G(\pi_1 Q_2)$ represents the deterministic part, with Q_2 defining the covariates influencing the purchasing levels and π_1 the set of parameters associated with them; π_2 is a coefficient; and u is disturbance with $u \sim N(0, \sigma^2)$. Expression (VI.11) is the unconditional mean expression and can be estimated consistently by a standard SUR technique with all the sample values (zero and positive).

[55] Chi-squared results obtained for the different regression models used here were between 200 and 4,500 above the critical value of 40.11.

This procedure is used for both the first and the second stages of the decision process to be estimated here so, in empirical terms, the estimation is composed of the following expressions:

$$P(w_k > 0) = \Phi(\pi_{0k} Q_k) \tag{VI.12}$$

$$w_k = \Phi(\pi_{0k} Q_k)\left[\alpha_k^* + \beta_k \log(M) + \sum_r \tau_{kr} h_r + \psi_k(Z)\right] + \pi_{2k}\phi(\pi_{0k} Q_k) + \varepsilon_k \tag{VI.13}$$

$$P(w_{kj} > 0) = \Phi(\pi_{0kj} Q_{kj}) \tag{VI.14}$$

$$w_{kj} = \Phi(\pi_{0kj} Q_{kj})\left[\alpha_{kj}^* + \beta_{kj} \log(m_k) + \sum_r \tau_{kjr}(h_r) + \psi_{kj}(\lambda)\right] + \pi_{2kj}\phi(\pi_{0kj} Q_{kj}) + \varepsilon_{kj} \tag{VI.15}$$

where (VI.12) is a PROBIT model used to estimate the decision to purchase or not to purchase for a particular consumption group; (VI.13) is a SUR–GLS model used to estimate the share purchased of each consumption group; (VI.14) is, again, a PROBIT model used to estimate the decision to purchase or not to purchase from formal or informal markets once the household has decided to purchase a particular consumption group; and (VI.15) is, again, a SUR–GLS model used to estimate how much of a particular consumption group is allocated to formal or informal markets.

This method is subsequently referred to as the S&Y and its estimation needs to be undertaken with caution, bearing in mind three complications that arise. First, the introduction of correctional terms in the system modifies the basic insights of the SUR specification mentioned in Chapter 5. Basically, the matrix of the variables included on the right-hand side of the regression models is no longer the same, so efficiency losses of not estimating (VI.13) and (VI.15) as systems will be greater and parameter estimates will no longer be similar to independent OLS estimations. Second, although dropping a category is still the correct way to accommodate the model for adding-up (as in Yen *et al.*, 2003; Yen and Lin, 2006), if we consider the arguments of Yen and Chen (2002), Chen and Chen (2002) and Caillavet (2005), the results will not be invariant to the category dropped so interpretation of results must be undertaken with caution and must take into account the fact that they are conditional on the deletion decision adopted. In this application, the first stage drops the ‘others’ category (OT) and the second the semi-formal market. Third, as in the Heckman procedure, the standard errors will be heteroskedastic and the covariance matrix may be incorrect given the introduction of predicted terms, so robust techniques need to be implemented. In this application[56] the Probit and SUR parts of the model are estimated

[56] Other solutions explored were using the Murphy and Topel (1985) variance–covariance matrix or bootstrapping standard errors for the second stage. The results were no different to those to be presented here, but they impede to recover estimates of the first stage and implement Weesie’s (1999) method for

considering the White/Huber correction and the between-model covariance matrices are estimated using Weesie's (1999) procedure.

To derive marginal effects when the S&Y method is estimated, both the first part (Probit results) and second part (SUR results) must be taken into account. The procedure is explained in Su and Yen (2000). Using (VI.11) as the general form of the model, the correct formulae take the form:

$$\frac{\partial w}{\partial x} = \Phi \pi_{1,x} + \phi \pi_{0,x} [G(\pi_1 Q_2) - \pi_2 (\pi_0 Q_1)] \qquad \text{(VI.16a)}$$

or

$$\frac{\partial w}{\partial x} = \pi_{1,x} \Theta = \pi_{1,x} \left[\Phi + \phi(\pi_{0,x} / \pi_{1,x})[G(\pi_1 Q_2) - \pi_2 (\pi_0 Q_1)] \right] \qquad \text{(VI.16b)}$$

This means that, for the elasticities of interest in this application, previous expressions are modified multiplying by Θ each of the parameters of interest. Note that each of the terms can be recovered from previous estimations or set at the sample means, so they are fully identifiable. However, we should also note that, for unconditional elasticities, the results for both the first stage (consumption groups) and second stage (markets inside groups) need to be combined. This is done via the non-linear combination of paramenter; the joint vce is computed using the Delta method.

In sum, the models to be estimated in this chapter follow the S&Y procedure that takes into account most of the econometric problems described previously, except for the endogeneity of total expenditure and labour supply. Therefore, more than causal effects, in this chapter the interpretation of results is undertaken in terms of the correlations, with parameter estimates that may be downward-biased, as discussed in Chapter 5. For this reason, for comparative purposes between the aggregate and disaggregated models, this chapter will rely on the first empirical exercise of the previous chapter, which shares similar methodological issues. More robust methods of estimation, like the implementation of IV models using S&Y procedures, are consigned to further research.

joint vce of the two parts of the model so, following this, a general White/Huber correction method was considered to be a good solution for inference purposes.

The specification of the demand equations to be estimated here is the same as that used in the previous chapter. This means that the definition of income used is the natural logarithm of annual *per capita* expenditure (for the first-stage model) and the natural logarithm of annual *per capita* expenditure on each consumption group (for the second-stage model). The hours of work of the seven earning opportunities and their participation dummies are also included in the models of each stage. The other variables included follow specification (1) of the previous chapter. This specification[57] is maintained at both the first stage (decision on goods) and the second stage (decision on markets). The exact definition of the variables used in this application and the basic tabulations employing the fully usable estimating sample are presented in Appendix 4.

The Probit models that supplement the SUR models include the same variables as before, plus additional identifying covariates assumed to be related to the probability of purchasing, but not with the amount allocated. In general, the additional variables for the first stage (consumption groups) were proxies for the supply conditions measured at the district level and proxies of consumption adequacy at the household level. The identifying variables were different for each of the consumption groups[58] modelled. In the second stage (markets), each of previous variables was used for its relevant sub-system plus an additional set of variables (common to all the sub-systems) oriented to identify the decision on market participation. The variables were characteristic of the district where the household lives (considering that more vulnerable districts have a higher probability of favoring the emergence of informal outlets), of the access conditions (like distance and quality of roads), of some subjective well-being indicators (as a crude proxy of stigma considering that those households which consider themselves as poor are more prone to purchasing informal goods), and some indicators of confidence in the government (as a crude proxy of tax morale).

In general, global adjustment of the PROBIT models, Wald tests of global significance for the identification covariates, individual significance and the correctness of the signs in most cases give support to the adequacy of the correctional terms. Estimates were also performed for different

[57] Specifications (2) to (4) were also explored but, since the results show minimal changes to those presented here, they are omitted to save space.

[58] Fon (proxies of consumption adequacy on buying food are used as well as proxies of the incidence of the agricultural, fishery and livestock sectors in the district), Foff (existence and number of restaurants or other places that sell food and the incidence of the restaurant sector in the economy of the district), CC (consumption adequacy for clothing and the incidence of the commercial sector in the district), HEA (consumption adequacy for health, proxies of the health condition of the family, the existence and number of health-service providers and the incidence of the health sector in the district), TC (consumption adequacy for transport, existence and number of transport and communication providers in the district, proxies of transport facilities as well as infrastructure and the incidence of the transport and communication sector in the district), ED (proxies of consumption adequacy on the consumption of education, proxies of number and incidence of education providers and the incidence of the educational sector in the district).

combinations of the identification variables used here as well as other candidates, without any material changes to the conclusions presented later.[59] Therefore, it was possible to conclude that the results were not sensitive to the identification assumptions made. As a result, without theoretical guidance on the identification variables that must be used and with a lack of better data (for example to construct better proxies of stigma, tax morale or perceived risk), the exercises performed give enough confidence in the strategy, at least for the purely correctional perspective followed here.

VI.3. The results

In this section the influence of income and hours of work is presented and discussed for each part of the decision process of the individual: the decision on which goods to purchase (first stage) and on where these good will be bought (second stage). Similarly, the joint estimates (first and second stage) for unconditional elasticities are derived and compared with the aggregate model.

First stage: the decision on which goods to purchase

In Table VI.1, the PROBIT and the SUR coefficients of the S&Y estimator applied on consumption groups are presented, with an estimation sample of 7,253 observations (once the dataset is cleaned of missing values). To save space, the focus is put only on the results for expenditure and hours of work. The full set of regression coefficients are presented in Appendix 5. Although the PROBIT and SUR results are not of particular interest (they are just intermediate estimates used to derive marginal effects), it is important to note, for the selected parameters, that the coefficients are, in general, well-determined (significant and with the correct signs) in both cases. We can also see that the Probit estimates shown here follow, in general, the same direction as the SUR estimates, with the exception of food in the household and food out of the household estimates. In these two cases, income increases the probability of purchasing these goods, but reduces the latent budget shares. Finally, the selection term indicates, in general, that the unobservables in the model have a positive influence on consumption with the exception of food consumed outside the house and of health.

In Table VI.2 the marginal effects and elasticities are presented. As in the previous chapter, the marginal effects are computed for the budget shares and elasticities for the consumption level. In terms of total expenditure, we can see that additional resources enjoyed by the family reduce the share of food in the household (Fon), have no effect on food consumed outside the house (Foff)

[59] These preliminary results are not presented in order to save space.

and increase clothing and personal care (CC), health goods and services (HEA), transport and communication (TC) and education and culture expenditure (ED). Thus, in terms of the elasticities computed and considering Deaton and Muellbauer's (1980a) discussion, Fon could be classified as a normal necessity (elasticity below 1), Foff as unitary (elasticity not significantly different to 1) and the other consumption groups as luxuries (elasticities above 1). This is a standard result in the literature on demand estimation in developing countries and particularly in Peru (see Monge, 2007). Point estimates reveal that a 10% increase in total expenditure will have the lowest effect on food to be consumed in the house (7.7% expansion) and the highest on education and culture (almost 14%).

Table VI.1: Probit and SUR coefficients, decision on consumption groups

	Probit estimates					
	Fon	**Foff**	**CC**	**HEA**	**TC**	**ED**
L(expenditure)	0.52061***	0.69847***	0.90564***	0.52624***	0.89699***	0.44086***
	(0.08090)	*(0.04344)*	*(0.11204)*	*(0.03819)*	*(0.06636)*	*(0.04926)*
Hrs_ff	-0.00468	0.00394**	0.00623	-0.00089	-0.00215	-0.00451**
	(0.00295)	*(0.00173)*	*(0.00774)*	*(0.00122)*	*(0.00489)*	*(0.00201)*
Hrs_if	-0.01237***	0.01248***	n.a.	-0.00174	-0.00368	-0.00834***
	(0.00314)	*(0.00351)*	*(n.a.)*	*(0.00203)*	*(0.00354)*	*(0.00244)*
Hrs_ii	-0.00456***	0.00513***	0.00604	-0.00140	0.00393*	-0.00177
	(0.00160)	*(0.00146)*	*(0.00478)*	*(0.00102)*	*(0.00218)*	*(0.00124)*
Hrs_sf	-0.00667	-0.00223	-0.01281***	-0.00195	-0.00538	0.00157
	(0.00497)	*(0.00221)*	*(0.00420)*	*(0.00201)*	*(0.00335)*	*(0.00262)*
Hrs_si	-0.00422**	0.00549***	-0.00174	0.00121	0.00456***	-0.00191*
	(0.00198)	*(0.00102)*	*(0.00301)*	*(0.00076)*	*(0.00137)*	*(0.00102)*
Observations	7,253	7,253	7,253	7,253	7,253	7,253
Pseudo LogL	-404.675	-2220.828	-211.496	-2807.751	-764.385	-1400.052
Chi2 (hours)	29.04***	56.08***	11.76**	6.54	18.35***	22.51***
	SUR estimates					
	Fon	**Foff**	**CC**	**HEA**	**TC**	**ED**
L(expenditure)	-0.08205***	-0.02567***	0.00796***	0.00515*	0.03033***	0.02868***
	(0.00434)	*(0.00416)*	*(0.00170)*	*(0.00303)*	*(0.00277)*	*(0.00235)*
Hrs_ff	-0.00052***	0.00069***	0.00001	-0.00001	0.00013**	-0.00017***
	(0.00007)	*(0.00007)*	*(0.00004)*	*(0.00005)*	*(0.00005)*	*(0.00006)*
Hrs_if	-0.00077***	0.00093***	0.00009	-0.00005	0.00017*	-0.00019*
	(0.00014)	*(0.00013)*	*(0.00009)*	*(0.00008)*	*(0.00010)*	*(0.00010)*
Hrs_ii	-0.00064***	0.00100***	0.00005	0.00002	0.00010**	-0.00035***
	(0.00007)	*(0.00007)*	*(0.00003)*	*(0.00004)*	*(0.00004)*	*(0.00003)*
Hrs_sf	-0.00068***	0.00011	-0.00005	0.00015	0.00013	-0.00001
	(0.00016)	*(0.00014)*	*(0.00008)*	*(0.00010)*	*(0.00011)*	*(0.00010)*
Hrs_si	-0.00061***	0.00057***	0.00014***	-0.00004	0.00001	-0.00009***
	(0.00006)	*(0.00006)*	*(0.00003)*	*(0.00003)*	*(0.00004)*	*(0.00003)*
PHI	0.20134***	-0.03038*	0.10260***	-0.03036***	0.03529**	0.01777*
	(0.05094)	*(0.01556)*	*(0.02660)*	*(0.01102)*	*(0.01550)*	*(0.01014)*
Observations	7,253					
Pseudo LogL	48858.641					
Chi2 (hours)	249.16***	374.61***	27.46***	4.14	14.99**	121.84***
Chi2 (hours), system	662.36***					

Note: Robust standard errors in parenthesis with Weesie (1999) corrections for between-model (PROBIT and SUR) vce. *** p<0.01, ** p<0.05, * p<0.1.

In the case of hours of work, differentiated effects are also found. It is therefore better to organize the discussion for each consumption group. In the case of Fon, all the hours of work considered in the estimation reduce consumption and, when formally tested, the point estimates prove to be similar.[60] In Pollak's (1969) terms, this means that labour and Fon consumption can be classified as quasi-substitutes. Therefore, despite the sector to which the household allocates its labour, hours of work have a negative influence on food consumption, possibly revealing the negative association between working and purchasing activities. One additional hour spent by the family will generate a decrease of between 0.05 and 0.07 of a percentage point in the budget share. Similar results are shown for ED. This result can be interpreted again in terms of the substitution between working and purchasing activities, mainly if these activities must be done during free-time. The average point estimate reveals that one additional working hour reduces ED consumption by 0.08 of one percentage point, with all the earning opportunities with statistically similar effects[61] except for Hrs_ii.

Table VI.2: Marginal effects and elasticities, decision on consumption groups

	Marginal effect					
	Fon	**Foff**	**CC**	**HEA**	**TC**	**ED**
L(expenditure)	-0.08360***	-0.00014	0.00436***	0.01603***	0.03159***	0.02755***
	(0.00444)	*(0.00365)*	*(0.00161)*	*(0.00245)*	*(0.00250)*	*(0.00197)*
Hrs_ff	-0.00049***	0.00072***	-0.00002	-0.00003	0.00012**	-0.00017***
	(0.00007)	*(0.00009)*	*(0.00005)*	*(0.00005)*	*(0.00005)*	*(0.00005)*
Hrs_if	-0.00069***	0.00120***	0.00009	-0.00008	0.00016*	-0.00020**
	(0.00014)	*(0.00017)*	*(0.00009)*	*(0.00009)*	*(0.00009)*	*(0.00009)*
Hrs_ii	-0.00060***	0.00103***	0.00002	-0.00002	0.00011**	-0.00032***
	(0.00007)	*(0.00008)*	*(0.00004)*	*(0.00005)*	*(0.00004)*	*(0.00003)*
Hrs_sf	-0.00064***	0.00003	0.00000	0.00008	0.00011	0.00000
	(0.00016)	*(0.00016)*	*(0.00008)*	*(0.00010)*	*(0.00011)*	*(0.00009)*
Hrs_si	-0.00058***	0.00066***	0.00015***	0.00000	0.00002	-0.00009***
	(0.00006)	*(0.00007)*	*(0.00003)*	*(0.00004)*	*(0.00004)*	*(0.00003)*
	Elasticities					
	Fon	**Foff**	**CC**	**HEA**	**TC**	**ED**
Expenditure	0.770***	0.999	1.049***	1.309***	1.231***	1.408***
	(0.01222)	*(0.03041)*	*(0.01820)*	*(0.04734)*	*(0.01833)*	*(0.02911)*
Hrs_ff	-0.028***	0.125***	-0.004	-0.012	0.018**	-0.051***
	(0.00412)	*(0.01571)*	*(0.01138)*	*(0.02096)*	*(0.00828)*	*(0.01653)*
Hrs_if	-0.010***	0.054***	0.006	-0.008	0.006*	-0.016**
	(0.00206)	*(0.00785)*	*(0.00540)*	*(0.00918)*	*(0.00374)*	*(0.00702)*
Hrs_ii	-0.031***	0.159***	0.005	-0.006	0.015**	-0.088***
	(0.00339)	*(0.01312)*	*(0.00775)*	*(0.01646)*	*(0.00578)*	*(0.00824)*
Hrs_sf	-0.014***	0.002	0.000	0.012	0.006	0.000
	(0.00340)	*(0.01037)*	*(0.00695)*	*(0.01573)*	*(0.00620)*	*(0.01047)*
Hrs_si	-0.047***	0.162***	0.049***	-0.003	0.004	-0.037***
	(0.00478)	*(0.01683)*	*(0.01007)*	*(0.02034)*	*(0.00770)*	*(0.01193)*

Note: Robust standard errors in parenthesis with Weesie (1999) corrections for between-model (PROBIT and SUR) vce. *** p<0.01, ** p<0.05, * p<0.1. In the case of expenditure elasticities, significance tests are of the form: different to 1.

The situation of CC and HEA is different. These two consumption groups seem to be less sensitive to the hours the family spend at work. Despite the positive effect of Hrs_si on CC in all other cases, hours of work have no influence on demand patterns. Again using Pollak's (1969) quantity

[60] Chi-squared results: 1.87 (ff vs if), 1.50 (ff vs ii), 0.78 (ff vs sf), 1.11 (ff vs si), 0.36 (if vs ii), 0.07 (if vs sf), 0.56 (if vs si), 0.04 (ii vs sf), 0.06 (ii vs si), 0.11 (sf vs si).

[61] Chi-squared results: 0.11 (ff vs if), 6.66*** (ff vs ii), 1.96 (ff vs si), 1.78 (if vs ii), 1.56 (if vs si), 34.47*** (ii vs si).

effects interpretation, it seems that clothing and personal care, and health goods and services can be classified as neutral with respect to leisure. The result for CC is difficult to interpret but, for HEA, it is pretty natural to assume that the health condition of the family and not necessarily working more or less hours will be correlated with the propensity to spend on health.

At the other extreme, Foff and TC seem to be quantity-complements with labour supply. When significant, hours of work are positively correlated to demand allocation in these two groups. This is, again, an intuitive result considering the close relation of these types of good on the working activities of the family: meals at work in the case of Foff and travel to the workplace in the case of TC. However, differences between earning opportunities should be noted. For example, for Foff, when formally tested, the effects are dissimilar between the earning opportunities[62]. The results suggest, in general, that effects of the wage-earning sector are higher than those of the self-employment sector. One additional working hour in the wage-earning sector will increase Foff's budget share by on average about 0.10 of one percentage point and one additional working hour in the self-employment sector will increase Foff's budget share on average by around 0.03 of one percentage point. This result is revealing higher pronpensity to buy food near their workplaces for wage-earners. Some possible explanations could be less-flexible work schedules for these types of worker or better facilities of this type of consumption in modern firms (in the case of Hrs_ff). By contrast, the self-employed, with more-flexible schedules or working near their homes, are more likely to return home to eat.

Among the wage-earners, informal workers show higher values than formal ones. One additional hour as a formal wage-earner will increase Foff's budget share by 0.07 of one percentage point while one additional hour as an informal wage-earner increases budget shares by 0.10 of one percentage point. This result is probably associated with even more inflexible schedules for informal wage-earners. A similar situation is observed with the self-employed, where the informal segment shows a marginal effect not only higher than its formal counterpart, but also statistically equal to zero.

The results for TC show a similar pattern. In this case, we find that working more hours in the wage-earning sector increases the consumption of transportation, but working more hours in the self-employment sector does not. These results can again be interpreted in terms of the less-flexible work schedules of wage-earners or their location decisions. For example, as long as people working in the wage-earning sector need to travel to their jobs and the self-employed do

---

[62] Chi-squared results: 6.22** (ff vs if), 6.18** (ff vs ii), 14.99*** (ff vs sf), 0.29 (ff vs si), 0.87 (if vs ii), 24.23*** (if vs sf), 8.51*** (if vs si), 31.25*** (ii vs sf), 11.46*** (ii vs si), 13.61*** (sf vs si).

not (maybe because they work closer to home), then one additional hour spent in the first case (maybe because of an additional member also working there) will increase transport expenditure, though this is not necessarily so for the second case. In terms of the point estimates, there are no differences between the formal and the informal segments of the wage-earning sector.[63]

In sum, the first-stage results help us to arrive at three preliminary conclusions. First, consumption groups can be classified in three types: necessities (Fon), unitary (Foff) and luxuries (CC, HEA, TC and ED). Second, the estimations also permit an ordering between those groups which are quantity-neutral for leisure (CC and HEA), quantity-complements (Fon and ED) and quantity-substitutes (Foff and TC). Third, differentiated effects of labour supply are found only in those consumption groups classified as substitutes for leisure. In this case, working an additional hour in the wage-earning sector generally has a greater effect than working it in the self-employment sector.

Second stage: the decision on where the goods will be bought

Here the results of the second stage of the model are presented and discussed. The interest here is to analyse, once the individual has decided to purchase a particular group of goods, where these goods are going to be bought (on informal or formal markets). The analysis is conditional on the decisions made at the first stage, so demand responses identified are derived using group expenditure. In Tables VI.3 and VI.4, the PROBIT and SUR results, respectively, are presented. To save space, the focus is again only on the results for expenditure and hours of work, with the full set of regression results presented in Appendix 5.

As already mentioned, the PROBIT and SUR coefficients are just intermediate results used to construct the marginal effects; however, the PROBIT model could provide interesting information about the differentiated behaviour of the linkage hypothesis at the zero-consumption limit. In contrast to what occurred at the previous stage, here parameters seem to differ (both in significance and in sign) in both parts of the estimation. As expected, the SUR coefficients are generally in line with the marginal effects constructed, with PROBIT estimates showing major differences mainly in terms of the labour supply variables. Therefore, it should be acknowledged that the validity of the linkage hypothesis will be different not only between consumption groups, but also in terms of the choice decisions about a particular good. The subsequent analysis will concentrate on the marginal effects computed in Table VI.5 and the elasticities reported in Table VI.6; however, these differences will be reported when relevant. Also, it is worth noting that, in

[63] Chi-squared results: 0.16 (ff vs if), 0.01 (ff vs ii), 0.23 (if vs ii).

contrast to the previous stage, the correctional terms are not always significant, but when significant, they yield the expected positive sign in four out of eight cases.

In terms of group expenditure, the marginal effect on Fon is positive and statistically similar[64] for both informal and formal shares. Therefore we could argue, for Peruvian consumers, that, when the family decides to purchase a greater amount of food, they obtain it through both formal and informal markets. However, once elasticities are computed, it is possible to verify that formal markets will grow faster[65] than informal ones. This is an intuitive result considering the discussion of the previous chapter, where higher total expenditure was correlated with more formal purchasing baskets. Similar results are shown for Foff, TC and ED, but with different point estimates. For Foff, increasing the allocation of resources in this consumption group reduces both formal and informal market shares, but the decline observed in the informal market is greater.[66] With this result, although the elasticities are below 1, it is possible to verify that formal markets are still growing faster[67] than informal ones. In TC and ED, the results are clearer and similar to those from the aggregate model. In both cases, marginal effects on informal consumption are negative with positive effects on its formal counterpart. Therefore, in terms of elasticities, a higher allocation of resources in these groups will generate increases that are more than proportional in formal and less than proportional in informal options.

The case for CC and HEA is different. In the latter, group expenditure has a positive effect only on formal shares, with informal ones being practically invariant. This result may be in line with previous ones but, when elasticities are computed, they prove to be statistically similar,[68] so it is preferable to conclude that, in the case of HEA, both formal and informal markets grow at the same speed for any increase in group expenditure. A similar outcome applies for CC. Here group expenditure increases only informal shares but, once the elasticities are computed, they are not statistically different.[69]

[64] Chi-squared results: 0.72
[65] Chi-squared results: 14.26***
[66] Chi-squared results: 22.80***
[67] Chi-squared results: 51.57***
[68] Chi-squared results: 0.05
[69] Chi-squared results: 1.37

Table VI.3: Probit coefficients, second-stage estimations

	Fon		Foff		CC		HEA		TC		ED	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expenditure)	0.57678***	0.54711***	0.03803**	0.19543***	0.40686***	0.36147***	0.21079***	0.50681***	0.37532***	0.77407***	0.04890***	0.73460***
	(0.04131)	*(0.03205)*	*(0.01529)*	*(0.01682)*	*(0.02279)*	*(0.02160)*	*(0.01452)*	*(0.02024)*	*(0.02412)*	*(0.02813)*	*(0.01629)*	*(0.03922)*
Hrs_ff	0.00065	0.00221**	0.00217**	0.00363***	-0.00157	0.00147	-0.00003	-0.00114	0.00154	0.00017	-0.00020	-0.00491**
	(0.00249)	*(0.00094)*	*(0.00096)*	*(0.00132)*	*(0.00111)*	*(0.00090)*	*(0.00102)*	*(0.00155)*	*(0.00156)*	*(0.00176)*	*(0.00101)*	*(0.00237)*
Hrs_if	-0.00296	-0.00152	0.00129	0.00065	-0.00294	-0.00130	-0.00075	0.00138	-0.00075	-0.00083	0.00329*	-0.00673**
	(0.00322)	*(0.00167)*	*(0.00177)*	*(0.00184)*	*(0.00199)*	*(0.00186)*	*(0.00191)*	*(0.00235)*	*(0.00250)*	*(0.00226)*	*(0.00174)*	*(0.00295)*
Hrs_ii	-0.00076	-0.00140*	0.00132*	-0.00176**	-0.00015	-0.00069	0.00015	0.00083	0.00263**	-0.00211**	-0.00098	-0.00348**
	(0.00164)	*(0.00073)*	*(0.00077)*	*(0.00077)*	*(0.00097)*	*(0.00077)*	*(0.00084)*	*(0.00122)*	*(0.00102)*	*(0.00090)*	*(0.00079)*	*(0.00147)*
Hrs_sf	0.00472	0.00164	-0.00041	0.00176	-0.00303*	-0.00103	-0.00147	0.00505*	-0.00536***	0.00102	0.00371**	0.00253
	(0.00487)	*(0.00159)*	*(0.00166)*	*(0.00201)*	*(0.00173)*	*(0.00157)*	*(0.00167)*	*(0.00270)*	*(0.00194)*	*(0.00255)*	*(0.00180)*	*(0.00316)*
Hrs_si	0.00174	0.00104*	0.00338***	-0.00116*	0.00043	-0.00044	0.00054	0.00047	-0.00170**	-0.00012	0.00284***	-0.00318***
	(0.00162)	*(0.00060)*	*(0.00067)*	*(0.00063)*	*(0.00078)*	*(0.00062)*	*(0.00067)*	*(0.00094)*	*(0.00082)*	*(0.00078)*	*(0.00067)*	*(0.00115)*
Observations	7,145	7,145	6,272	6,272	7,181	7,181	6,021	6,021	6,967	6,967	6,570	6,570
Pseudo LogL	-626.346	-4082.925	-3843.844	-2999.566	-2800.179	-3659.533	-2753.117	-1480.547	-1895.040	-1964.155	-3378.221	-897.204
Chi2 (hours)	3.57	14.43**	31.06***	18.17***	7.57	5.15	1.65	5.40	19.57***	5.79	27.41***	20.69***

Note: Robust standard errors in parenthesis with Weesie (1999) corrections for between-model (PROBIT and SUR) vce. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. L(expenditure) corresponds to group expenditure in each case.

Table VI.4: SUR coefficients, second-stage estimations

	Fon		Foff		CC		HEA		TC		ED	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expenditure)	-0.01111	0.01161	-0.13601***	-0.07062***	-0.08257***	-0.11007***	-0.07617***	0.00938	-0.07099***	0.13331***	-0.14352***	0.14999***
	(0.00720)	*(0.01386)*	*(0.00659)*	*(0.00701)*	*(0.00656)*	*(0.01116)*	*(0.01278)*	*(0.00610)*	*(0.00679)*	*(0.00652)*	*(0.00426)*	*(0.00352)*
Hrs_ff	0.00025	0.00002	0.00016	-0.00009	-0.00018	0.00016	-0.00033	-0.00017	0.00045**	-0.00033*	0.00054***	-0.00022
	(0.00017)	*(0.00018)*	*(0.00032)*	*(0.00027)*	*(0.00023)*	*(0.00023)*	*(0.00049)*	*(0.00023)*	*(0.00018)*	*(0.00017)*	*(0.00017)*	*(0.00017)*
Hrs_if	0.00047	0.00056*	0.00088	-0.00091*	0.00012	-0.00031	0.00051	0.00026	0.00079**	-0.00073**	0.00039	-0.00039
	(0.00033)	*(0.00029)*	*(0.00055)*	*(0.00051)*	*(0.00047)*	*(0.00054)*	*(0.00081)*	*(0.00041)*	*(0.00033)*	*(0.00031)*	*(0.00037)*	*(0.00034)*
Hrs_ii	-0.00001	-0.00001	0.00036	-0.00151***	-0.00021	-0.00006	-0.00054	0.00026	0.00089***	-0.00080***	0.00050**	-0.00048***
	(0.00015)	*(0.00014)*	*(0.00024)*	*(0.00026)*	*(0.00018)*	*(0.00023)*	*(0.00037)*	*(0.00017)*	*(0.00016)*	*(0.00016)*	*(0.00020)*	*(0.00015)*
Hrs_sf	0.00165***	0.00004	0.00022	0.00095*	0.00001	0.00005	0.00020	0.00034	-0.00095***	0.00058*	0.00040	0.00014
	(0.00031)	*(0.00032)*	*(0.00063)*	*(0.00049)*	*(0.00043)*	*(0.00043)*	*(0.00090)*	*(0.00035)*	*(0.00033)*	*(0.00031)*	*(0.00032)*	*(0.00029)*
Hrs_si	0.00044***	-0.00011	0.00113***	-0.00009	0.00004	-0.00009	0.00004	0.00001	0.00014	-0.00012	0.00051***	-0.00060***
	(0.00011)	*(0.00011)*	*(0.00021)*	*(0.00021)*	*(0.00015)*	*(0.00019)*	*(0.00033)*	*(0.00014)*	*(0.00015)*	*(0.00015)*	*(0.00014)*	*(0.00012)*
PHI	-0.22651***	-0.04543*	-0.04389	-0.05292	-0.41433***	-0.07401**	0.04710	0.16085***	0.44836***	0.56496***	-0.00381	0.39844***
	(0.06863)	*(0.02421)*	*(0.04509)*	*(0.05273)*	*(0.03763)*	*(0.03021)*	*(0.05461)*	*(0.06201)*	*(0.02923)*	*(0.02370)*	*(0.03288)*	*(0.03602)*
Observations	7,145		6,272		7,181		6,021		6,967		6,570	
Pseudo LogL	4043.816		-3196.704		2225.751		309.848		2056.629		29.746	
Chi2 (hours)	48.29***	4.63	31.92***	38.77***	2.09	0.98	3.27	4.42	50.90***	35.99***	28.79***	37.18***
Chi2 (hours), system	61.38***		92.75***		2.87		10.12		56.21***		50.32***	

Note: Robust standard errors in parenthesis with Weesie (1999) corrections for between-model (PROBIT and SUR) vce. *** p<0.01, ** p<0.05, * p<0.1. L(expenditure) corresponds to group expenditure in each case.

Table VI.5: Marginal effects, second-stage estimations

	Fon		Foff		CC		HEA		TC		ED	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expenditure)	0.02057***	0.03076***	-0.07314***	-0.01488**	0.02697***	-0.00261	0.00556	0.04824***	-0.07576***	0.05882***	-0.10224***	0.10779***
	(0.00774)	*(0.00729)*	*(0.00603)*	*(0.00737)*	*(0.00857)*	*(0.00563)*	*(0.00509)*	*(0.00529)*	*(0.00568)*	*(0.00617)*	*(0.00440)*	*(0.00422)*
Hrs_ff	0.00028	0.00011	0.00048	0.00066	-0.00052	0.00017	-0.00007	-0.00024	0.00035**	-0.00029*	0.00039*	-0.00004
	(0.00023)	*(0.00010)*	*(0.00031)*	*(0.00041)*	*(0.00039)*	*(0.00013)*	*(0.00017)*	*(0.00029)*	*(0.00015)*	*(0.00016)*	*(0.00021)*	*(0.00015)*
Hrs_if	0.00030	0.00019	0.00075	-0.00057	-0.00059	-0.00020	0.00003	0.00034	0.00073**	-0.00056**	0.00066	-0.00012
	(0.00040)	*(0.00018)*	*(0.00056)*	*(0.00066)*	*(0.00072)*	*(0.00027)*	*(0.00032)*	*(0.00050)*	*(0.00028)*	*(0.00025)*	*(0.00041)*	*(0.00028)*
Hrs_ii	-0.00005	-0.00007	0.00045*	-0.00151***	-0.00021	-0.00008	-0.00009	0.00030	0.00070***	-0.00053***	0.00026	-0.00031**
	(0.00018)	*(0.00009)*	*(0.00024)*	*(0.00031)*	*(0.00032)*	*(0.00012)*	*(0.00014)*	*(0.00023)*	*(0.00013)*	*(0.00013)*	*(0.00021)*	*(0.00012)*
Hrs_sf	0.00186***	0.00010	0.00006	0.00108	-0.00070	-0.00007	-0.00010	0.00070	-0.00066**	0.00042	0.00071*	0.00004
	(0.00047)	*(0.00019)*	*(0.00059)*	*(0.00071)*	*(0.00065)*	*(0.00023)*	*(0.00031)*	*(0.00046)*	*(0.00028)*	*(0.00026)*	*(0.00038)*	*(0.00025)*
Hrs_si	0.00052***	0.00000	0.00126***	-0.00030	0.00014	-0.00006	0.00006	0.00004	0.00018	-0.00009	0.00070***	-0.00042***
	(0.00016)	*(0.00007)*	*(0.00021)*	*(0.00026)*	*(0.00027)*	*(0.00010)*	*(0.00012)*	*(0.00018)*	*(0.00012)*	*(0.00011)*	*(0.00016)*	*(0.00010)*

Note: Robust standard errors in parenthesis with Weesie (1999) corrections for between-model (PROBIT and SUR) vce. *** p<0.01, ** p<0.05, * p<0.1. L(expenditure) corresponds to group expenditure in each case.

Table VI.6: Elasticities, second-stage estimations

	Fon		Foff		CC		HEA		TC		ED	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
Expenditure	1.034***	1.474***	0.738***	0.972**	1.065***	0.962	1.079	1.062***	0.834***	1.132***	0.633***	1.227***
	(0.01282)	*(0.11249)*	*(0.02164)*	*(0.01404)*	*(0.02054)*	*(0.08251)*	*(0.07205)*	*(0.00682)*	*(0.01247)*	*(0.01386)*	*(0.01580)*	*(0.00889)*
Hrs_ff	0.010	0.037	0.039	0.029	-0.026	0.053	-0.021	-0.007	0.017**	-0.014*	0.031*	-0.002
	(0.00797)	*(0.03352)*	*(0.02545)*	*(0.01777)*	*(0.01954)*	*(0.03884)*	*(0.05328)*	*(0.00832)*	*(0.00723)*	*(0.00781)*	*(0.01657)*	*(0.00701)*
Hrs_if	0.003	0.016	0.016	-0.006	-0.008	-0.016	0.002	0.003	0.009**	-0.007**	0.013	-0.001
	(0.00356)	*(0.01513)*	*(0.01178)*	*(0.00739)*	*(0.00942)*	*(0.02138)*	*(0.02618)*	*(0.00376)*	*(0.00340)*	*(0.00311)*	*(0.00835)*	*(0.00331)*
Hrs_ii	-0.002	-0.020	0.032*	-0.058***	-0.009	-0.021	-0.026	0.008	0.029***	-0.022***	0.018	-0.012**
	(0.00570)	*(0.02473)*	*(0.01713)*	*(0.01193)*	*(0.01421)*	*(0.03214)*	*(0.03837)*	*(0.00585)*	*(0.00542)*	*(0.00550)*	*(0.01452)*	*(0.00498)*
Hrs_sf	0.024***	0.012	0.002	0.016	-0.013	-0.008	-0.012	0.008	-0.012**	0.007	0.021*	0.001
	(0.00616)	*(0.02336)*	*(0.01661)*	*(0.01069)*	*(0.01239)*	*(0.02644)*	*(0.03624)*	*(0.00489)*	*(0.00488)*	*(0.00468)*	*(0.01103)*	*(0.00427)*
Hrs_si	0.025***	0.000	0.139***	-0.018	0.010	-0.028	0.027	0.002	0.012	-0.006	0.075***	-0.027***
	(0.00757)	*(0.02998)*	*(0.02283)*	*(0.01495)*	*(0.01865)*	*(0.04169)*	*(0.05055)*	*(0.00707)*	*(0.00799)*	*(0.00724)*	*(0.01699)*	*(0.00639)*

Note: Robust standard errors in parenthesis with Weesie (1999) corrections for between-model (PROBIT and SUR) vce. *** p<0.01, ** p<0.05, * p<0.1. In the case of expenditure elasticities, significance tests are of the form: different to 1. Expenditure corresponds to group expenditure in each case.

The results for Fon and TC are comparable to those of Böhme and Thiele (2011, 2012b). Using their aggregate results for the African countries modelled, the computed conditional elasticities are qualitatively similar to those presented here: informal channels of commercialization generally show lower elasticities than formal channels for both types of good. However, quantitatively, the contrasts are greater. For food consumption, the authors found both formal and informal elasticities below 1, with average gaps lower than those presented here. In the case of transport and communication, the results are more comparable: elasticities for formal consumption are above 1 and informal ones in general below 1, with gaps slightly wider than those presented here.

Turning our attention to labour supply variables, again differences between goods are to be found. For Fon, the effects seem to be concentrated only on the informal equation, where working more hours in the self-employment sector has a positive effect on informal Fon purchases. This is an interesting finding since it reveals that, although working more hours (in general) has a negative effect on food consumption, once the household has decided to purchase food, working more hours in the self-employment sector will increase food bought from informal markets. By contrast, working more hours in the wage-earning sector has no influence on either formal or informal consumption. The result reveals a particular association between working in the self-employment sector and buying food from informal outlets and could be interpreted in favour of the linkage hypothesis considering the close relation between formal and informal self-employment. As mentioned, using anecdotal evidence it is possible to infer that most formal self-employment has an informal past, so consumption patterns could be similar considering that informational advantages gained during the informal period and networks developed may well be strong enough to prevail during the formal period. However, for a shortrun model where working transitions are not being explicitly analysed, it is better to interpret this evidence as not being in line with the proposed hypothesis.

The results are clearer in the case of Foff. The statistically significant marginal effects detected are very close to those encountered in the aggregate version of the model: informal hours of work in both the wage-earning (Hrs_ii) and the self-employment (Hrs_si) sectors increase informal consumption shares, with the latter effect greater, but the former also reducing formal shares.[70] Therefore, compared with previous estimates, these results are more clearly in line with the linkage hypothesis: any additional working hour has a positive effect on Foff purchases, but when this hour comes from the informal sector it will not only increase Foff consumption, but will also

[70] Chi-squared results: 7.50***

skew it in favour of informal-market purchases. This is an expected outcome considering that informational advantages and other reasons supporting the linkage hypothesis are probably better grounded for meals at work. For example, food vendors are normally located close to workplaces and they tend to share most of the characteristics of their clients (including their formal or informal condition), thus reducing true access costs within sectors. Also, repeated consumption helps to develop confidence links between market participants with potential influence on searching costs.

Weaker evidence, but still in line with the linkage hypothesis, is provided by TC and ED. In the first case (TC), it seems that wage-earning opportunities (both formal and informal) increase the consumption of informal and reduce that of formal transportation. The first effect is comparable to the evidence encountered previously for group consumption, where a positive correlation with the wage-earning sector was detected. This result is possibly explained by the characteristics of the transportation market in Peruvian cities, highly dominated by informal providers. Therefore, it is understandable that the greater number of trips to work generated by increasing the hours worked in the wage-earning sector (the less flexible one) will result in informal consumption. By contrast, although formal providers do not dominate any segment of the market, they are more common for longer trips or weekend taxi services, a type of consumption possibly used out of working time or during the vacation. Also private transportation (mainly formal) could be behind this result.

When formally tested, the marginal effects between formal and informal wage-earners are not significantly different, either in the formal or in the formal equation, so no evidence in favour of the linkage hypothesis is found for this labour opportunity.[71] Different is the case of the self-employment sector. Here, the only detectable effect is the negative linkage of Hrs_sf with informal consumption – a result that could be interpreted in terms of this hypothesis. Interestingly enough, when the SUR model is analysed separately, the evidence in favour of the linkage hypothesis is stronger. The results in this case go in the same direction as previous ones but, when the coefficients are tested, the effects of informal wage-earners on informal and formal consumption seem to be stronger[72] (in absolute terms). Similarly, the negative effect of Hrs_sf on informal TC consumption is still detected in the SUR model, but is accompanied by its positive effect on formal TC.

[71] Chi-squared results. Informal equation: 1.43 (ff vs if), 3.26* (ff vs ii), 0.01 (if vs ii). Formal equation: 0.83 (ff vs if), 1.38 (ff vs ii), 0.01 (if vs ii).
[72] Chi-squared results. Informal equation: 0.84 (ff vs if), 3.60* (ff vs ii); Formal equation: 1.32 (ff vs if), 4.29** (ff vs ii).

In the second case (ED), both formal and informal earning opportunities (Hrs_ff, Hrs_sf and Hrs_si) show a positive and statistically similar[73] effect on informal ED consumption, but only informal labour supply (Hrs_ii and Hrs_si) reduces its formal counterpart. These results are mainly driven by SUR estimates, since the PROBIT models reveal similar effects[74] across all the earning opportunities in both formal and informal markets. Therefore, in this case, the weak evidence in favour of the linkage hypothesis comes in the form of informal labour opportunities (in both the waged and the self-employment sectors) reducing formal allocations.

In terms of CC and HEA, something very similar to the first-stage results is observed, since hours of work have no effect not only on the amount of CC and HEA, but also on the decision as to where these goods will be purchased. In consequence, it is not possible to arrive at any conclusions in terms of the linkage hypothesis in these last two cases.

In sum, the second stage of the model reveals that formalization of the consumption baskets, given higher group expenditure, is not equally applicable across all consumption groups. The evidence of computed elasticities reveals formal markets growing faster than informal ones for four out of six goods: Fon, Foff, TC and ED. Similarly, evidence in favor of the linkage hypothesis is not equally applicable to all goods considered in this estimation. Considering the results of marginal effects, only in one group (Foff) is strong evidence found, with similar patterns to those described in the aggregated version of the model. Weaker evidence, but still in line with the linkage hypothesis, was found in the case of TC and ED, with different patterns of effects. In the first case, this is mainly through the negative effect of formal self-employment on informal consumption and, in the second case, it is driven by the negative effect of informal labor supply (in both the wage-earning and the self-employment sectors) on formal consumption. Interestingly enough, two of the three cases where the hypothesis was validated (Foff and TC) were previously identified as quantity-substitutes for leisure.

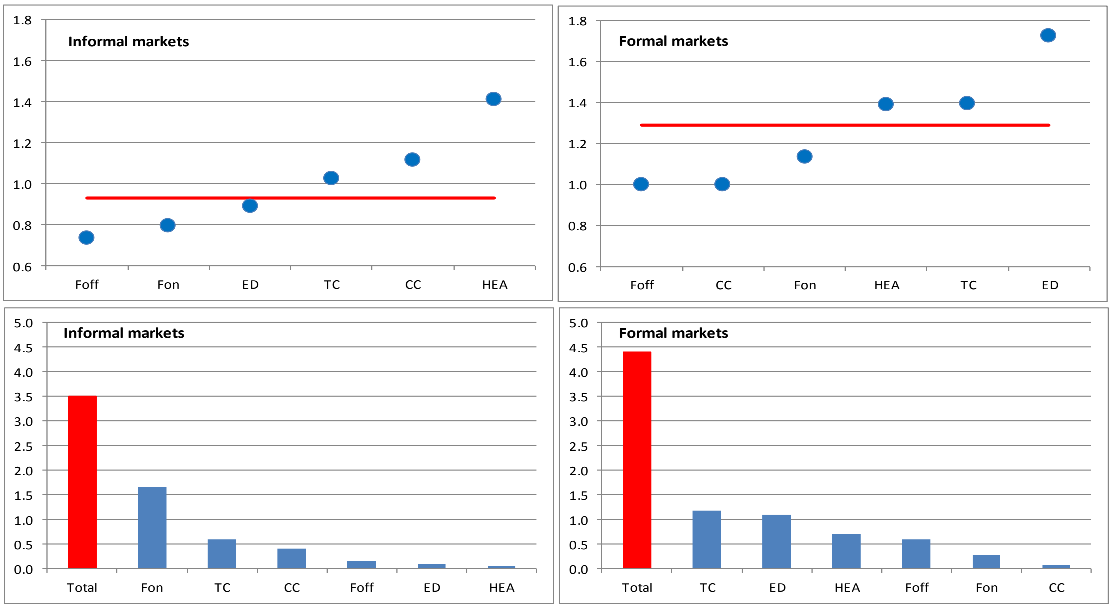
First and second stages together for policy simulations

Now that the results for the first and second stages of the model have been established, it is possible to compute the unconditional elasticities merging them, using the formulas already presented in the previous section. The objective is to render comparable the disaggregated version of the model with the aggregated one in order to compare both results and assess how the public policy suggestions obtained in Chapter 5 behave by consumption group. For the next exercises,

[73] Chi-squared results: Informal equation: 0.57 (ff vs sf), 1.48 (ff vs si), 0.00 (sf vs si).
[74] Chi-squared results: Informal equation: 0.03 (if vs sf), 0.06 (if vs si), 0.20 (sf vs si). Formal equation: 0.24 (ff vs if), 0.28 (ff vs ii), 0.46 (ff vs si), 1.02 (if vs ii), 1.27 (if vs si), 0.03 (ii vs si).

the results for the aggregated version of the model correspond to the SUR–GLS estimates in earlier chapters, considering the three-market definition and under specification (1). This version of the model is the one that shares most of the characteristics of the disaggregated estimations presented here. The comparison must take into account that both SUR–GLS and S&Y are possibly downwardly biased but, in general, it is expected that the bias will be in the same direction and of highly comparable size across the models. Therefore, the conclusion driven by the comparison will be still accurate.

Figure VI.1: Unconditional income elasticities vs aggregated model elasticities



Note: Implied marginal effects are computed using the full sample means: 8,007 observations for the aggregated model and 7,253 observations for the disaggregated model. When computed elasticities are not significantly different to 1, this value is used. Aggregate results used for comparison purposes are SUR–GLS estimates under specification (1) and using the three-market definition. In the top graphs the blue dots correspond to elasticities computed for each consumption group and the red lines to elasticities computed for the whole formal or informal market. In the lower graphs, the blue lines are the marginal effects (for a S/.10 increase in income in consumption levels) for consumption groups; the red lines stand for the marginal effects for the whole formal or informal markets.

In Figure VI.1, the expenditure effects are compared. In the top panels, the results for elasticities are presented. The red line corresponds to the expenditure elasticity for formal and informal markets computed for the aggregated model in Chapter 5.[75] The blue dots correspond to the unconditional elasticities computed for market allocation within each of the consumption groups using the formula (VI.7). The vertical axis in this case is measured in percentage changes (%). In the panels at the bottom of the graph these elasticities are used to compute the implied marginal effects in terms of consumption levels for a S/.10 increase in total expenditure. The red bar

[75] It is worth noting that the results in Chapter 5 are broadly comparable to the implied weighted average values for formal and informal markets computed using the seven commodities in this chapter. In fact the weighted elasticity for formal markets is around 1.19 (very close to the 1.29 of Chapter 5) and the weighted elasticity for informal markets is around 0.89 (very close to the 0.93 of Chapter 5). The weighted results are computed using the adding-up property for the 'others' group (OT) and weights are computed from those presented in Appendix 4.

corresponds to the estimate for the aggregated model and the blue bars to consumption groups. The vertical axis in this case is measured in *New Soles* (S/.).

One first, important, result to note is that the value of total expenditure elasticities is always higher than zero. Therefore, again, the inferiority of informal goods is discarded, even for particular consumption groups. In general, formal goods are always either classified as unitary (with income elasticities equal to 1) or as luxuries (with income elasticities greater than 1). This is an expected result. However, although informal consumption was previously classified as a necessity (elasticity lower than 1), not all of the informal goods show this same result. Some of the consumption groups included (Fon, Foff and ED) replicate it, but others (TC, CC and HEA) show unconditional elasticities greater than 1. Therefore, some consumption groups included in the analysis, even if they are bought at informal markets, keep their status as luxuries (as in the first stage).

However, note that, despite this classification, as identified for the aggregate model, informal elasticities are generally lower than formal ones, revealing similar patterns to those in the previous chapter – which means that income growth is associated with the formalization of the consumption basket even for particular goods. The only exceptions found in this application are CC, where informal elasticity is higher, and HEA, where formal and informal elasticities are very similar. In consequence, income growth formalizes the consumption basket for Peruvian households (result in Chapter 5), but its impact will be different across consumption groups. In this application, formalization occurs mainly in four out of the six groups (Fon, Foff, TC and ED) where positive gaps[76] between formal and informal unconditional elasticities are identified. Interestingly enough, this gap seems to be wider when the consumption groups under analysis are luxuries (ED and TC) than when they are necessities or unitary (Foff and Fon), maybe because of the greater differences between formal and informal options for more complex goods.

Using the results from the lower panel, the re-allocation of expenditures for a S/.10 increase in total expenditure can be examined. As mentioned in the previous chapter, these additional resources will be allocated in the following way: S/.4.5 on formal consumption and S/.3.5 on informal markets. Considering formal allocation, of the additional S/.4.5 purchased, around 65% is explained by education and cultural expenditure, transport and communication and health expenditure (27, 25 and 16%, respectively). In the case of the additional S/.3.5 expenditure on informal consumption, around 65% is explained by food consumption and transport and

[76] Chi-squared results for equality in elasticities: 21.29*** (Fon), 67.71*** (Foff), 2.12 (C), 0.06 (HEA), 244.21*** (TC), 1011.78*** (ED).

communication (47 and 17%, respectively). This information could be used to infer the main channels of engagement (in terms of consumption groups) of the family with the formal and informal economy respectively.

The unconditional elasticities for hours of work computed using formula (VI.8) are presented in Table VI.7. These results are compared with the aggregated version of the model in order to analyse whether the predicted responses on aggregate consumption for different changes in labour supply are more or less replicated across consumption groups. To facilitate the analysis, the comparable results are highlighted. First, in the case of Hrs_ff, the aggregate model found that the consumption basket becomes more formal as the allocation of time to this earning opportunity increases. The results obtained showed a positive association between Hrs_ff and formal markets and a negative association with informal ones. In the disaggregate model, a formalization of the consumption basket similar to this effect is identified only for Fon. In this case a positive association of Hrs_ff with formal consumption is identified, with informal purchases kept almost invariant. In the case of Foff, some evidence along these lines is also identified, with formal markets increasing more than informal ones but, when formally tested,[77] the elasticities are not statistically different. The other consumption groups show contrary or neutral results.

Second, in the case of Hrs_ii, the aggregate model shows that working more hours in this earning opportunity generates an informalization of the consumption basket via the reduction of formal purchases, with informal ones practically invariant. A similar effect is detected in ED. In this case, Hrs_ii reduces consumption at formal and informal markets, but with greater reductions[78] found in the former. In the case of Foff and TC, we can argue for an informalization of the budgets driven by Hrs_ii but, in this case, because of the greater elasticities of informal consumption when compared with its formal counterpart. For TC, Hrs_ii increases informal consumption, with formal counterparts practically invariant and for Foff, although both effects (at formal and informal markets) are positive, the results for informal goods are statistically[79] higher. In the case of Fon, some evidence along these lines is also identified, with formal markets decreasing more than informal ones; however, when formally tested,[80] the elasticities are not statistically different. Other groups also show neutral effects.

---

[77] Chi-squared results: 0.34
[78] Chi-squared results: 27.22***
[79] Chi-squared results: 6.71***
[80] Chi-squared results: 1.62

Table VI.7: Hours' unconditional elasticities (pure effect) vs aggregated model results

	AGGREGATED		Fon		Foff		CC		HEA		TC		ED	
	Informal	**Formal**	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
Hrs_ff	**-0.014*** **	**0.016*** **	-0.019**	-0.005	0.132***	0.150***	-0.030	0.050	-0.034	-0.019	0.031***	0.006	-0.002	-0.065***
	(0.00478)	***(0.00611)***	*(0.00821)*	*(0.03149)*	*(0.02491)*	*(0.01816)*	*(0.01964)*	*(0.03528)*	*(0.05524)*	*(0.01793)*	*(0.00899)*	*(0.01037)*	*(0.01850)*	*(0.01968)*
Hrs_if	**0.002**	**0.002**	-0.008**	0.001	0.056***	0.046***	-0.002	-0.011	-0.006	-0.006	0.014***	0.000	0.003	-0.021**
	(0.00272)	***(0.00288)***	*(0.00330)*	*(0.01236)*	*(0.01153)*	*(0.00764)*	*(0.00883)*	*(0.01922)*	*(0.02394)*	*(0.00840)*	*(0.00362)*	*(0.00442)*	*(0.00806)*	*(0.00851)*
Hrs_ii	**-0.006**	**-0.021*** **	-0.033***	-0.065***	0.149***	0.096***	-0.004	-0.016	-0.032	0.002	0.041***	-0.006	-0.038***	-0.120***
	(0.00394)	***(0.00398)***	*(0.00557)*	*(0.02352)*	*(0.01597)*	*(0.01247)*	*(0.01471)*	*(0.02879)*	*(0.03760)*	*(0.01391)*	*(0.00626)*	*(0.00760)*	*(0.01333)*	*(0.01036)*
Hrs_sf	**0.001**	**0.015*** **	0.010	-0.009	0.003	0.018	-0.014	-0.009	0.001	0.020*	-0.006	0.015**	0.021*	0.000
	(0.00351)	***(0.00495)***	*(0.00648)*	*(0.02052)*	*(0.01545)*	*(0.01263)*	*(0.01163)*	*(0.02288)*	*(0.03570)*	*(0.01149)*	*(0.00549)*	*(0.00666)*	*(0.01085)*	*(0.01232)*
Hrs_si	**0.021*** **	**0.006**	-0.023***	-0.069***	0.259***	0.140***	0.061***	0.019	0.024	-0.001	0.015*	-0.001	0.052***	-0.072***
	(0.00499)	***(0.00564)***	*(0.00773)*	*(0.02649)*	*(0.02263)*	*(0.01776)*	*(0.01951)*	*(0.03867)*	*(0.04845)*	*(0.01619)*	*(0.00898)*	*(0.00939)*	*(0.01668)*	*(0.01393)*

Note: Robust standard errors in parenthesis with Weesie (1999) corrections for between-model (first- and second-stage) vce. *** p<0.01, ** p<0.05, * p<0.1. Aggregate results used for comparison purposes are SUR–GLS estimates under specification (1) and using the three-market definition.

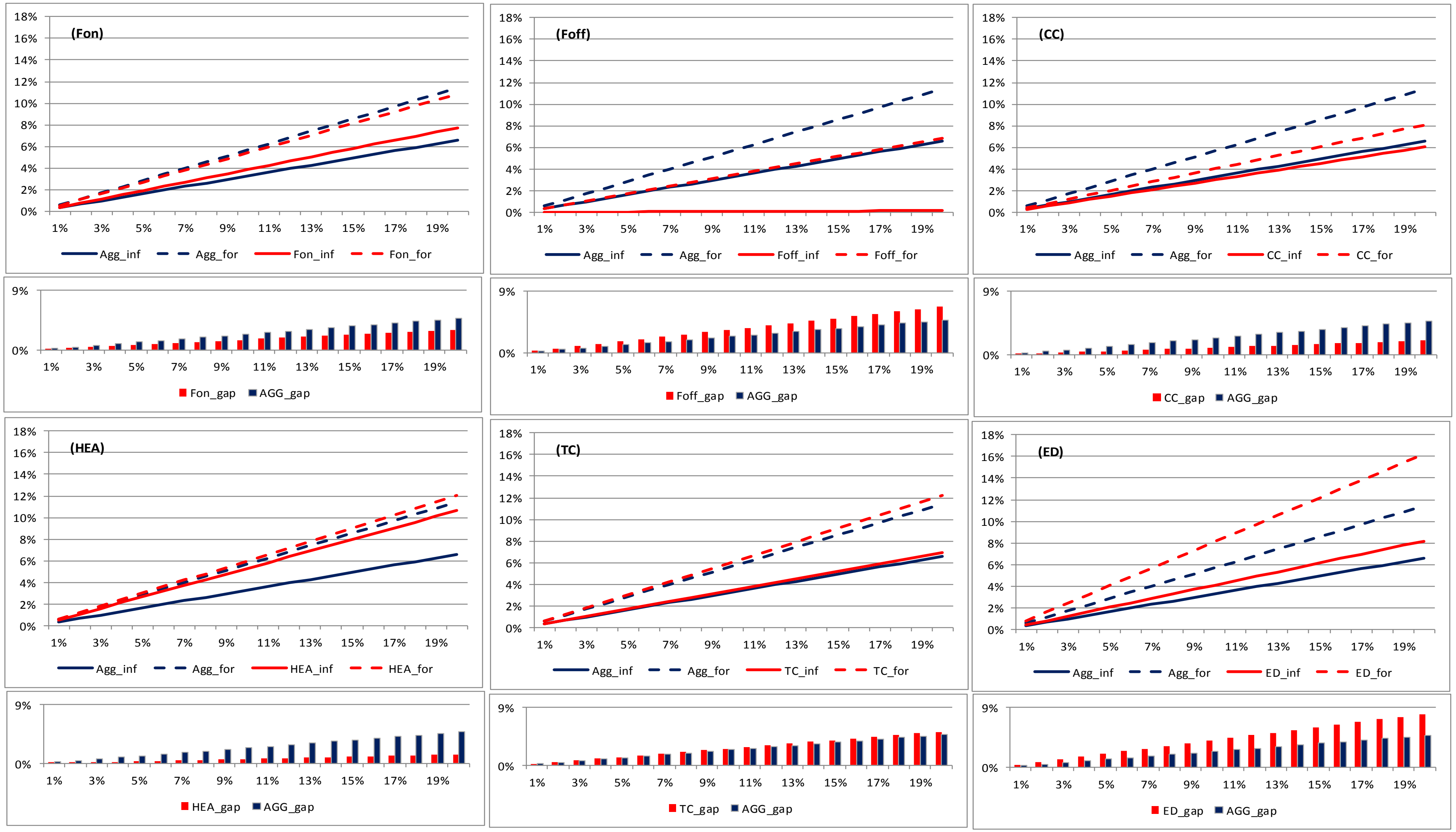
Third, in the case of Hrs_sf, the aggregate model showed that, *ceteris paribus*, working more hours in this sector will increase formal consumption, with informal consumption practically invariant. Similar effects are detected for HEA and TC. In both cases, formal consumption increases with hours worked in this sector, with informal consumption practically invariant (as in the aggregate case). In the other consumption groups, neutral (Fon, Foff, CC) or contrary results (ED) are found.

Fourth, in the case of Hrs_si, the aggregate model shows that working more hours as an informal self-employer is associated with an informalization of the consumption basket. On average and *ceteris paribus*, Hrs_si increases consumption from informal markets, with formal consumption practically invariant. A similar result is encountered in CC and TC. Results in the same direction are also found for ED, where the positive influence on informal consumption is also accompanied by a negative influence on formal consumption. Similarly, for Foff, although both formal and informal consumption increases, the results are statistically higher[81] in the latter. In the case of Fon, point estimates show that informal consumption decreases less than does its formal counterpart, but the differences are statistically[82] insignificant. In the other cases (HEA), neutral results are also found.

With these results, the overall formalization effect is simulated. For this purpose, full-effect hours' elasticities are computed using expression (VI.9) and used to estimate market responses to the supply-side policy already explained in Chapter 5. The results are presented in Figure VI.2. In the panels at the top, the policy responses in percentage changes of formal and informal consumption (horizontal axis) are presented according to the reduction in percentage terms of hours worked in the informal sector (vertical axis). In the panels at the bottom, the percentage gaps are presented. The results in blue correspond to the aggregate model (benchmark) and in red to the simulation for the particular consumption group modelled. The evidence suggests that there are two consumption groups where the formalization policy is more effective than the aggregate benchmark (ED and Foff); one consumption group where the policy is more or less equally effective (TC); and three groups where the policy is less effective (Fon, CC and HEA). As expected, the consumption groups where the formalization policy is more or equally effective are those where the income elasticity gaps are wider or the linkage effect clearer.

[81] Chi-squared results: 15.94***
[82] Chi-squared results: 2.61

Figure VI.2: Policy simulations (formalization policy on the supply-side)



Note: Average hours of work are computed the maximum sample available for estimation in this chapter (7,253 observations). In all cases, the aggregated results are those presented in Chapter 5: SUR–GLS under specification (1) for the three-market definition. The vertical axis displays the expected growth of expenditure in each sector and the vertical axis the decrease in hours worked in the informal sector (policy result).

VI.4. Concluding remarks

The estimations in this chapter have resulted in four additional conclusions on consumption allocation between formal and informal markets. First, for the first-stage results, consumption groups can be classified in three types: necessities (Fon), unitary (Foff) and luxuries (CC, HEA, TC and ED). The estimation also permits an ordering between those groups which are quantity-neutral for leisure (CC and HEA), quantity-complements (Fon and ED) and quantity-substitutes (Foff and TC). Interestingly enough, the differentiate effects of labour supply are found only on those consumption groups classified as substitutes for leisure. In these cases, working an additional hour in the wage-earning sector has a generally greater effect than it would in the self-employment sector.

Second, as far as the second-stage results are concerned, although the linkage hypothesis is supported for aggregate formal and informal consumption, when particular consumption groups are modelled, the evidence is found only for some of them. In this application, in three out of the six consumption groups, correlations between working and purchasing in the informal (or formal) sector are detected. The strongest evidence is found in food to be consumed outside the house (Foff), an intuitive result considering that this consumption group is mainly composed of workplace meals – a type of consumption where informational advantages are probably better grounded and the location decisions of vendors could generate stronger linkages between market participants. At the same time, the hypothesis is verified in its weaker version in the case of transport and communication (TC) and education and culture (ED). In the other consumption groups no conclusive evidence was found. Interestingly enough, two out of the three cases where the hypothesis was validated (Foff and TC) correspond to consumption groups previously classified as quantity-substitutes for leisure.

Also, it should be acknowledged that the way in which this hypothesis is verified differs according to the consumption groups considered. In the case of Foff, the pattern of results is highly comparable with the aggregated model: positive linkages of informal workers (in both the wage-earning and the self-employment sectors) with informal markets and negative linkages of informal wage-earners with formal consumption. For ED, this last result was identified though also accompanied by the negative effect for the informal self-employed. In the case of TC, the linkage comes mainly in the form of the formal self-employed purchasing less-informal goods.

Third, inferiority of informal goods is again not supported by the data and, in this case, even for particular consumption groups. Group expenditure increases consumption for both formal and informal market allocation, with higher effects detected for the former than for the latter. This evidence was found in four out of six consumption groups (Fon, Foff, TC and ED). In the rest (CC and HEA), group expenditure elasticities for formal and informal markets were found to be similar. Once these results were expressed in unconditional terms in order to compute total expenditure elasticities, formal expenditure was always classified as unitary or luxuries. By contrast, informal markets were classified as necessities in three cases (Fon, Foff and ED) and luxuries in three cases (TC, CC and HEA). Therefore, for some consumption groups included in the analysis, even if bought at informal markets, they keep their condition of luxuries (as in the first stage). However, it should be noted that, despite this classification, as identified for the aggregate model and in the first-stage results, unconditional elasticities for informal markets are still lower than formal ones. This result has been interpreted previously as the formalization effect on the consumption basket produced by increasing family incomes. This formalization effect is found to be higher in the cases of ED and TC and lower in the cases of CC and HEA.

Also, due to the unconditional income effects, it was possible to detect the main channels of engagement with informal and formal markets for a given increase in total expenditure. In the case of informal markets, as expected, the main channel identified is food in household consumption (Fon) with transport and communication (TC) and clothing and personal care (CC) having secondary importance. In the case of the formal market, the main channels detected were education and cultural expenditure (ED) and transport and communication (TC), with health goods and services (HEA) having secondary importance. These results more or less confirm similar commentaries done in the descriptive analysis in Chapter 4.

Fourth, where the policy simulations are concerned, we can see that, in all the groups, the results are similar to those of the aggregate version of the model: replacing informal hours of work by formal options generates more formal consumption baskets. However, there are two groups (Foff and ED) where the policy seems to be more effective, one group (TC) where it is equally effective and three groups (Fon, CC and HEA) where it is less effective when compared to aggregate estimates. The three groups where policy seems to be equally or more effective are those where evidence of the linkage hypothesis is found or where the income gaps are wider.

VII. ESTIMATION OF PRICE ELASTICITIES

VII.1. Introduction

Until now, the estimation procedures presented here have been based on Engel curves where, for simplicity (and lack of information), prices were assumed constant and equal to 1 across all consumers. Although, as suggested in Deaton (1997), this procedure is adequate for estimating expenditure elasticities and the marginal effects of additional demand shifters (as hours worked), it will impede the modelling of complete consumer behaviour. In fact, the lack of own-price and cross-price elasticities is an important gap in the literature when the demand-side of the informal economy is studied (Reilly *et al*., 2006). In this chapter, this weakness is partially addressed. The objective is to test the assumption of imperfect substitution between markets and to reveal which sectors are either more or less responsive to price changes. This last result is particularly important since it will be informative for policy purposes as long as it helps to infer the effectiveness of using demand-side policies (like price changes) in order to reduce (or increase) informal (or formal) expenditure. Similarly, based on the computed compensated elasticities, the chapter produces welfare effects associated with these policies, for which purpose compensated variations are calculated.

The estimation follows a linear approximation of the Almost Ideal Demand System (AIDS). As discussed in Chapter 4, given the available data the application is undertaken only for food consumption (specifically, food to be consumed within the household), which is the category where data on prices (or unit values) were obtainable for the different markets identified in the sample. Therefore, the results presented here are conditional to this group under standard weak separability assumptions. Focusing only on food is adequate for two reasons as far as this exercise is concerned. First, because this is the main expenditure group for Peruvian urban families, and explains almost 33% of total purchases. Second, using only a particular commodity group like food makes interpretation of the results more transparent since the goods supplied by formal and informal outlets can be assumed to be more homogenous.

The chapter is organized as follows. In Section 2, the empirical strategy is discussed. In Section 3, the econometric results for the different estimations of own-price and cross-price elasticities are presented. Section 4 then implements these results in a set of policy simulations for different price changes and the compensated variations associated with them are computed and discussed. Finally, Section 5 summarizes the main findings in the form of concluding remarks.

VII.2. Empirical specification and econometric issues

The empirical specification used in this chapter follows the AIDS demand function already presented in (III.32a)-(III.32b). Allowing for some modification in terms of the notation that will be used in this chapter, the final empirical expression to be estimated here is

$$w_j = \alpha_j + \sum_g \gamma_{jg} \log(p_g) + \beta_j \log\left(\frac{m}{P^*}\right) + \tau_j(A) + e_j \qquad \text{(VII.1a)}$$

where

$$\log P^* = \sum_g \bar{w}_g \log(p_g) \qquad \text{(VII.1b)}$$

w_j or w_g is the budget share of formal, semi-formal or informal food consumption, p_g are the corresponding prices, m is food expenditure in *per-capita* terms, A is a set of additional covariates (including labour supply), e_j is the error term, with $e_j \sim N(0, \sigma^2_{e_\rho})$ and α_j, γ_{jg}, β_j and τ_j are the parameters to be estimated. As stated in Chapter 3, expression (VII.1b) is the stone price index used by Deaton and Muellbauer (1980a) in the derivation of the linear approximation of the AIDS. However, for reasons given in Moschini (1995), stone price indexes may actually be inadequate, given that they are not invariant to units of measurement. The practical solutions given by Moschini consider the log-linear approximation of the Laspeyres index which, in turn, is just the stone price index with constant weights. Although Moschini's (1995) arguments are placed more in a time-series context (while this study is placed in a cross-section framework), sample means (constant across the cross-section) are used in this application, as in some of the examples mentioned in Deaton (1997).

Based on (VII.1a) and (VII.1b), income elasticities are derived, as in the previous chapters, with Marshallian own-price and cross-price elasticities following Chalfant's (1987) approximation (see Green and Alston, 1990) for the linear version of the AIDS. Finally, using the Slutsky identity (Gravelle and Rees, 2004), Hicksian elasticities are computed. The expressions reached are:

$$\eta_j = \frac{\partial c_j}{\partial m}\frac{m}{c_j} = \frac{\beta_j}{w_j} + 1 \qquad \text{(VII.2)}$$

$$\varepsilon^O_{jg} = \frac{\partial c^O_j}{\partial p_g}\frac{p_g}{c_j} = -\delta_{jg} + \frac{\gamma_{jg}}{\hat{w}_j} - \beta_j\left(\frac{\hat{w}_g}{\hat{w}_j}\right) \qquad \text{(VII.3)}$$

$$\varepsilon_{jg}^{H} = \frac{\partial c_j^H}{\partial p_g}\frac{p_g}{c_j} = \varepsilon_{jg}^{O} + \eta_j w_g \tag{VII.4}$$

Expression (VII.2) is the expenditure elasticity, (VII.3) is the ordinary price elasticity derived considering that $\partial \log(P^*)/\partial \log(p_g) = \overline{w}_g$ (expenditure shares are constant) and (VII.4) is the Hicksian price elasticity. Note that δ_{jg} is the Kronecker delta where $\delta_{jg} = 1$ when $j = g$ and $\delta_{jg} = 0$ when $j \neq g$.

Using expression (VII.4) it is also possible to compute welfare effects from price changes in the form of compensating variations (CV). As defined in Gravelle and Rees (2004), the CV is defined as the amount of money which must be taken from the consumer in the final situation (post-price change, labelled as 1) in order to make him or her as well off as in the initial situation (pre-price change, labelled as 0). Therefore, using the notation of the authors, the CV can be defined as

$$CV = M^0 - c(p^1, u^0) = c(p^0, u^0) - c(p^1, u^0) \tag{VII.6}$$

In terms of the expenditure or cost function (c) with u, p and M used to defined the utility level, prices and total expenditure (or incomes), respectively. The computation of the CV in this application follows Niimi's (2007) suggestion based on the second-order Taylor expansion of the minimum expenditure function which, in terms of the budget share specification used here, takes the form (see also Friedman and Levinsohn, 2002):

$$\Delta \log(c) \approx \sum_j w_j \Delta \log(p_j) + \frac{1}{2}\sum_j \sum_g w_j \varepsilon_{jg}^{H} \Delta \log(p_j) \Delta \log(p_g) \tag{VII.7}$$

As commented in Niimi (2007), this formulation has the property to incorporate household behavioural responses to price changes and could be considered superior when compared with the first-order approximation explained in Deaton and Muellbauer (1980b).

All these expressions are identifiable provided that the variables included in the estimation can be set at the sample means and standard errors calculated using the Delta method. However, considering the discussion in Chapter 6, it is important to note that all these expressions are conditional on group expenditure. Without knowledge of prices in the other groups, unconditional elasticities cannot be computed, so the analysis in this chapter will be carried out only at this level.

The specification of demand function (VII.1a)-(VII.1b) is similar to the Working–Leser Engel curves already estimated in Chapters 5 and 6 in terms of the variables included. This means that the definition of m is total expenditure on food to be consumed within the household (Fon) expressed in annual *per-capita* terms; estimations are done under the four specifications already explained in Chapter 5 for the control variables included in A. However, in this case prices are explicitly included in the model. For this purpose, unit values computed as district means are used in the estimation. Following Niimi's (2007) conclusions on the different correctional procedures for unit values, in the absence of market prices this formulation could be considered a second-best option.

Details of the computation of prices have been already explained in Chapter 4. As mentioned there, the computation is based on the available data on the values and volumes of 49 food products purchased at three different types of market (formal, semi-formal and informal, according to the classification of the outlet proposed), so it was possible to construct 147 individual unit values at the household level. These data were then used to compute district means and, at this level, to aggregate them (using geometric indexes) in terms of overall formal, semi-formal and informal food unit values. A standard imputation technique was used for the missing values that appear during computation (5% at the last stage of aggregation) in order to construct a complete vector of these three prices for each district in the sample and perform the estimations proposed here. However, considering the arbitrary decisions made in terms of the weighting factors used during unit values' computation and following Niimi (2007), who shows that results can be sensitive to weight choices, three versions are presented. A base version (UV-1) with weights defined at the district level and two robust checks: (UV-2) with weights at the departmental level and (UV-3) with weights at the national level. The decision on the base also follows Niimi (2007), who demonstrated that the first version outperforms in reducing quality biases.

As in Chapter 6, the main econometric issue to handle during estimation is the censoring bias that will emerge given the high importance of zero-level consumption. The solution proposed is to again rely on the S&Y estimation procedure. This means that the final specification of the model is, in fact, a two-equation procedure with a first-stage PROBIT model for the decision to purchase or not from formal and informal markets and a second-stage SUR–GLS model for the amounts allocated in each. Using previous notation, the estimable equations take the form:

$$P(w_{kj} > 0) = \Phi(\pi_{0kj} Q_{kj}) \qquad \text{(VI.8)}$$

$$w_{kj} = \Phi(\pi_{0kj} Q_{kj}) \left[\alpha_j + \sum_g \gamma_{jg} \log(p_g) + \beta_j \log\left(\frac{m}{P*} \right) + \tau_j(A) + e \right] + \dots$$
$$\dots + \pi_{2kj} \phi(\pi_{0kj} Q_{kj}) + \varepsilon_{kj} \tag{VI.9}$$

where (VI.8) is the PROBIT model and (VI.9) is the SUR–GLS model. Therefore, to derive marginal effects, again the Su and Yen (2000) procedure needs to be implemented, so the corrected formulas for (VII.2) and (VII.3), once the two stages are taken into account, become:

$$\eta_j^{S\&Y} = \frac{\partial c_j}{\partial m} \frac{m}{c_j} = \frac{\Theta_j \beta_j}{w_j} + 1 \tag{VII.10}$$

$$\varepsilon_{jg}^{S\&Y(1),O} = \frac{\partial c_j^O}{\partial p_g} \frac{p_g}{c_j} = -\delta_{jg} + \frac{\Theta_j \gamma_{jg}}{\hat{w}_j} - \Theta_j \beta_j \left(\frac{\hat{w}_g}{\hat{w}_j} \right) \tag{VII.11}$$

with (VII.10) and (VII.11) used to find the adjusted version of expression (VII.4) and $\Theta = \Phi + \phi(\pi_{0,x} / \pi_{1,x})[(\pi_1 Q_2) - \pi_2(\pi_0 Q_1)]$ as defined in Chapter 6. Note also that, as explained there, the estimation must be made with caution, mainly in terms of the theoretical restrictions to be imposed in the model, like adding-up, homogeneity and symmetry. In the first case, dropping one of the consumption categories, as in Yen *et al*. (2003) and Yen and Lin (2006), is still considered an adequate solution (provided that the interpretation of the results is conditional on the decision adopted). In this chapter, to maintain coherence with previous ones, the semi-formal category is still used as the residual equation. In the second and third cases, as in Yen *et al*. (2003), Yen and Lin (2006), Alviola *et al*. (2010), Tafere *et al*. (2010), Boysen (2012) and Akaichi and Revoredo (2013), restrictions are imposed jointly[83] during estimation as parameter restrictions in the second stage of the S&Y estimator. However, as noted by Dong *et al*. (2004) and Barslund (2011), this procedure implies that demand restrictions will be accounted for only on the latent, and not on the observed shares.

To partially handle this issue, an alternative (possibly raw) procedure is implemented here, imposing restrictions on the marginal effects. This has no consequences on the homogeneity property as long it is an intra-equation restriction but, for symmetry (inter-equation restriction), it takes the form $\Phi_j \gamma_{jg} = \Phi_g \gamma_{gj}$ in the second stage of the S&Y estimator, using the sample means of Φ. To keep coherence in the derivation of marginal effects in this case, Φ must be fixed, so

[83] Semi-formal prices are included in each regression, but not reported here.

when this type of restriction is imposed, prices are dropped from the PROBIT equation (Yen *et al*., 2003; Yen and Lin, 2006; Barslund, 2011 also drop prices in their first-stage equations) and expenditure is included in nominal terms (as in Yen *et al*., 2003).[84] Under this assumption, expenditure elasticity is still (VII.10), but the corrected Marshallian elasticity takes the form:

$$\varepsilon_{jg}^{S\&Y(2),O} = \frac{\partial c_j^O}{\partial p_g}\frac{p_g}{c_j} = -\delta_{jg} + \frac{\Phi_j \gamma_{jg}}{\hat{w}_j} - \Phi_j \beta_j \left(\frac{\hat{w}_g}{\hat{w}_j}\right) \qquad \text{(VII.12)}$$

with the corresponding adjustment in Hicksian elasticities. For comparison purposes, this version of the model is also estimated with restrictions imposed on the latent shares (as in Yen *et al*., 2003). Note that expression (VII.7) must consider the new definitions of Hicksian price elasticities when computing CV.

VII.3. Estimation results and computed elasticities

In Table VII.1 the regression results and marginal effects for the food demand system using price vector UV-1 and specification (1) are presented. It is important to note that the sample sizes used in the estimation have changed when compared with those in Chapter 6 since, in this case, dropping some missing values for additional covariates in order to identify others groups' PROBIT models was not necessary. Basic tabulations of the covariates used in this chapter with this new sample size are presented in Appendix 6. Only the estimations for the main variables of interest are shown here in order to conserve space. Full estimates for both the PROBIT and the SUR stages are presented in Appendix 7. In general, the performance of the PROBIT models in this new estimation context is highly comparable to previous S&Y exercises, with adequate global adjustment, joint significance of the additional instruments included, individual significance and correctness of the signs on most of the additional variables included. In the case of the SUR estimates, most of the results for the additional variables are comparable to the effects (in sign and significance) already detected on the Engel curve in the last chapter (for example hours of work), so the robustness of the inclusion of prices could be claimed.

The results are organized into three methods of imposing the demand restrictions: restriction 1 means a correctly specified PROBIT (with prices) and restrictions imposed in the latent shares, restriction 2 means allowing a misspecification of the PROBIT (dropping prices) and restrictions

[84] Yen *et al*. (2003) use nominal incomes and, like Yen and Lin (2006) and Barslund (2011), also drop total expenditure from their PROBIT equation. This last procedure was also implemented, but since it could generate severe misspecification problems in the PROBIT, this model was discarded during estimation. However, results for the elasticities to be presented show minor changes when compared with this alternative procedure.

still imposed in the latent share, and restriction 3 means a misspecified PROBIT (without prices) and restrictions imposed in the marginal effects. Without more guidance about the correct way to impose restrictions within the S&Y framework, it is better to hold the different ways of imposing the restriction throughout the application for comparative purposes. The results presented in the PROBIT and SUR models are of no particular interest since they are just intermediate regressions in order to derive marginal effects; however, in general, they show the expected results in the core variables.

Table VII.1: Regression results and marginal effects for price vector UV-1 under different methods of imposing demand restrictions, specification (1)

		Restriction 1		Restriction 2		Restriction 3	
		Informal	Formal	Informal	Formal	Informal	Formal
Probit	L(expend_Fon)	0.57087***	0.56310***	0.56693***	0.55184***	0.56693***	0.55184***
		(0.03863)	*(0.03096)*	*(0.03851)*	*(0.03062)*	*(0.03851)*	*(0.03062)*
	L(P_informal)	-0.35757	0.38637**	n.a.	n.a.	n.a.	n.a.
		(0.26768)	*(0.17089)*				
	L(P_formal)	0.10512	-0.46372***	n.a.	n.a.	n.a.	n.a.
		(0.11170)	*(0.06111)*				
	Observations	7,513	7,513	7,513	7,513	7,513	7,513
	Pseudo LogL	-692.833	-4228.773	-697.076	-4266.134	-697.076	-4266.134
SUR	L(expend_Fon)	-0.00596	-0.00366	-0.00675	-0.00268	-0.00704	0.00207
		(0.00720)	*(0.01331)*	*(0.00718)*	*(0.01300)*	*(0.00716)*	*(0.01300)*
	L(P_informal)	-0.14533***	0.07775***	-0.17200***	0.07993***	-0.19317***	0.16278***
		(0.02502)	*(0.00910)*	*(0.02398)*	*(0.00909)*	*(0.02442)*	*(0.01843)*
	L(P_formal)	0.07775***	-0.01935*	0.07993***	-0.06040***	0.07944***	-0.06631***
		(0.00910)	*(0.01113)*	*(0.00909)*	*(0.00951)*	*(0.00900)*	*(0.00933)*
	PHI	-0.13311**	-0.05838**	-0.14399**	-0.05859***	-0.14704**	-0.04373*
		(0.06728)	*(0.02300)*	*(0.06633)*	*(0.02224)*	*(0.06627)*	*(0.02237)*
	Observations	7,513	7,513	7,513	7,513	7,513	7,513
	Pseudo LogL	4124.244		4134.204		4152.785	
Marginal Effects	L(expend_Fon)	0.02034***	0.02648***	0.02026***	0.02655***	0.02019***	0.02688***
		(0.00723)	*(0.00708)*	*(0.00726)*	*(0.00690)*	*(0.00727)*	*(0.00686)*
	L(P_informal)	-0.15693***	0.05604***	-0.16639***	0.03773***	-0.18686***	0.07684***
		(0.02851)	*(0.00987)*	*(0.02319)*	*(0.00429)*	*(0.02362)*	*(0.00870)*
	L(P_formal)	0.08002***	-0.03235***	0.07732***	-0.02852***	0.07684***	-0.03130***
		(0.01075)	*(0.00685)*	*(0.00879)*	*(0.00449)*	*(0.00870)*	*(0.00440)*

Note: Robust standard errors in parenthesis with within-model adjustment using Weesie (1999). *** p<0.01, ** p<0.05, * p<0.1. Prices used here correspond to O1_uv in Appendix 6 with restriction 1 using the correspondent adjusted version of L(expend_Fon).

In the case of total Fon expenditure, this increases the probability of buying both formal and informal goods at highly comparable levels, but is not significant in the SUR, revealing less influence in the latent dependant variable. When computed, the marginal effects, as expected, show no noticeable differences between the various methods of imposing restrictions, with point estimates around 0.020 (informal equation) and 0.027 (formal equation). This result was already obtained in the Engel curve model (Chapter 4). The conclusions obtained there hold once prices are included in the estimation: increasing the allocation of goods on food consumption tends to increase both formal and informal shares in highly comparable proportions: differences in the marginal effect are still non-significant.[85]

However, as discussed in Chapter 4, this does not mean that income elasticities are similar. As we will see later, elasticities are still in the order of 1.04 (informal equation) and 1.40 (formal equation). The results are not discussed in depth here, but it is important to remember their implications: as found earlier, formal consumption will increase faster than informal for any percentage change in household food consumption. However, this does not necessarily imply that formal and informal goods are both luxuries. As in Chapter 6, we can see that, once corrected by total expenditure, it is possible to demonstrate that informal food consumption shows an elasticity below 1 (necessity) and formal food consumption an elasticity above 1 (luxury) with values around 0.8 and 1.2, respectively. It is important to bear these results in mind, given their influence when Marshallian price elasticities are derived.

For prices, the results are again as expected in terms of the signs displayed. Own-price effects are negative on both the probability of purchasing and the latent estimates. This is confirmed by the estimation of marginal effects. Therefore, increasing the prices in each sector will reduce the probability that a household buys goods there and will lead to reductions in the shares they bought. It should also be noted that own-price marginal effects are generally higher[86] in the informal than in the formal equation. This possibly reveals that informal markets tend to be more sensitive to own-price shocks than formal ones. However, a result like this does not necessarily relate to what happens in elasticity terms as long as marginal effects need to be corrected by consumption shares which are also higher for informal markets. In the case of cross-price effects, results reveal the expected positive coefficients on both PROBIT and SUR models, revealing substitution between sectors. This is confirmed when marginal effects are computed. Effects on the informal equation are again slightly higher,[87] but it is still preferable to wait until proper elasticities are computed to conclude on the relative responsiveness of formal and informal consumption to price shocks.

An additional issue to consider is the differences in the point estimates across methods associated with prices. As expected, they are more notorious than those mentioned for group expenditure,

[85] Chi squared results: 0.29 (R1), 0.31 (R2), 0.35 (R3).
[86] Chi-squared results: 17.91*** (R1), 33.89*** (R2), 42.13*** (R3).
[87] Chi-squared results: 4.85** (R1), 77.34*** (R2), 0.00 (R3)

mainly in terms of the cross-price effects in the formal equation. Focusing just on the marginal effects for own-price parameters, we find informal equation values of between -0.157 and -0.187 and formal ones of between -0.029 and -0.031. By contrast, the cross-price effects range between 0.077 and 0.080 (informal equation) and between 0.038 and 0.077 (formal equation). Although, as discussed in Deaton and Muelbauer (1980b), marginal effects not are necessarily directly related to elasticities (compensated or uncompensated), it is expected that point estimates will differ between the methods of imposing the restriction; however, it is possible to tentatively conclude that, in this application, differences will be small.

Table VII.2: Elasticities for price vector UV-1 under different methods of imposing demand restrictions, specification (1)

	Expenditure Elasticities		Hicksian Elasticities				Marshallian Elasticities			
	Informal consumption	Formal consumption	Informal consumption		Formal consumption		Informal consumption		Formal consumption	
			Own-price	Cross-Price	Own-price	Cross-Price	Own-price	Cross-Price	Own-price	Cross-Price
Restriction 1	1.034***	1.392***	-0.661***	0.201***	-1.411***	1.430***	-1.282***	0.131***	-1.505***	0.594***
	(0.01205)	*(0.10472)*	*(0.04749)*	*(0.01791)*	*(0.10133)*	*(0.14606)*	*(0.04761)*	*(0.01791)*	*(0.10389)*	*(0.15641)*
Restriction 2	1.034***	1.393***	-0.650***	0.199***	-1.327***	1.406***	-1.271***	0.130***	-1.421***	0.570***
	(0.01209)	*(0.10208)*	*(0.04028)*	*(0.01458)*	*(0.06639)*	*(0.06550)*	*(0.04008)*	*(0.01460)*	*(0.06775)*	*(0.07891)*
Restriction 3	1.034***	1.400***	-0.684***	0.199***	-1.370***	1.968***	-1.305***	0.129***	-1.464***	1.129***
	(0.01210)	*(0.10150)*	*(0.04105)*	*(0.01444)*	*(0.06510)*	*(0.13022)*	*(0.04077)*	*(0.01447)*	*(0.06665)*	*(0.13101)*

Note: Robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Tests for expenditure elasticities are of the form ‘different from 1’. See footnote in Table VII.1 for additional details.

In Table VII.2 Marshallian and Hicksian elasticities are estimated for the different methods used in this application. A general inspection of the overall results is that they reveal higher elasticities (own-price and cross-price) when compared with other studies using the same methodology (see Yen *et al*., 2003; Yen and Lin, 2006; Alviola *et al.*, 2010; Tafere *et al*., 2010; Boysen, 2012; Akaichi and Revoredo, 2013). One explanation behind it is that, in contrast, to previous studies that model price effects across different food consumption groups in this application the model is done for food purchased in different markets. Therefore, high price responses is an expected result as long as substitution possibilities are higher for one good purchased in different markets than for two different goods purchased even in the same market (see Restrepo-Echavarría, 2011; Charlot *et al*., 2013 for similar explanations). Other possibility is that the results might be driven by quality effects on unit values, which are used instead of market prices. Following Deaton’s (1997) discussion, with quality shading it is possible to verify an overestimation of price responses when unit values are used. Although this application uses district means which, in turn, will reduce the size of the bias to a greater extent than alternative procedures (Niimi, 2007), some influence could be still present. However, without better available data, this issue should be taken into account in the interpretation of results.

The results confirm the normality of both formal and informal consumption, with positive income effects and negative compensated own-price effects. The demand observed displays a negative relationship with prices for both formal and informal goods. In this context, as expected, Hicksian estimates are lower in absolute terms given the influence of income elasticities on the Marshallian computation, with higher effects on informal markets than on their formal counterparts. Although expenditure and uncompensated elasticities are lower in the former case, the proportion of the informal consumption share (higher than formal) compensates for this difference (60% vs 7%, see Chapter 4). As a result, own-price Marshallian elasticities are closer between sectors than the differences displayed in pure substitution terms.

Point estimates show that, for a 1% increase of own-price, the consumption of food bought from informal markets decreases by around 0.665% (Hicksian) and around 1.286% (Marshallian); consumption bought from formal markets will decrease by about 1.369% (Hicksian) and around 1.463% (Marshallian), considering the mean of the different estimation methods. When tested, Marshallian elasticities for both markets are statistically higher than 1 (in absolute terms) [88], with formal values higher than informal ones.[89] The implication of the finding is interesting as it reveals elastic observed demand curves. Although a segmentation hypothesis is not being properly tested, a result of high own-price elasticities (possibly driven by high substitution possibilities across both markets) casts doubts on its validity. Similarly, the results define a formal sector more responsive to own-price effects, a result possibly driven by better substitution possibilities for the formal sector in informal markets than the converse.

In order to characterize demand patterns, it is better to focus on compensated demands. In this case, Hicksian own-price elasticities for the formal equation are still greater than 1, but informal elasticities are below 1 (in absolute terms).[90] Therefore, in utility constant terms, the evidence supports the notion that informal goods are price-inelastic while formal ones are still price-elastic, or, similarly, that the gaps for the pure substitution effects encountered are higher than the gross own-price effects. In terms of the compensated cross-price elasticities, imperfect substitution between formal and informal markets is supported by the data at least for food to be consumed within the household. Point estimates (on average across methods) reveal that, for 1% increase in formal prices, informal consumption will increase by 0.200%. In the case of formal consumption,

[88] Chi-squared results for the Marshallian demand (informal consumption): 35.03*** (R1), 45.60*** (R2), 55.78*** (R3). Chi-squared results for the Marshallian demand (formal consumption): 23.65** (R1), 38.56** (R2), 48.50*** (R3).
[89] Chi-squared tests for equality of own-price Marshallian elasticities: 3.76* (R1), 3.57* (R2), 4.21** (R3).
[90] Chi-squared results for the Hicksian demand (informal consumption): 50.93*** (R1), 75.46*** (R2), 59.28*** (R3). Chi-squared results for the Hicksian demand (formal consumption): 16.47*** (R1), 24.20*** (R2), 32.25** (R3). Chi-squared tests for the equality of own-price Hicksian elasticities: 44.31*** (R1), 75.26*** (R2), 80.58*** (R3).

a similar shock in informal prices will generate an increase of 1.601%. Interestingly enough, compensated cross-price elasticities for formal consumption are statistically higher[91] when compared to informal ones. Therefore, we can also argue that, considering utility constant demand functions, households' adjustment of formal consumption to informal prices changes will be higher (in percentage terms) than the converse for a similar price shock. The data also support the find that compensated own-price elasticities are higher than cross-price effects for informal consumption, but not necessarily for formal.[92] In the latter, similar own-price and cross-price elasticities are detected at least in two out of three methods. The results for the uncompensated Marshallian elasticities are along the same lines: gross substitution between markets with formal consumption being more responsive.[93]

Table VII.3: Marginal effects for price vectors UV-2 and UV-3 under different methods of imposing demand restrictions, specification (1)

		Restriction 1		Restriction 2		Restriction 3	
		Informal	Formal	Informal	Formal	Informal	Formal
UV-2	L(expend_Fon)	0.01921***	0.02862***	0.01964***	0.02876***	0.01984***	0.02891***
		(0.00739)	*(0.00730)*	*(0.00745)*	*(0.00713)*	*(0.00744)*	*(0.00712)*
	L(P_informal)	-0.11429***	0.03409**	-0.10418***	0.03259***	-0.12387***	0.06781***
		(0.03896)	*(0.01344)*	*(0.02958)*	*(0.00433)*	*(0.03017)*	*(0.00853)*
	L(P_formal)	0.07713***	-0.00921	0.06678***	-0.00784	0.06781***	-0.01081**
		(0.01203)	*(0.00752)*	*(0.00888)*	*(0.00532)*	*(0.00853)*	*(0.00500)*
UV-3	L(expend_Fon)	0.02244***	0.02717***	0.02277***	0.02784***	0.02306***	0.02778***
		(0.00727)	*(0.00728)*	*(0.00738)*	*(0.00716)*	*(0.00736)*	*(0.00714)*
	L(P_informal)	-0.32094***	0.06879***	-0.31208***	0.02226***	-0.33374***	0.05805***
		(0.05416)	*(0.01713)*	*(0.04445)*	*(0.00559)*	*(0.04501)*	*(0.01134)*
	L(P_formal)	0.04593***	-0.01283	0.04561***	-0.01683**	0.05805***	-0.02076***
		(0.01371)	*(0.00874)*	*(0.01145)*	*(0.00677)*	*(0.01134)*	*(0.00629)*

Note: Robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. See footnote in Table VII.1 for additional details. Prices used here correspond to O2_uv and O3_uv in Appendix 6, with restriction 1 using the correspondent adjusted version of L(expend_Fon).

In order to check the robustness of the results in Table VII.3, the estimations for the alternative price definitions are presented. Note that only marginal effects are included to conserve space. Full estimation results (PROBIT and SUR) are presented in Appendix 7. In general, the marginal effects of income are highly comparable across the three definitions of prices, so the previous conclusions still hold. Similarly, the price effect (own- and cross- in both equations) is qualitatively similar between definitions, although we can see that own-price effects for formal

[91] Chi-squared tests for the equality of cross-price Hicksian elasticities: 78.12*** (R1), 551.66*** (R2), 233.03*** (R3).
[92] Chi-squared test of own vs cross (informal): 91.29***(R1), 120.87*** (R2), 145.17*** (R3). Chi-squared test of own vs cross (formal): 0.01 (R1), 0.89 (R2), 20.17*** (R3).
[93] Chi-squared tests for the equality of cross-price Marshallian elasticities: 9.46*** (R1), 40.43*** (R2), 71.67*** (R3).

consumption are not well determined in two out of three cases under UV-2 and in one out of three for UV-3. Also, point value estimates differ. In general, UV-2 reveals lower effects in absolute terms across all the estimates presented and UV-3 higher effects related to informal prices and lower effects related to formal ones. The highest mean average differences are encountered in the own-price effects displayed, similar to that found in Niimi (2007) related to changing the weighting factors associated with price construction. As also found there, the mean differences encountered are not necessarily all positive or all negative for own-price and cross-price elasticities, so is not possible to provide concrete arguments about the direction of the quality biases. In the absence of market-price data, we cannot explore these gaps in any depth.

Table VII.4: Own-price and cross-price elasticities for price vectors UV-2 and UV-3 under different methods of imposing demand restrictions, specification (1)

		Expenditure Elasticities		Hicksian Elasticities				Marshallian Elasticities			
		Informal	Formal	Informal consumption		Formal consumption		Informal consumption		Formal consumption	
				Own-price	Cross-Price	Own-price	Cross-Price	Own-price	Cross-Price	Own-price	Cross-Price
UV-2	R.1	1.032***	1.424***	-0.590***	0.196***	-1.069***	1.105***	-1.210***	0.126***	-1.165***	0.250
		(0.01232)	*(0.10807)*	*(0.06491)*	*(0.02004)*	*(0.11130)*	*(0.19888)*	*(0.06524)*	*(0.02002)*	*(0.11353)*	*(0.20777)*
	R.2	1.033***	1.426***	-0.545***	0.182***	-1.021***	1.324***	-1.162***	0.113***	-1.118***	0.468***
		(0.01240)	*(0.10552)*	*(0.05060)*	*(0.01475)*	*(0.07866)*	*(0.06597)*	*(0.05056)*	*(0.01475)*	*(0.07996)*	*(0.08153)*
	R.3	1.033***	1.428***	-0.575***	0.184***	-1.067***	1.827***	-1.196***	0.114***	-1.164***	0.970***
		(0.01239)	*(0.10534)*	*(0.05161)*	*(0.01417)*	*(0.07398)*	*(0.12758)*	*(0.05152)*	*(0.01419)*	*(0.07554)*	*(0.12876)*
UV-3	R.1	1.037***	1.402***	-0.934***	0.144***	-1.122***	1.618***	-1.557***	0.074***	-1.217***	0.777***
		(0.01211)	*(0.10772)*	*(0.09022)*	*(0.02283)*	*(0.12940)*	*(0.25356)*	*(0.09071)*	*(0.02284)*	*(0.13095)*	*(0.26149)*
	R.2	1.038***	1.412***	-0.890***	0.147***	-1.153***	1.183***	-1.513***	0.077***	-1.248***	0.335***
		(0.01229)	*(0.10595)*	*(0.07496)*	*(0.01900)*	*(0.10016)*	*(0.08432)*	*(0.07504)*	*(0.01902)*	*(0.10080)*	*(0.09803)*
	R.3	1.038***	1.411***	-0.927***	0.168***	-1.212***	1.705***	-1.550***	0.097***	-1.307***	0.858***
		(0.01225)	*(0.10559)*	*(0.07596)*	*(0.01883)*	*(0.09297)*	*(0.16904)*	*(0.07601)*	*(0.01885)*	*(0.09378)*	*(0.17181)*

Note: Robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Tests for expenditure elasticities are in the form ‘different to 1’. See footnote in Table VII.3 for additional details.

In Table VII.4, the computation of Hicksian and Marshallian price elasticities are presented for the two new definitions of prices. In all cases the expected signs emerge: negative compensated and uncompensated own-price elasticities and positive compensated and uncompensated cross-price ones. Therefore, imperfect substitution between channels of commercialization is a result robust to the changes in the definition of unit values used here. Also, in general, the main characterization of demand patterns for the formal and informal sector still holds, with price-elastic Marshallian demand curves for both formal and informal consumption and price-elastic Hicksian demands only for the formal sector. Informal consumption again prove to be price-inelastic in utility constant terms.

However, in line with the marginal effects constructed previously, the point estimates differ. The most important changes in terms of the qualitative conclusions displayed earlier relate to the differences between uncompensated own-price elasticities. Considering the new definitions of prices, it is not possible to argue that formal own-price effects are higher than their informal counterparts. In fact, under UV-3, although both own-price elasticities are still greater than 1, now the price elasticities for informal consumption are statistically higher.[94] In the case of UV-2, the elasticities for formal consumption are not statistically different to 1, with informal ones still higher than unity.[95] In this last case, it is strange that, when the point values of both elasticities are compared, they are statistically similar – a result possibly driven by the lowest precision of the estimates of own-price effects on the formal equation mentioned earlier. The other conclusions hold, despite the differences in point value estimates displayed.

In Tables VII.5 and VII.6, additional robust checks are presented, varying the specification of the demand equations. As in Chapter 5, specification (2) refers to the inclusion of controls for location proxies, specification (3) to ethnic and cultural backgrounds and specification (4) to enforcement. The estimations are carried out using the base-price vector (UV-1) under the different methodologies of imposing demand restrictions. Only the marginal effects and elasticities are presented. Full versions of the PROBIT and SUR models are presented in Appendix 7. The general information that this new robust check provides is that previous results are not sensitive to the specification used and only minor differences in the point value estimates are observed. The only important change observed is the reduction in the total food expenditure marginal effect under specification (2), but without the material consequences of the calculations of price elasticities. The other results are still roughly the same as before, with negative own-price and positive cross-price effects on both formal and informal shares.

As a result, when Marshallian elasticities are computed, uncompensated elastic demands are still found for formal and informal consumption, with higher values in the former than in the latter (this last result is less clear under specification 4). Similarly, inelastic compensated demand curves are found for informal consumption while formal ones are still elastic. Cross-price elasticities still show imperfect substitution across markets, with formal consumption more responsive to informal price shocks than the reverse. This result is also observed in Marshallian demands.

[94] Chi-squared results. Different to 1 (informal): 37.71*** (R1), 46.76*** (R2), 52.34*** (R3); Different to 1 (formal): 2.75* (R1), 6.07** (R2), 10.74*** (R3); Different informal vs formal: 4.33** (R1), 4.26** (R2), 3.84* (R3).

[95] Chi-squared results. Different to 1 (informal): 10.32*** (R1), 10.32*** (R2), 14.39*** (R3); Different to 1 (formal): 2.11 (R1), 2.16 (R2), 4.70** (R3); Different informal vs formal: 0.11 (R1), 0.21 (R2), 0.11 (R3).

Table VII.5: Marginal effects for price vector UV-1 under different methods of imposing demand restrictions and specifications

		Restriction 1		Restriction 2		Restriction 3	
		Informal	Formal	Informal	Formal	Informal	Formal
Specification (2)	L(expend_Fon)	0.01232*	0.02734***	0.01252*	0.02718***	0.01272*	0.02725***
		(0.00683)	*(0.00690)*	*(0.00684)*	*(0.00660)*	*(0.00683)*	*(0.00656)*
	L(P_informal)	-0.15018***	0.04598***	-0.15561***	0.03383***	-0.17616***	0.07363***
		(0.02396)	*(0.00705)*	*(0.02126)*	*(0.00429)*	*(0.02170)*	*(0.00856)*
	L(P_formal)	0.07126***	-0.04044***	0.06932***	-0.03818***	0.07363***	-0.04151***
		(0.00971)	*(0.00698)*	*(0.00878)*	*(0.00471)*	*(0.00856)*	*(0.00460)*
Specification (3)	L(expend_Fon)	0.02099***	0.02638***	0.02085***	0.02642***	0.02072***	0.02672***
		(0.00727)	*(0.00708)*	*(0.00730)*	*(0.00701)*	*(0.00730)*	*(0.00697)*
	L(P_informal)	-0.14852***	0.05075***	-0.15840***	0.03496***	-0.17888***	0.07283***
		(0.02917)	*(0.00958)*	*(0.02352)*	*(0.00461)*	*(0.02409)*	*(0.00922)*
	L(P_formal)	0.07399***	-0.03026***	0.07161***	-0.02804***	0.07283***	-0.03112***
		(0.01154)	*(0.00670)*	*(0.00944)*	*(0.00468)*	*(0.00922)*	*(0.00462)*
Specification (4)	L(expend_Fon)	0.02051***	0.02688***	0.02067***	0.02680***	0.02051***	0.02698***
		(0.00727)	*(0.00710)*	*(0.00728)*	*(0.00692)*	*(0.00729)*	*(0.00690)*
	L(P_informal)	-0.15913***	0.04230***	-0.16980***	0.03218***	-0.18503***	0.06174***
		(0.02895)	*(0.00862)*	*(0.02340)*	*(0.00425)*	*(0.02383)*	*(0.00857)*
	L(P_formal)	0.06673***	-0.02606***	0.06594***	-0.02398***	0.06174***	-0.02566***
		(0.01065)	*(0.00639)*	*(0.00871)*	*(0.00437)*	*(0.00857)*	*(0.00424)*

Note: Robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. See footnote in Table VII.1 for additional details. Prices used here correspond to O1_uv in Appendix 6, with restriction 1, using the correspondent adjusted version of L(expend_Fon).

In sum, the general balance of evidence presented here reveals interesting patterns in characterizing formal and informal consumption in terms of their demand properties. First, purchases on both formal and informal markets show behaviour coherent with demand theory, with negative own-price elasticities for both Marshallian and Hicksian versions. Second, the results suggest that both markets are elastic in uncompensated terms: own-price elasticities around or greater than 1, with formal markets more responsive than informal ones, although this last result is not robust when the definition of prices changes. The evidence in utility constant terms is clearer. It was found that, in compensated terms, formal markets are more sensitive to own-price shocks than informal ones for all specifications and exercises presented here. In fact, it was verified that compensated demand curves for informal markets are price-inelastic, with their formal counterparts being price-elastic. Third, imperfect substitution across markets is supported by the data, but the computed effects are not equal. It was verified that formal markets are more sensitive to informal price shocks than the other way around. This conclusion holds for both compensated and uncompensated demands.

Table VII.6: Marginal effects for price vector UV-1 under different methods of imposing demand restrictions and specifications

		Expenditure Elasticities		Hicksian Elasticities				Marshallian Elasticities			
		Informal	Formal	Informal consumption		Formal consumption		Informal consumption		Formal consumption	
				Own-price	Cross-Price	Own-price	Cross-Price	Own-price	Cross-Price	Own-price	Cross-Price
Spec. (2)	R.1	1.021*	1.405***	-0.650***	0.186***	-1.531***	1.281***	-1.263***	0.117***	-1.626***	0.438***
		(0.01138)	*(0.10205)*	*(0.03992)*	*(0.01618)*	*(0.10323)*	*(0.10426)*	*(0.04021)*	*(0.01617)*	*(0.10663)*	*(0.11690)*
	R.2	1.021*	1.402***	-0.642***	0.185***	-1.480***	1.258***	-1.255***	0.116***	-1.575***	0.416***
		(0.01140)	*(0.09767)*	*(0.03675)*	*(0.01460)*	*(0.06958)*	*(0.06453)*	*(0.03672)*	*(0.01460)*	*(0.07139)*	*(0.07862)*
	R.3	1.021*	1.403***	-0.677***	0.192***	-1.532***	1.823***	-1.290***	0.123***	-1.626***	0.981***
		(0.01139)	*(0.09706)*	*(0.03750)*	*(0.01423)*	*(0.06794)*	*(0.12756)*	*(0.03741)*	*(0.01425)*	*(0.06987)*	*(0.12938)*
Spec. (3)	R.1	1.035***	1.390***	-0.647***	0.191***	-1.380***	1.351***	-1.268***	0.121***	-1.474***	0.517***
		(0.01211)	*(0.10479)*	*(0.04860)*	*(0.01922)*	*(0.09910)*	*(0.14182)*	*(0.04860)*	*(0.01925)*	*(0.10145)*	*(0.15252)*
	R.2	1.035***	1.391***	-0.636***	0.190***	-1.321***	1.356***	-1.257***	0.120***	-1.415***	0.521***
		(0.01215)	*(0.10371)*	*(0.04072)*	*(0.01566)*	*(0.06918)*	*(0.07008)*	*(0.04040)*	*(0.01570)*	*(0.07011)*	*(0.08314)*
	R.3	1.035***	1.395***	-0.670***	0.192***	-1.367***	1.904***	-1.291***	0.122***	-1.462***	1.067***
		(0.01216)	*(0.10316)*	*(0.04171)*	*(0.01531)*	*(0.06829)*	*(0.13785)*	*(0.04129)*	*(0.01536)*	*(0.06940)*	*(0.13893)*
Spec. (4)	R.1	1.034***	1.398***	-0.665***	0.179***	-1.318***	1.226***	-1.286***	0.109***	-1.413***	0.387***
		(0.01211)	*(0.10512)*	*(0.04822)*	*(0.01775)*	*(0.09460)*	*(0.12762)*	*(0.04835)*	*(0.01775)*	*(0.09735)*	*(0.13931)*
	R.2	1.034***	1.397***	-0.656***	0.180***	-1.265***	1.275***	-1.277***	0.111***	-1.359***	0.437***
		(0.01213)	*(0.10237)*	*(0.04046)*	*(0.01445)*	*(0.06456)*	*(0.06425)*	*(0.04031)*	*(0.01446)*	*(0.06589)*	*(0.08002)*
	R.3	1.034***	1.399***	-0.681***	0.173***	-1.291***	1.702***	-1.302***	0.104***	-1.385***	0.862***
		(0.01214)	*(0.10210)*	*(0.04122)*	*(0.01422)*	*(0.06272)*	*(0.12789)*	*(0.04101)*	*(0.01424)*	*(0.06426)*	*(0.13055)*

Note: Robust standard errors in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Tests for expenditure elasticities are in the form 'different to 1'. See footnote in Table VII.5 for additional details.

Therefore, we can argue that formal markets, in contrast to informal ones, are generally more sensitive to price shocks, at least for food consumption and under the assumptions made here. This is an expected result considering the higher importance of informal expenditure compared to formal. It is also an intuitive result, given the stylized facts that support the idea that informal consumption is a better substitute for formal purchases than the other way around. This can be explained if the variety of goods sold on informal markets is greater or if the purchasing possibilities offered there are more flexible, in order to better take into account consumers' needs. Some evidence on these lines is presented by de Soto (1986), Dasgupta (1992) and Böhme and Thiele (2012b). Similarly, if there are confidence links between market participants in the informal sector that better explain consumption, limited formal supply and relatively high price gaps compared to the formal sector then the differences in demand responses, as found here, are a reasonable result.

VII.4. Policy simulations and welfare analysis

The objective of this section is to analyse the welfare effects of certain government policy alternatives in order to formalize the economy using the relative price channel. To do this, the predicted changes in formal and informal consumption for the different policy scenarios are presented, as well as the welfare effects associated with them, in the form of compensated variations (CV). Predicted consumption is obtained from Marshallian elasticities and compensated variations from Hicksian elasticities for an arbitrary range of simulation of +/- 20% change on formal and informal prices. The analysis is restricted to the base-price vector (UV-1) using the base specification (1) and considering the three methods of imposing restrictions.

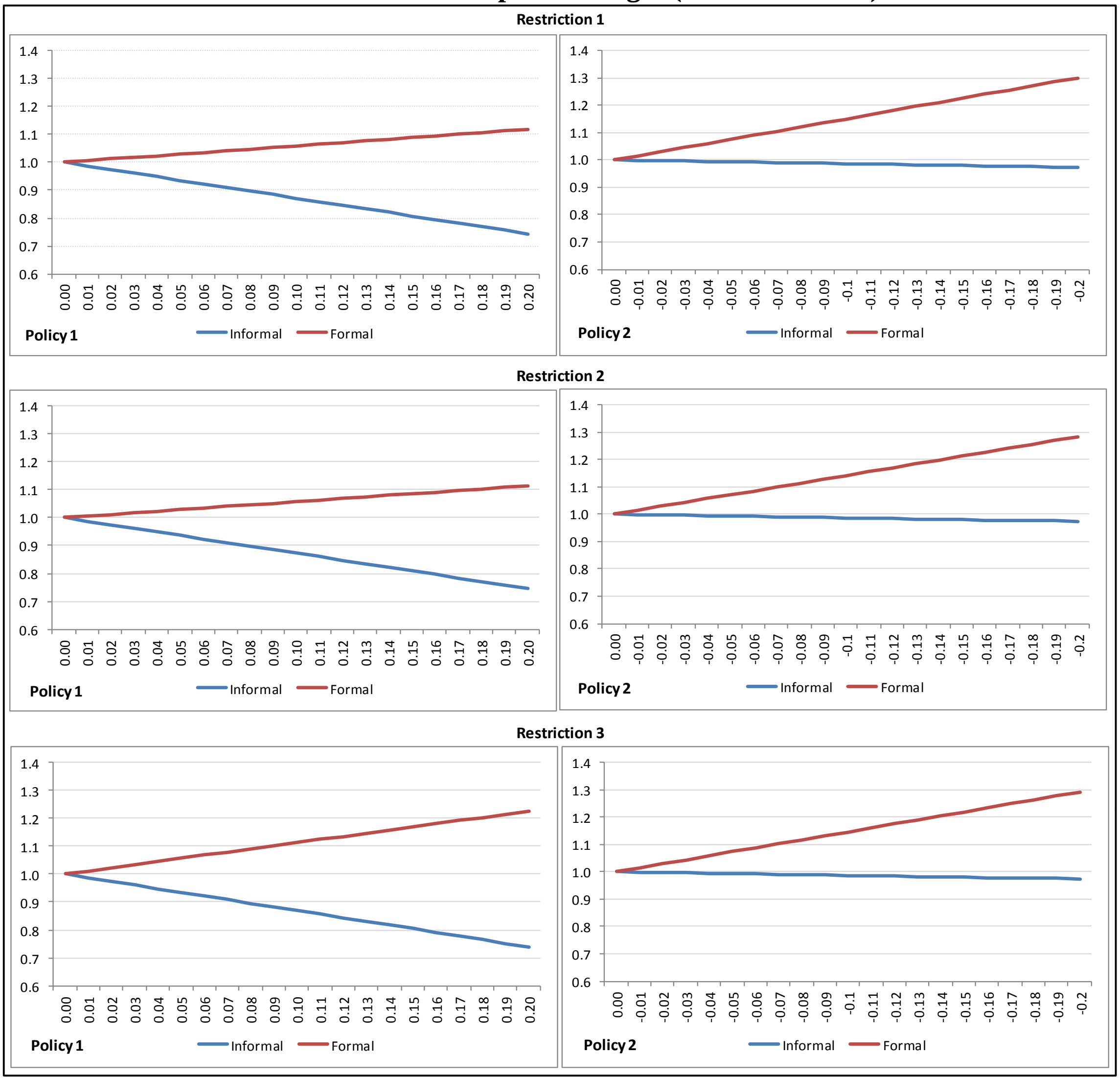
Two scenarios are constructed. The first (Policy 1) refers to an increase in informal prices, holding constant formal ones. A scenario like this could be understood in the context of local goverment enforcement activities designed to increase the detection risks and operational costs of informal sellers in both sidewalk commerce and city markets. Using anecdotal evidence and local news reports, we can argue that public policies like this are very common in Peruvian cities and are generally welcomed by the neighbours within the parameters of public ornament, safety and health considerations. Using the results presented here, it will be possible to measure the other side of this formalization policy: the negative welfare effect, given the price increase.

The second policy (Policy 2) refers to a price reduction in formal markets, holding constant informal prices.[96] A result like this could be a consequence of more competition in the supermarket segment, using policies to promote investments in the modern sector. Some alternatives (also discussed in de Soto, 1986) refer to the reduction of public barriers of entry in the form of less complicated and more predictable bureaucratic procedures, lower registration fees and quicker processing of licence requests. Also, as noted in Arellano and Burgos (2010), recent economic growth and household income expansion are helping the increase of the formal retail food market. Anecdotal evidence and market reports show that the supermarket sector has been one of the more dynamic ones, with a pattern of increase of 19 new outlets every year from 2005 (47 supermarkets) to 2014 (220 supermarkets).[97] This change has been faster in Lima than in other cities and it is probable that it will be faster in bigger cities than in smaller ones. If we consider similar market reports, we can see that this pattern will not stop in the short term, generating additional competition in this segment and the consequent price reduction. Using the results presented here it is possible to inspect the possible welfare effects of this development of the market.

[96] It is important to recognize that a scenario like the second one also implies some enforcement activities done by the government as long as it is expected that informal sellers could react to lower formal prices. The reaction function is not modelled here, so a result like this could not be explored properly. However, in preliminary exercises this possibility was (roughly) taken into account holding constant relative prices. This simulation is not presented here for the sake of brevity, but the results showed that allowing informal sellers to react could almost entirely offset the purposes of the policy.

[97] The figures presented here only consider the ‘big name’ supermarkets in the segment (Wong, Plaza Vea, Metro and Tottus).

Figure VII.1: Predicted consumption levels for formal and informal consumption for different scenarios of price changes (Policies 1 and 2)



Note: The horizontal axis refers to price changes and the vertical axis to predicted consumption levels normalized to 1 for the actual sample mean.

In Figure VII.1, the results of both policies are presented in terms of their formalization effectiveness. The panels on the left show the results of Policy 1 and those on the right the results for Policy 2. Note, also, that the figure is divided into three parts enabling consideration of the predictions under the different methods of imposing restrictions. The horizontal axis always refers to the different percentage changes in prices calculated from the actual mean values (formal or informal). The vertical axis is the predicted consumption levels, normalized to 1 for the sample mean value, which is used as the origin of the policy simulation.

As expected, policies that work through own-price mechanisms are more effective in reducing (or increasing) own consumption. For example, a 20% reduction in informal prices will lead to an almost 30% decrease on informal consumption. Similarly, a reduction of 20% in formal prices will generate an increase in formal consumption of a similar percentage. A result like this follows on from the previous discussion on Marshallian own-price elasticities: similar to and greater than 1 in both sectors. However, the side-effect generated by each policy is in fact different, in line with the differences encountered in cross-price effects. For example, the 20% shock on informal prices could generate increases in formal consumption of between 10% (under restrictions 1 and 2) and 20% (under restriction 3). These results are higher than the side-effects generated by formal prices, where informal consumption remains practically unchanged in all the estimations. Therefore, if the objective of the government is to promote formal and reduce informal consumption, when only one price instrument is used, the results fall in favour of policies that work through informal prices (policy 1).

Table VII.7: Computed compensating variations for Policy 1 and Policy 2

	Results from the formula						As a % of total expenditure					
	Restriction 1		Restriction 2		Restriction 3		Restriction 1		Restriction 2		Restriction 3	
	Policy 1	Policy 2	Policy 1	Policy 2	Policy 1	Policy 2	Policy 1	Policy 2	Policy 1	Policy 2	Policy 1	Policy 2
5%	0.029	-0.004	0.029	-0.004	0.029	-0.004	1.0%	-0.1%	1.0%	-0.1%	1.0%	-0.1%
10%	0.055	-0.008	0.055	-0.008	0.055	-0.008	1.9%	-0.3%	1.9%	-0.2%	1.9%	-0.3%
15%	0.080	-0.012	0.080	-0.012	0.080	-0.012	2.7%	-0.4%	2.7%	-0.4%	2.7%	-0.4%
20%	0.103	-0.017	0.103	-0.017	0.103	-0.017	3.6%	-0.6%	3.6%	-0.6%	3.6%	-0.6%
Decile 1	0.093	-0.006	0.093	-0.006	0.092	-0.006	4.5%	-0.3%	4.5%	-0.3%	4.5%	-0.3%
Decile 2	0.097	-0.007	0.097	-0.007	0.096	-0.007	4.6%	-0.3%	4.6%	-0.3%	4.5%	-0.3%
Decile 3	0.103	-0.008	0.103	-0.008	0.102	-0.008	4.5%	-0.3%	4.5%	-0.3%	4.5%	-0.3%
Decile 4	0.105	-0.010	0.105	-0.010	0.104	-0.010	4.4%	-0.4%	4.4%	-0.4%	4.3%	-0.4%
Decile 5	0.107	-0.011	0.107	-0.011	0.106	-0.011	4.2%	-0.4%	4.2%	-0.4%	4.2%	-0.4%
Decile 6	0.112	-0.013	0.112	-0.013	0.112	-0.013	4.3%	-0.5%	4.3%	-0.5%	4.2%	-0.5%
Decile 7	0.111	-0.014	0.111	-0.014	0.111	-0.014	4.1%	-0.5%	4.1%	-0.5%	4.1%	-0.5%
Decile 8	0.112	-0.021	0.112	-0.020	0.112	-0.020	3.9%	-0.7%	3.9%	-0.7%	3.9%	-0.7%
Decile 9	0.103	-0.029	0.103	-0.028	0.103	-0.028	3.1%	-0.8%	3.1%	-0.8%	3.0%	-0.8%
Decile 10	0.087	-0.057	0.087	-0.056	0.087	-0.057	2.0%	-1.2%	2.0%	-1.2%	2.0%	-1.2%

However, as long as a policy like this is based on price increases, we can expect there to be negative welfare effects. In Table VII.7, the results for the compensating variation estimation are presented. Following the definition of CV used here (the amount of money that needs to be transferred to a household in order for the latter to be indifferent to the shock experienced), positive values refer to negative welfare effects and negative values to positive ones. The calculations are done for four critical price changes (5%, 10%, 15% and 20%) for the whole sample; then the price change of 20% is used to inspect the distributional impacts. The CV is thus computed for the different deciles of total *per-capita* expenditure using the estimated elasticities and changing the budget shares. The first six columns of the table are the results of applying the formula (VII.7) and the final six are the income changes expressed as a percentage of total expenditure. As shown, the different restrictions imposed reveal no material differences in the computation of CVs.

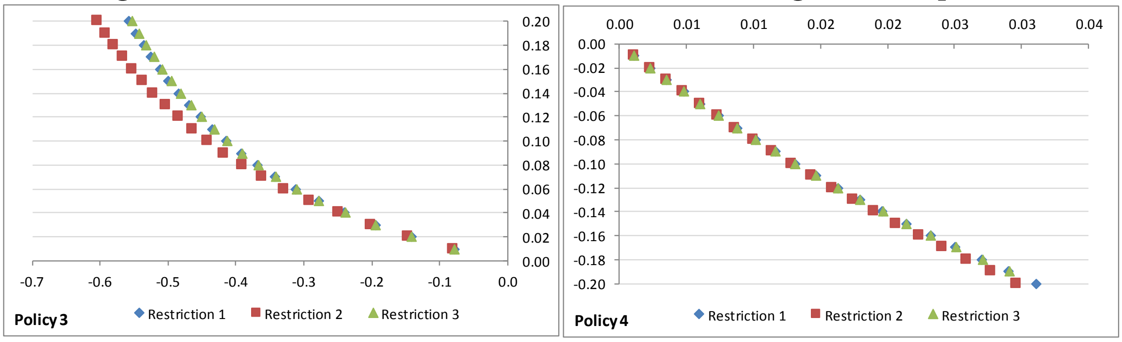
For interpretational purposes it is better to focus on the last six columns. In the top panel the results for the whole sample are considered. Under Policy 1, welfare losses are expected. For a price increase of 20%, on average the amount of money to be transferred to the households in order to offset their welfare loss is around 3.6% of total expenditure. By contrast, under Policy 2, as long as it is based on price reductions, welfare gains can be expected. The average savings for households in this case are around 0.6% of total expenditure. Interestingly enough, although the policies are based on similar price shocks, the welfare gains and losses are not equal: welfare gains for formal price reductions are lower in absolute terms than welfare losses when informal prices increase. This suggests that the welfare effects of informal prices are greater (in any direction) when compared to those produced by formal prices. This is an expected result because, although compensated price effects are greater for formal consumption, household consumption is biased in favour of informal markets.

At the bottom of the panel the distributional impacts are shown. The results suggest that the two policies will be unevenly distributed, with greater negative (positive) impacts in the case of Policy 1 (Policy 2) on the poorest (richest) individuals in the sample. A result like this is explained by the different sizes of informal consumption shares across the distribution; as stated in Chapter 5, they are highly correlated with poverty levels. Therefore, those households with high initial informal consumption (mostly the poor) will be more affected by a policy that increases informal prices and those households with higher formal consumption (mostly the rich) will benefit more from a policy that reduces formal prices. A result like this demonstrates the negative welfare consequences of using an increase in informal prices as a formalization policy.

However, it is obvious that a government could obtain better results than Policies 1 and 2. From average welfare gains in the case of Policy 2 and average welfare losses in the case of Policy 1, it is possible to imagine a wide range of intermediate situations that combine both policies in order to minimize the welfare impacts of formalization. From simulations displayed previously on the effect on total consumption, combined policies will also have a greater effect on both increasing and reducing formal and informal purchases, respectively. To construct these alternative scenarios, expression VII.7 is fixed at zero, and solved for price changes in the formal and informal sector. The result generates multiple possibilities; here only the two extremes are outlined. The first extreme is Policy 3, where price changes in the informal sector are set as before and price changes in the formal sector are found in order to make average household welfare invariant. The second extreme is Policy 4, which solves the problem the other way around: the formal price change is set as before and informal price adjustments are found.

The different combinations are shown in Figure VII.2. The panel on the left shows the results for Policy 3 and the panel on the right the results for Policy 4. In both cases the vertical axis is the fixed-price change. This means that, under Policy 3, the increases are between 1% and 20% in informal prices and, under Policy 4, the reductions are between 1% and 20% of formal ones. In the horizontal axis the compensating price changes are shown. Policy 3 shows the minimum reduction that needs to be implemented on formal prices to keep welfare unchanged. Policy 4 gives the maximum tolerable increases that could be implemented on informal prices without affecting welfare levels. In line with the differences in compensated own-price and cross-price elasticities between formal and informal, the price combinations will be different in these two scenarios.

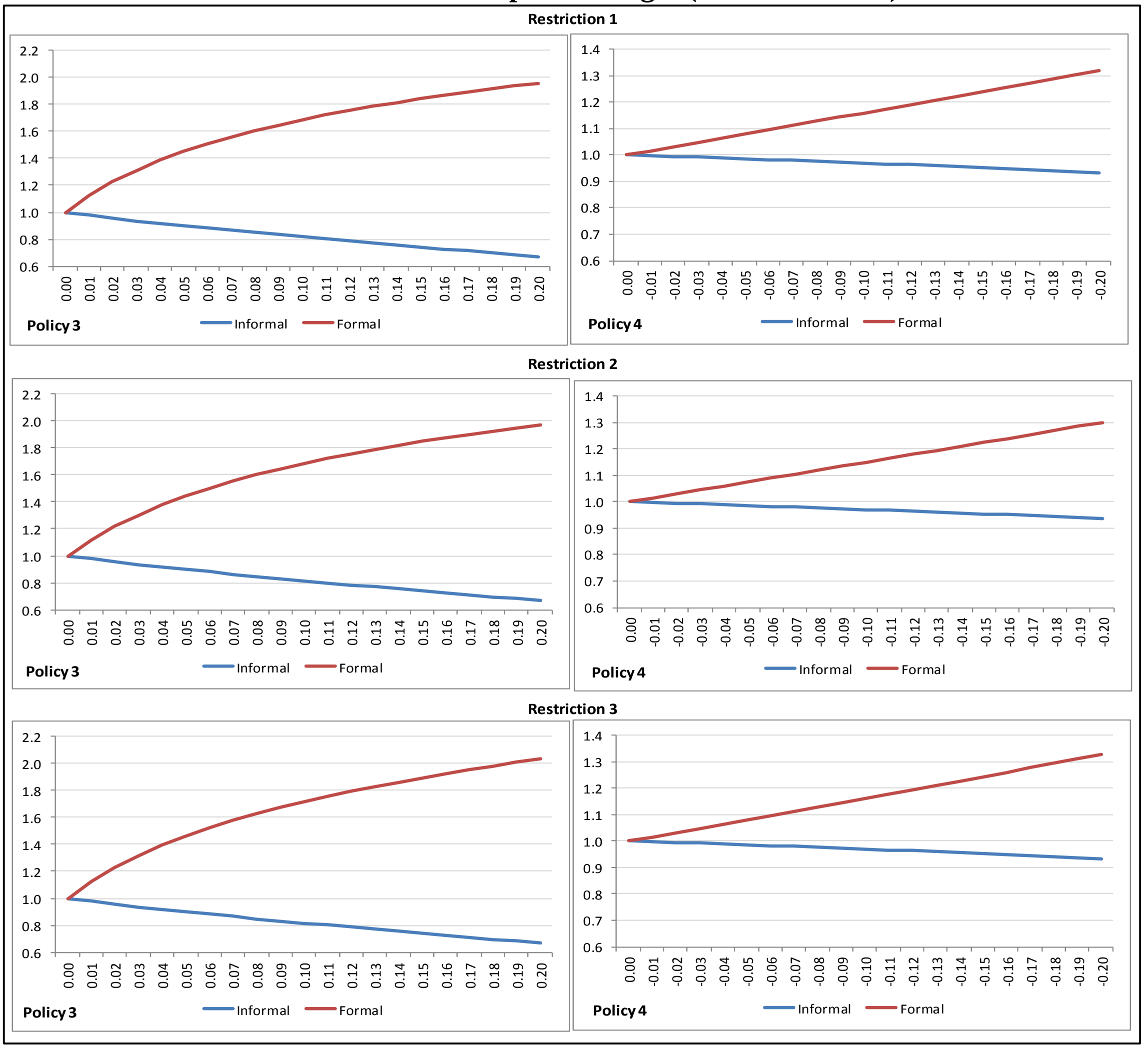
Figure VII.2: Combination of formal and informal changes that keep CV=0



Note: price-change combinations found setting equation (VII.6) to zero. Results found using the non-linear GRG (Generalized Reduction Gradient) algorithm. For Policy 3, the horizontal axis refers to changes in formal prices and the vertical axis to changes in informal ones. For Policy 4, the horizontal axis refers to changes in informal prices and the vertical axis to changes in formal ones.

The information provided is useful for policy design. For example, if the government designs a policy that allows a price increase in the informal sector of 10%, a reduction of almost 40% in formal prices is needed to keep household welfare (on average) unaltered. If the policy generates a 20% increase, the compensated reduction in formal prices climbs to almost 60%. By contrast, if new economic conditions (like increasing competence) or direct policies help to reduce formal prices in the economy by around 10%, this gives the government some space to increase informal prices via enforcement policies of around 1.5% and, if the price reduction is about 20%, the compensating price change climbs to 3%. The implications of the results are interesting since they reveal that even small changes in informal prices need higher changes in formal prices to keep welfare unaltered; conversely, even a strong reduction on formal prices will give a small space in which to increase informal prices. This is just another way to see the differentiated welfare effects produced by changing formal and informal prices.

Figure VII.3: Predicted consumption levels for formal and informal consumption for different scenarios of price changes (Policies 3 and 4)



Note: The horizontal axis refers to price changes and the vertical axis to predicted consumption levels normalized to 1 for the actual sample mean.

In Figure VII.3, both policies are used to simulate their effectiveness in the formalization of household consumption. It is obvious that, given the higher price change, the reductions on informal consumption and increase on its formal counterpart will be higher under Policy 3 when compared with any alternative. However, under Policy 4, the formalization of consumption also outperforms previous alternatives. Therefore, the results could be used to suggest the existence of a range of policies that work on moving both prices and that will not only produce lower welfare impacts, but will also improve the formalization objectives of the government. However, these policies are not free of costs. In Table VII.8, the distributional welfare effects are computed. As shown there, although the compensated variation is set to zero for the whole sample mean estimate, when the information is opened by deciles, we must acknowledge the existence of winners and losers under Policies 3 and 4.

Table VII.8: Computed compensating variations for Policy 3 and Policy 4

	Results from the formula						As a % of total expenditure					
	Restriction 1		Restriction 2		Restriction 3		Restriction 1		Restriction 2		Restriction 3	
	Policy 3	Policy 4	Policy 3	Policy 4	Policy 3	Policy 4	Policy 3	Policy 4	Policy 3	Policy 4	Policy 3	Policy 4
Decile 1	0.054	0.010	0.054	0.010	0.054	0.010	2.6%	0.5%	2.6%	0.5%	2.6%	0.5%
Decile 2	0.052	0.010	0.052	0.010	0.052	0.010	2.4%	0.4%	2.4%	0.4%	2.4%	0.5%
Decile 3	0.049	0.009	0.048	0.009	0.049	0.009	2.1%	0.4%	2.1%	0.4%	2.1%	0.4%
Decile 4	0.043	0.008	0.043	0.008	0.043	0.008	1.7%	0.3%	1.7%	0.3%	1.7%	0.3%
Decile 5	0.039	0.007	0.039	0.007	0.039	0.007	1.5%	0.3%	1.5%	0.3%	1.5%	0.3%
Decile 6	0.033	0.006	0.033	0.006	0.033	0.006	1.2%	0.2%	1.2%	0.2%	1.2%	0.2%
Decile 7	0.023	0.004	0.023	0.004	0.023	0.004	0.8%	0.2%	0.8%	0.1%	0.8%	0.2%
Decile 8	-0.008	-0.002	-0.008	-0.001	-0.008	-0.002	-0.3%	-0.1%	-0.3%	0.0%	-0.3%	-0.1%
Decile 9	-0.060	-0.011	-0.059	-0.011	-0.060	-0.011	-1.6%	-0.3%	-1.6%	-0.3%	-1.6%	-0.3%
Decile 10	-0.227	-0.043	-0.224	-0.041	-0.226	-0.043	-4.3%	-0.9%	-4.3%	-0.9%	-4.3%	-0.9%

Note: Computed for price changes of 20% with the corresponding adjustment.

As expected, the winners are those households located at the top end of the distribution, where welfare gains are concentrated; the losers are those located at the bottom, with positive values for CV. In fact, in the two first deciles (the poorest individuals), a 20% increase in informal prices and a reduction in formal ones of between 50% and 60% (Policy 3) will generate welfare loses of around 2.5% of total expenditure. By contrast, at the top end of the distribution (decile 10), the income gains are around 4.3% of total expenditure. Similarly, using Policy 4, a 20% reduction in formal prices, accompanied by the compensation of a 3% increase in informal prices, generates negative welfare effects for the poorest individuals (deciles 1 and 2) of around 0.5%, with welfare gains in decile 10 closer to 1% of total expenditure. Therefore, although Policies 3 and 4 improve the formalization objectives of the government, without average welfare effects they are unable to avoid the negative distributional effects of the formalization policy.

Moreover, even if the government use income transferences to reduce the negative distributional impacts, as analysed in Chapter 5, this could partially offset the original intention of the policy (reduce informality and increase formality). Therefore, although it is possible to suggest the existence of space for optimal policies that combine price and income mechanisms in order to fulfil formalization objectives without distributional consequences, the results of Chapters 5 and 7 suggest that this space is narrow so its design needs to be extremely carefully thought out. In the general case, it is better to keep in mind that formalization policies will cause negative welfare effects concentrated mainly on the poor.

VII.5. Concluding remarks

The inclusion of prices in the analysis helps us to reach two additional conclusions on informal consumption in developing countries. First, the results are coherent with expected demand patterns: negative compensated and uncompensated own-price elasticities and positive income ones. In general, we can clearly see own-price elastic observed demand curves for both formal and informal channels of commercialization, with point estimates around or above 1, something that could be used as preliminary evidence of high substitution possibilities across sectors. Considering utility constant demand curves, only formal markets prove to be own-price elastic, with informal ones being own-price inelastic. Compensated cross-price elasticities also confirm that formal and informal consumption can be treated as imperfect substitutes, with higher substitution effects in the case of formal consumption than of informal ones.

This last result is interesting since the data in general support the notion that formal markets are more sensitive to price shocks than informal ones. This is an intuitive result, if we think that informal channels of commercialization could be considered more flexible and diverse; so informal to formal substitution possibilities could be greater than the other way around. And with the presence of confidence links between market participants in the informal sector explaining at least part of the consumption patterns there, it seems obvious to expect such a result.

Second, in terms of the policy simulations undertaken here, policies that combine informal price increases with formal price reductions will naturally be the most effective in terms of formalization and will help to minimize the welfare impacts. In principle, informal price increases will help to reduce informal and to slightly increase formal consumption, while a formal price reduction will reinforce formal consumption promotion. At the same time, price-change combinations could be designed to maintain almost unaltered welfare effects, with price reductions in formal markets compensating increases on informal price. However, these policies, even when they are carefully designed, are not free of costs. In the cases presented in this application, the main cost is the unequal distribution effects that they will cause, with welfare losses for the poorest individuals in the sample and welfare gains for the richest. Such a result is explained by the uneven distribution of informal consumption shares across the income distribution – higher for poorer households.

VIII. INTRA-HOUSEHOLD BARGAINING AND INFORMAL CONSUMPTION

VIII.1. Introduction

The first three empirical models of informal consumption estimated in this research have been constructed on the assumption of the unitary utility maximization process. In this chapter this assumption is relaxed via the inclusion of bargaining structures in the demand analysis proposed up to now. The objective is to analyse whether it is possible to identify differentials in the preferences of the different household members in terms of resource allocation between formal (better quality) and informal (lesser quality) markets. It is reasonable to assume that it will be possible. For example, a common result in this literature (as shown in Thomas and Chen, 1994; Hoddinott and Haddad, 1995; Thomas, 1997; Phipps and Burton, 1998; Quisumbing and de la Brière, 2000; Quisumbing and Malucccio, 2000) is that some family members' (wives) expenditure is more pro-welfare than the expenditure of other members (husbands). This result has already been suggested by anthropological tradition (see, for example, Ortner, 1974; Rosaldo, 1974) where the role of the woman in the family (mother: closer to domestic activities and household needs) allows us to infer *ex-ante* this type of preference ordering. If this is true, a natural extension of the argument is that females' influence on resource allocation will not only be in the form of 'more' pro-welfare goods, but also of trying to improve the 'quality' of the purchasing basket of the family (i.e. reducing the share of products obtained from informal outlets, for example). This is the kind of hypothesis that will be investigated here.

The chapter uses an AIDS specification (Deaton and Muellbauer, 1980a) where a cooperative bargaining structure is imposed in the form of a sharing rule (Chiappori, 1988, 1992; Bourguignon *et al*., 1993; Chiappori, 1997; Browning and Chiappori, 1998) in order to test the effects of member income shares on the allocation of consumption across formal and informal markets. Three complementary empirical exercises are proposed.

The first exercise follows on from previous work by Lancaster *et al*. (2006) and Monge (2007), who derive theoretical household demands from individual demands leading to non-linear specifications of Working–Leser Engel functions. As discussed in Basu (2006) and Lancaster *et al*. (2006), it is plausible to assume that the effects of bargaining, indeed any bargaining, will depend on the relative level of power enjoyed by individuals. For example, as the above authors explain, the effects of bargaining on resource allocation will be different in a situation where power is balanced or when it is biased against one of the members. Linear approximations overlook this possibility and lose much of the information that this specification could provide. However, this approach makes any important econometric problems that might emerge less tractable – basically, endogeneity in the measure of power. This is the starting point for the second exercise, which uses a linear approximation of the Engel curve, as in Hoddinott and Haddad (1995). Here the authors simply extend the empirical household demand function, including a measure of bargaining power inside the family. The effects of bargaining can then be studied simply by checking the sign and significance of this additional variable. The third exercise refers to the disaggregated version of the model (similar to that presented in Chapter 6, but for collective models). As long as the previous two exercises are done for aggregated market consumption, an in-depth exploration inside consumption groups is needed in order to check whether bargaining on market allocation is more relevant for some goods than for others. This model is again estimated using a non-linear approximation.

The chapter is organized as follows. In Section 2, the theoretical background for collective models is presented in order to justify the inclusion of measures of bargaining power in a household's demand. Section 3 discusses the data used and the calculation of the main variables. In Section 4, the empirical specification and the econometric issues around the estimation strategy are presented in order to derive the effect of income distribution between family members on budget allocation. The estimation results are also discussed and then, in Section 5, the main conclusions of the application are summarized.

VIII.2. Theoretical background[98]

As discussed in the surveys presented in Haddad *et al*. (1997), Xu (2007) and Doss (2011) from both a theoretical and empirical point of view, unitary models seem to be less accurate when analysing the consumption decision process of a family if compared, for example, with collective approaches. In the former, the household is assumed to act as a unique and homogenous decision-making agent while, in the latter, it is recognized as a complex organization of individuals who eventually need to bargain in order to reach common decision choices (Alderman *et al*., 1995).

The simplest version of the collective models is the sharing rule approach. The framework is based on the notion of cooperation, used in a game theory sense, and the idea that, in such a decision-making process, interaction will produce Pareto-efficient results. The basic assumption of the model is then that households will never adopt decisions that are Pareto-dominated (Chiappori, 1997). However, instead of being explicitly modelled, household behaviour is assumed to be in the form of an optimal distribution of total income between family members that

[98] The theoretical discussion is based on my MA thesis (University of Sussex), included in the references as Monge (2007). This article was partially published after completion of my MA in *Revista Apuntes* No. 55, second semester of 2004 (Universidad del Pacífico), given calendar delays in the review. Supporting documentation of this is available on request.

will produce the expected outcomes. As originally proposed by Chiappori (1988 and 1992), Bourguignon *et al*. (1993), Browning *et al*. (1994), Chiappori (1997) and Browning and Chiappori (1998), the model departs from a household composed by k members, a set of goods (x_i), a set of prices (p_i) and individual determinants like λ_j. Then preferences will be represented by:

$$\begin{aligned} U_1 &= U_1(x_{1i}, \lambda_1) \\ U_2 &= U_2(x_{2i}, \lambda_2) \\ &\ldots \\ U_k &= U_k(x_{ki}, \lambda_k) \end{aligned} \qquad \text{(VIII.1)}$$

In order to allow for interaction between members, the maximization problem within this framework is solved in a two-step procedure (Lancaster *et al*., 2003). In the first step, the overall household income ($M = \sum_{j=1}^{k} Y_j$) is pooled and distributed between the members. The amounts $Y_j^* = F_j(M, p_1, p_2, ..., p_q)$ received by each member are known as the sharing rule, which depends on total income and prices. These amounts are set according to the Pareto efficiency property (each member agrees upon the amount he or she is permitted to spend). In the second stage, once the sharing rule is identified, each individual solves his own maximization problem in the form

$$\text{Max} \quad V_j = V_j[U_1(x_{1i}, \lambda_1), U_2(x_{2i}, \lambda_2), ..., U_k(x_{ki}, \lambda_k)] \qquad \text{(VIII.2)}$$

$$\text{s.t.} \quad \sum_{i=1}^{q} p_{ji} x_{ji} = Y_j^*$$

which is a caring or non-paternalistic representation of preferences for each individual. As shown by Bourguignon *et al*. (1993) and Browning and Chiappori (1998), this maximization problem within an assumption of efficiency can be simplified by

$$\text{Max} \quad W = \sum_{j=1}^{k} \theta_j W_j(V_1, V_2, ..., V_k) \qquad \text{(VIII.3)}$$

$$\text{s.t.} \quad M = \sum_{j=1}^{k} Y_j^* = \sum_{i=1}^{q} p_{1i} x_{1i} + \sum_{i=1}^{q} p_{2i} x_{2i} + ... + \sum_{i=1}^{q} p_{ki} x_{ki}$$

$$\sum_{j=1}^{k} \theta_j = 1$$

where θ_j are scalars that lie between 0 and 1 and are set arbitrarily by the household depending on the kind of cooperative agreement reached. A particular point on the Pareto frontier represents specific values for θ_j which correspond to the weight of given to each member for the welfare of the household in this particular situation. Under such conditions and considering k household members, a possible sharing rule will take the form $Y_j^* = (\theta_j)(M)$. In substituting and solving the maximization problem, the household demand functions achieved take the form:

$$\sum_{j=1}^{k} x_{ji} = x_i^* = x_i(Y_j^*, p_1, p_2, ..., p_q; \lambda_j) = x_i[(\theta_j)M, p_1, p_2, ..., p_q; \lambda_j] \tag{VIII.4}$$

The basic distinction between functions (VIII.4) and those presented in Chapter 3 is the introduction of the sharing rule as an additional argument; this is simply the way in which the distributional effects are recognized by the model. In fact, the term θ_j in a standard game theory framework can be considered as the true measure of the bargaining power of a particular member or the individual's ability to impose his/her decisions on the house. As noted by Sen (1985), Mencher (1988), Hoddinott and Haddad (1995), Riley (1997), Basu (2006) and Lancaster *et al*. (2006), this ability will depend on the actual contribution of each member to the family budget. Therefore, as will be seen later, in empirical terms θ_j can be interpreted as the income share of each of the j household members.

The marriage-market or divorce-threat approach extends the sharing rule model in order to give an economic rationale to the agreement reached and to explicitly model the behavioural characteristics of the household bargaining process.[99] According to Manser and Brown (1980) and McElroy and Horney (1981), individuals can either reach a binding agreement between different options or they can disagree. In the former case, each member cooperates and allocates the resources in a way that maximizes their collective utility function. In the latter case, a threat point is characterized by the payoffs associated with disagreement (divorce situation). As a result, using standard game-theory solutions, the bargaining power will be determined by the threat-point shifters or, in the words of McElroy (1997), the extra-household environmental conditions (EEP). Therefore, in the context of the sharing rule approach, this means that $\theta_j = \theta_j(\delta_j)$ with

[99] A good discussion of the determinants of the equilibrium in marriage markets is provided in Lundberg and Pollak (1993, 1996) and McElroy (1997).

δ_j being the EEPs (Browning *et al*., 1994; Chiappori, 1997); i.e., the distribution of power depends on the general conditions faced by individuals on the marriage and re-marriage market.

In an alternative variant to these types of cooperative equilibrium, Lundberg and Pollak (1993) suggest that cooperative solutions need credible threats and for a day-by-day negotiation a divorce or decision to leave the household could be a non-credible threat. For this reason, in their separate spheres' model, the authors specify a non-cooperative Pareto-inefficient solution (with the inefficient provision of public goods) as the relevant threat point. Making the threat point internal to the household has important empirical implications, as it will generate demand functions that, in some circumstances, depend not on who receive an income after a household's dissolution, but on who actually receives an income within the family (Lundberg and Pollak, 1996). In such circumstances, δ_j could be supplemented by actual or observable characteristics of individuals beyond EEPs.

The non-cooperative solution is the starting point of Ulph's (1988) proposal. One implication of this model is that control over money in the household does not happen in a 'sharing rule' context; by contrast, individual budgeting as a reaction to other household members' decisions is a more accurate notion. At the same time, the idea of a collective utility function no longer applies, but the idea of power (i.e. the ability of a member to impose preferences) is still relevant. In the original derivation of the model, Ulph (1988) proposed the share of income that belongs to a particular member as a true measure of bargaining power. Using this definition, Ulph (1988) and Hoddinott and Haddad (1995) point out three interesting features of the model. First, if the income is skewed toward a specific member, we can see that the Nash equilibrium reached will show that individual's preferred allocation of resources. Second, as the income share of an individual rises, the goods preferred by that individual will rise, that of the other individual will fall and that of public goods will fall or increase depending on which individual more greatly dislikes the preferred goods of the other member. Third, if members strongly disagree on purchasing, they can make strategic purchases to pre-commit the household to a minimum level of consumption.

In sum, the demand analysis proposed by collective models in any of their variants recognizes that not only does total income influence demand, but also that whoever controls the money within the household can actually modify the resource allocation. This is possible, given the introduction of bargaining processes in the analysis of household expenditure, which provides several new insights into the nature of the utility maximization problem of the family. The theoretical basis for such collective models is built in contrast to the unsatisfactory answers given by the unitary models and thanks to the analytical tools derived from game theory. As a result, from a public policy perspective, several new options for demand-side policies are opened via, for example, targeted subsidies to particular family members, who presumably show preferences for socially desirable goods.

Formalization policies could exploit these distributional effects on demand if it were possible to demonstrate that family members bargain not only between goods, but also on the allocation of resources between different markets (less-informal or more-formal). The objective now is to investigate this issue. To do so, our previous discussion needs to be slightly modified, basically so that we start thinking of collective models not in terms of goods, but of markets. Therefore, the term x_{ji} used in this discussion is no longer the good i purchased by member j; instead it is the budget that member j allocates to market i ($i =$ formal, informal). This is the way that the model will be empirically estimated.

VIII.3. The Data

The database considered in this chapter is similar to that already used in the estimations of previous models (Chapters 5 to 7) and explained in detail in Chapter 4. As stated there, the total sample size for urban households is 10,808 observations. However, in order to fit with the theoretical discussion presented above, the size has been restricted in several ways. The restrictions imposed are similar to those used in Bourguignon *et al*. (1993), Hoddinott and Haddad (1995), Phipps and Burton (1998), Lancaster *et al*. (2006) and Monge (2007). Here, the selected families are those where the two parents (mother and father) are present and live together (married or unmarried), with each aged between 18 (adult) and 65 (retirement) years old and at least one having a positive income. Other members (sons, daughters, etc.) may or may not exist, but they cannot be income-earners. Therefore, in terms of the theoretical models presented above, in empirical terms a two-member-family decision-making structure is assumed.

The data restrictions introduced imply a reduction in the sample size to 3,159 observations, so it would seem reasonable here to record possible changes in the dependent variables: consumption on goods and markets. In Table VIII.1, total expenditure (purchasing) by consumption group for the unrestricted and restricted samples is presented. As shown, there are no important differences in terms of the mean value of total expenditure or of the distribution between groups. In Table VIII.2 we have similar comparisons, but the definitions of the markets used are presented. With the three- and five-market definitions, again the figures seem to be highly comparable, with a small redistribution away from informal and formal markets and in favour of the semi-formal category. However, budget shares still reveal that the major allocation is to informal markets when compared with formal ones. Similarly, levels of censoring in this case (aggregate markets) can still be neglected as long as they are below 1% in all cases.

Table VIII.1: Total expenditure (purchasing) by consumption group

Total expenditure (purchasing) by group

	Unrestricted sample				Restricted sample			
	Mean	**Prop. (%)**	**Bud. Share**	**% > 0**	**Mean**	**Prop. (%)**	**Bud. Share**	**% > 0**
Fon	4,856	32.7	0.362	97.2	5,086	34.9	0.399	98.2
	(3348.43)		*(0.17)*		*(2990.04)*		*(0.16)*	
Foff	1,790	12.0	0.130	85.4	1,389	9.5	0.099	84.6
	(2109.29)		*(0.14)*		*(1744.48)*		*(0.11)*	
CC	1,221	8.2	0.083	96.3	1,302	8.9	0.090	98.5
	(1391.87)		*(0.06)*		*(1388.13)*		*(0.06)*	
HEA	887	6.0	0.053	82.1	771	5.3	0.048	81.9
	(1761.88)		*(0.08)*		*(1732.53)*		*(0.07)*	
TC	2,316	15.6	0.136	94.5	2,142	14.7	0.121	94.8
	(4449.06)		*(0.1)*		*(6875.47)*		*(0.09)*	
ED	1,304	8.8	0.066	86.7	1,258	8.6	0.067	91.2
	(2721.7)		*(0.08)*		*(2581.86)*		*(0.08)*	
OT	2,494	16.8	0.172	99.6	2,611	17.9	0.176	99.9
	(3146.76)		*(0.1)*		*(3492.58)*		*(0.1)*	
Total (purch.)	**14,868**	**100.0**		**100.0**	**14,558**	**100.0**		**100.0**
	(12172.17)				*(13275.37)*			
Number obs.	**10,808**				**3,159**			

Note: expenditure in annual S/. (*Nuevos Soles*). Standard deviations in parentheses.

Table VIII.2: Total expenditure (purchasing) by markets (different definitions)

	Unrestricted sample				Restricted sample			
	Three-market approach				Three-market approach			
	Mean	**Prop. (%)**	**Bud. Share**	**% > 0**	**Mean**	**Prop. (%)**	**Bud. Share**	**% > 0**
Informal	5,496	37.0	0.390	99.4	5,367	36.9	0.401	99.8
	(4221.03)		*(0.18)*		*(3804.19)*		*(0.18)*	
Semi-formal	4,213	28.3	0.318	99.8	4,396	30.2	0.335	100.0
	(3930.)		*(0.19)*		*(4004.54)*		*(0.19)*	
Formal	5,158	34.7	0.292	99.1	4,795	32.9	0.264	99.9
	(7734.17)		*(0.18)*		*(9762.17)*		*(0.17)*	
Total (purch.)	**14,868**	**100.0**			**14,558**	**100.0**		
	(12172.17)				*(13275.37)*			
	Five-market approach				Five-market approach			
	Mean	**Prop. (%)**	**Bud. Share**	**% > 0**	**Mean**	**Prop. (%)**	**Bud. Share**	**% > 0**
Market 1	2,398	16.1	0.174	98.4	2,286	15.7	0.171	98.9
	(2323.02)		*(0.13)*		*(2207.32)*		*(0.13)*	
Market 2	3,098	20.8	0.216	90.4	3,081	21.2	0.230	92.8
	(3014.57)		*(0.17)*		*(2773.35)*		*(0.18)*	
Market 3	4,213	28.3	0.318	99.8	4,396	30.2	0.335	100.0
	(3930.)		*(0.19)*		*(4004.54)*		*(0.19)*	
Market 4	1,584	10.7	0.101	88.8	1,386	9.5	0.087	88.4
	(2115.82)		*(0.11)*		*(2008.18)*		*(0.09)*	
Market 5	3,575	24.0	0.191	98.0	3,410	23.4	0.177	99.5
	(6692.99)		*(0.15)*		*(8967.76)*		*(0.14)*	
Total (purch.)	**14,868**	**100.0**			**14,558**	**100.0**		
	(12172.17)				*(13275.37)*			
Number obs.	**10,808**				**3,159**			

Note: expenditure in annual S/. (*Nuevos Soles*). Standard deviations in parentheses.

In Table VIII.3, the information on expenditure across markets within each of the consumption groups is presented. As a consequence of the data restrictions imposed, the sample sizes for the disaggregated models have also fallen but note that, again, the sample means of consumption are highly comparable between the restricted and the unrestricted samples. Some other important changes to note are the growth in the consumption levels of Fon, CC and OT and a reduction in those of the other categories. And in terms of the markets within each consumption group, Fon, Foff and TC show a re-composition away from informal consumption, and ED away from formal consumption. Other consumption categories maintain their distribution between formal and informal markets almost unaltered. Similarly, we should note that censoring at this level of analysis is still an important issue that must be considered during estimation.

Table VIII.3: Purchasing markets by expenditure group

	Unrestricted sample					Restricted sample				
	Mean	Prop. (%)	B. Share	% > 0	Sample	Mean	Prop. (%)	B. Share	% > 0	Sample
Fon (>0)	**4,998**	**100.0**			**10,501**	**5,180**	**100.0**			**3,102**
Informal	3,153	63.1	0.600	96.6		3,124	60.3	0.597	97.9	
Semi-formal	1,388	27.8	0.329	92.9		1,578	30.5	0.337	93.6	
Formal	457	9.1	0.071	47.3		477	9.2	0.066	48.6	
Foff (>0)	**2,097**	**100.0**			**9,226**	**1,642**	**100.0**			**2,671**
Informal	396	18.9	0.264	57.0		289	17.6	0.277	53.4	
Semi-formal	607	29.0	0.212	31.8		436	26.5	0.182	25.0	
Formal	1,093	52.1	0.524	76.6		918	55.9	0.542	73.3	
CC (>0)	**1,267**	**100.0**			**10,412**	**1,322**	**100.0**			**3,112**
Informal	502	39.6	0.416	80.5		539	40.8	0.444	84.5	
Semi-formal	661	52.2	0.510	93.7		677	51.2	0.486	94.2	
Formal	104	8.2	0.073	29.8		106	8.0	0.070	31.7	
HEA (>0)	**1,080**	**100.0**			**8,878**	**942**	**100.0**			**2,586**
Informal	70	6.5	0.065	18.7		60	6.3	0.067	17.2	
Semi-formal	83	7.7	0.154	38.1		74	7.9	0.157	37.0	
Formal	927	85.8	0.781	89.4		808	85.8	0.776	89.2	
TC (>0)	**2,449**	**100.0**			**10,218**	**2,258**	**100.0**			**2,996**
Informal	943	38.5	0.467	89.0		799	35.4	0.472	87.5	
Semi-formal	157	6.4	0.095	62.5		107	4.8	0.073	54.7	
Formal	1,350	55.1	0.438	82.9		1,351	59.8	0.455	82.1	
ED (>0)	**1,504**	**100.0**			**9,374**	**1,380**	**100.0**			**2,880**
Informal	152	10.1	0.288	73.0		166	12.0	0.280	74.2	
Semi-formal	232	15.4	0.241	68.7		229	16.6	0.274	72.7	
Formal	1,119	74.5	0.471	85.6		985	71.4	0.446	91.1	
OT (>0)	**2,503**	**100.0**			**10,770**	**2,614**	**100.0**			**3,155**
Informal	532	21.3	0.199	85.2		567	21.7	0.205	87.6	
Semi-formal	1,296	51.8	0.495	98.7		1,441	55.1	0.523	99.3	
Formal	675	27.0	0.306	92.4		606	23.2	0.272	93.7	

Note: expenditure in annual S/. (*Nuevos Soles*). (>0) means that calculations are restricted to positive expenditure in the particular group.

A second important issue to consider for the empirical strategy in this chapter is the construction of individual incomes. As discussed in the previous section, the specification of the demand equation will exploit the notion of bargaining power interpreted as the control over household resources that each decision-making member of the family has. In this application, control over money is defined as the share of income of the husband (or father). Since the database used here identifies individual incomes, this variable was computed with minimal assumptions and following the National Statistics Institute methodology (INEI, in Spanish). To do this, the income module of the survey was used, whereby the different sources of income are recorded for each member.

Sources of income were classified as either assignable or non-assignable. The assignable income identified was labour income (monetary and non-monetary) from different labour options (wage-earners and self-employment) and from primary and secondary (moonlighting) activities. Remittances (national and foreign) received by each member from other households or from institutions (including the government) were also individually recorded, as were extraordinary incomes such as individual property rents or lottery gains, insurance, heritages, etc. However, as the definition of consumption used here is purchasing, only monetary incomes were used in the calculation of power. Non-monetary sources identified were payments or transfers in kind, and own-consumption, already excluded from consumption measures as they cannot be classified in terms of formal and informal consumption. Non-assignable sources of income identified were the imputed consumption of the household, imputed property rents and other government transfers in kind not recorded individually. The low level of importance of these sources to total income determines that their exclusion in the computation of power has little effect on the results.

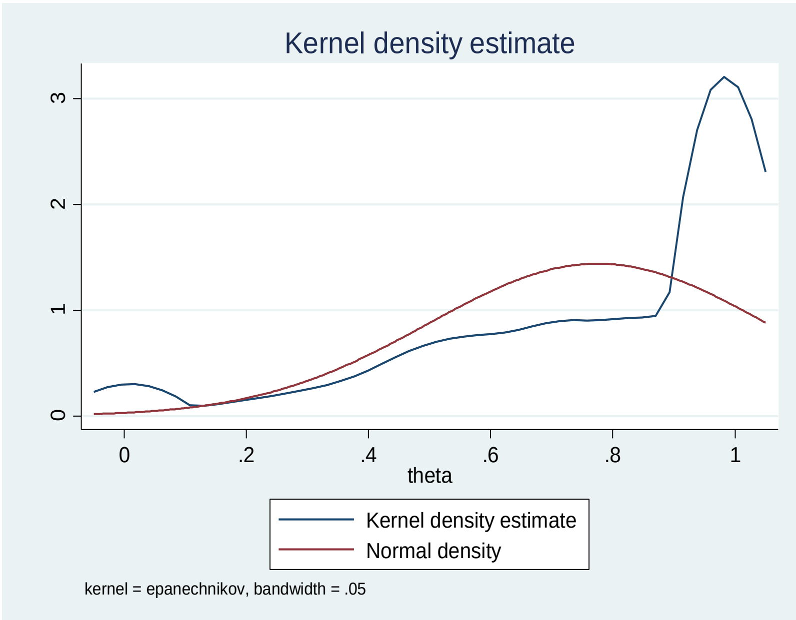
Table VIII.4: Total household income by source

	Unrestricted Sample		Restricted Sample	
	S/., annual	**% structure**	**S/., annual**	**% structure**
Total income	22,320		19,918	
	(21704.200)		*(20458.990)*	
Assignable	18,743	84.0	17,001	85.4
	(19525.710)		*(18223.980)*	
Monetary	17,354	92.6	15,935	93.7
	(19125.010)		*(17917.520)*	
Father	8,726	50.3	12,104	76.0
	(14292.110)		*(14741.740)*	
Mother	3,965	22.8	3,832	24.0
	(7602.009)		*(7436.498)*	
Son/Daughter	3,246	18.7	-	-
	(7651.893)			
Other	1,418	8.2	-	-
	(5121.533)			
Non-monetary	1,389	7.4	1,066	6.3
	(2205.097)		*(2039.443)*	
Non-assignable	3,577	16.0	2,917	14.6
	(4943.170)		*(4668.033)*	
Number obs.	**10,808**		**3,159**	

Note: Standard deviations in parentheses.

In Table VIII.4, the values for each source of income identified in the sample are presented for both the unrestricted and the restricted sample. As shown, in the unrestricted sample, total annual incomes are almost S/. 22,000, 85% of which could be assigned between members. From the total assignable income, 93% is monetary (source of the computation of bargaining power) so, in general terms, around 80% of total income was included in the estimation proposed. For household members, the major proportion of monetary-assignable income comes from the father (or husband), who accounts for almost 50% of this source. Lower values are identified for the other members: mother (23%), son/daughter (19%) and others (8%). Once the restrictions are imposed, total income falls to around S/. 20,000. As expected, there is also a re-composition of the structure of the income. Although, assignable income is still around 85% and monetary income is above 90%, control over money in the household changes in favour of the father. Now the monetary-assignable income of this member accounts for 75%, with that of mothers around 25%. Interestingly enough, these figures are similar to those of Monge (2007), who uses a different data sample and includes both monetary and non-monetary assignable income for the urban and rural areas in his calculation.

Figure VIII.1: Kernel density function for bargaining power and basic statistics



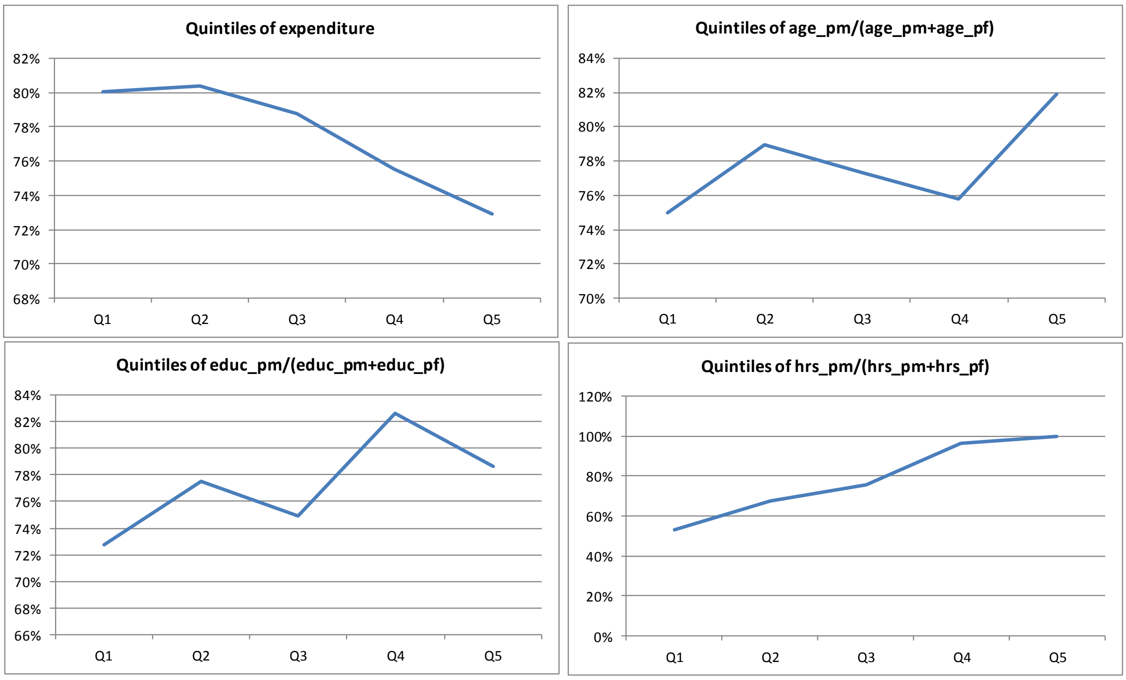
Basic statistics	
Mean	0.775
Median	0.894
Mode	1.000
Min	0.000
Max	1.000
S.D.	0.277
Variance	0.770
Skewness	-1.242
Kurtosis	3.739
P10%	0.382
P25%	0.612
P50%	0.894
P75%	1.000
P90%	1.000

Note: calculations based on the 3,159 observations of the restricted sample.

As mentioned, the variable of bargaining power is defined as the monetary-assignable income of the husband as a proportion of total monetary-assignable income. The basic statistics of this new variable introduced in the research are presented in Figure VIII.1, where the kernel density function and the normal approximation are also displayed. As shown, although some concentration of values is identified at around 0.0 and 0.5, the bargaining power used in the sample is highly skewed to the right, with an important concentration of values around 1 (almost 40% of the sample). This result is to be expected given that, in a developing country like Peru, the main earners in the household are typically men with better salaried jobs. By contrast, the labour inactivity of women determines, in most cases a monetary income very close to zero. The mean of the variable at the household level is around 0.78 and the median around 0.89. A similar distribution of power is identified in Hoddinott and Haddad (1995) for Côte d’Ivoire.

When this variable is correlated with total expenditure and other possible measures of power (like age, education and hours of work), the results which emerge are as expected. In Figure VIII.2, the results plot the average value of the bargaining power defined earlier for each of the quintiles defined for the variables of interest. As shown, as the family becomes richer, control over the father's resources is lower. There is also a positive correlation of the measure of power with the age and education gaps, although the relationship displayed in these cases is less clear. However, it should be noted that the differences across quintiles are small. For example, for the poorest households, the husband's share of the income is around 80% and for the richest the value is around 73%. Similarly, for those households with the highest age or education gaps, the extent of the change in the bargaining power variable goes from 75% to 82% and from 73% to 83%, respectively. However, it is different for hours of work where, as expected, both the positive correlation and the range of values are greater.

Figure VIII.2: Correlations of bargaining power with expenditure, age, education and hours of work



Note: the vertical axis refers to the bargaining power of the husband (in %) and the horizontal axis the quintiles for each variable considered.

VIII.4. Empirical specification and econometric issues

The empirical strategy used in this application derives from conclusions arrived at in Monge (2007), where the demand analysis presented by the author shows evidence in favour of a cooperative equilibrium for Peruvian households: income is not pooled between family members and the implied bargaining process attains Pareto-efficient outcomes. Under such circumstances, as explained in Lancaster *et al*. (2006), it is justifiable to use a sharing rule model in order to investigate the distributional effects on demand. As a result, the demand equation for market i will be (VIII.4). The objective now is to find a suitable specification for it. In order to do so and to maintain coherence in the specifications used thus far, Working–Leser Engel curves[100] will be used. In its most basic version, this specification takes the form:

$$\begin{aligned} w_a &= \alpha_a + \beta_a \ln(\theta M) + \varepsilon_a \\ w_b &= \alpha_b + \beta_b \ln[(1-\theta)M] + \varepsilon_b \end{aligned} \qquad \text{(VIII.5)}$$

for household members a and b (husbands and wives, respectively) where $w_a = C_a / Y_a^*$ and $w_b = C_b / Y_b^*$, with $Y_a^* = \theta M$ and $Y_b^* = (1-\theta)M$ being the shared expenditures for C_a and C_b being the formal or informal consumption levels for each member. In this setting, M is total expenditure, θ the measure of bargaining power and $\alpha_a, \alpha_b, \beta_a, \beta_b$ the set of estimable parameters. However, as individual consumption is not observed, the expression for household formal or informal consumption shares (w) is obtained, as in Lancaster *et al*. (2006) and Monge (2007), with the identity $w = \theta w_a + (1-\theta) w_b$. It is therefore possible to express it as:

$$\begin{aligned} & w = \alpha_b + (\alpha_a - \alpha_b)\theta + \beta_a \theta \ln(\theta M) + \beta_b (1-\theta) \ln[(1-\theta)M] + \ldots \\ & \ldots + \beta_1 z + \beta_2 \lambda + \varepsilon \end{aligned} \qquad \text{(VIII.6)}$$

which is the estimable equation used in this chapter for the non-linear model, where $\varepsilon = \theta \varepsilon_m + (1-\theta)\varepsilon_f$ and some household controls (z) and individual characteristics (λ) are already included in the pragmatic form suggested for the general case in Deaton (1997). The additional variables included in the estimation are similar to those already used in earlier chapters, but with some modifications in order to recognize individual characteristics and household controls. The base model includes demographics such as education, age, chronic illness and hours of work (and participation dummies). Four different specifications are then used: (1) the base model, (2) the base model with location variables, (3) the base model with cultural backgrounds and (4) the base model with enforcement variables. However, some difference with previous specifications should be noted. First, in terms of demographics, the dummy variable that identifies the existence of the spouse is dropped from the sample selection, as explained in the previous section. Second, both the education and the age of the head of the household are dropped since they are introduced as individual characteristics. Third, hours of work and participation dummies previously defined for seven different earning opportunities are now aggregated and included

[100] As in Chapter 5, the Engel curves used here are derived from the AIDS demand model, but neglecting the influence of prices. This means that the assumption that all households face the same price vectors is maintained. However, in contrast to those models, here total instead of *per-capita* expenditure is used, as it is less obvious how to impose distribution in a *per-capita* specification.

separately for both the father and the mother for the same reason. Fourth, as in Hoddinott and Haddad (1995) and Lancaster *et al*. (2006), the demographic structure of the household in terms of cohort proportions is introduced. The description of the variables and their basic tabulation (means and standard deviations) for the fully usable sample are shown in Appendix 8.

Beyond the general specification of the model, in order to estimate (VIII.6), a functional form for θ is needed. As mentioned, one proxy proposed in the literature is the observed control over resources computed as the ratio of the husband's income with total income (see Hayashi, 1995; Hoddinott and Haddad, 1995; Lancaster *et al*., 2006). Therefore, in line with earlier discussions, in this case θ takes the form $Y_a/(Y_a + Y_b)$, i.e. a function of observed incomes. Thus, the basic empirical equation to be estimated shows that bargaining power has a direct effect on demand allocation and an indirect effect via the sharing rule imposed. This issue is extensively discussed in Basu (2006) and Lancaster *et al*. (2006) and takes into account the fact that the effects of bargaining (or any bargaining at all) will depend on the distribution of power within the household.

However, this issue makes interpretation of the coefficients less transparent, so further computation is needed to reach to an adequate characterization of member-specific demand behaviour. The exercises proposed in Lancaster *et al*. (2006) are used to compute the marginal effect of θ on the budget share (w). This expression will take the form:

$$\frac{\partial w}{\partial \theta} = (\alpha_a - \alpha_b) + \beta_a(1 + \ln(\theta M)) - \beta_b(1 + \ln[(1-\theta)M]) \qquad \text{(VIII.7)}$$

with $\partial w / \partial \theta > 0$ used as evidence that giving more power to the husband will result in household budgets being biased in favour of w and $\partial w / \partial \theta < 0$ to conclude that giving more power to the wife will result in household budgets being biased in favour of w. The previous expression is computed as a non-linear combination of parameters and standard deviations, computed using the Delta method. However, hypothesis-testing might be complicated since it is data-dependant; it is therefore necessary to compute it for different realizations of θ. These results are supplemented by the calculation of predicted budget shares from (VIII.6), where all values are set to the sample means, except for θ.

The econometric issues that emerge during estimation are similar to those already discussed in earlier chapters. Therefore, for the estimation of the non-linear version of the aggregated model (the first exercise), Seemingly Unrelated Regression (SUR) with robust standard errors is used.

Therefore, the base estimator used in the application is the SUR-GLS already used in Chapter 5. For the disaggregated model (the third exercise), in light of the incidence of zero-consumption for different endogenous variables, the S&Y estimator is used.

However, some differences from earlier estimations should be considered. Although the groups used are similar to those in Chapter 6 and all the consumption categories are estimated as a system (SUR) in the second stage of the model, the first stage and the correction covariates are calculated only for the problematic categories (those with censoring above 5%). In contrast to the estimations in Chapter 6, the lowest sample size considered here prevents us from fitting PROBIT estimations when the censoring level is low. In these cases, the implied assumption is that the correction terms $\Phi \approx 1$ and $\phi \approx 0$ (see Wooldridge, 2002). With the final sample size used in the estimation, the problematic cases detected (where censoring needs to be corrected) are, for the model of consumption groups, food outside the house (with censoring around 15%) and health goods and services (with censoring around 18%). In the models of markets within groups, the only unproblematic categories were informal food consumption within the house (with censoring below 1%) and formal education and culture (with censoring around 5%).

In the second exercise, the endogeneity of total expenditure and bargaining power is explicitly handled, so IV techniques are used. In the case of total expenditure, the way that simultaneity bias emerges in this demand framework is the same as in Chapter 5, so similar reasons for instrumenting it in the present context are taken into account. The case for bargaining power is less obvious. In the original sharing rule model presented in Bourguignon *et al*. (1993), Browning *et al*. (1994) and Browning and Chappori (1998), the authors justify how the sharing rule composed by total incomes is exogenous. This line of thinking permits von Braun (1988), García (1990), Bourguignon *et al*. (1993), Browning *et al*. (1994), Hayashi (1995) and Phipps and Burton (1998) to run cooperative bargaining models under the assumption of the exogeneity of individual incomes. However, this argument is criticized by Hoddinott and Haddad (1995) and, more explicitly, by Basu (2006). These authors recognize that, as long the power variable (in most applications) is composed of labour income, similar arguments to those used to justify simultaneity in labour supply applies here. Moreover, even beyond this argument, the notion of power will be simultaneously determined in the household as long as it is plausible to assume some feedback between choices undertaken and the ability to impose one's own decisions.

However, implementing this in the non-linear setting of specification (VIII.6) is not an easy task. For this reason, the strategy followed was to estimate a linear version of the model using the simplified specification proposed by Hoddinott and Haddad (1995). According to the authors a suitable specification for investigating bargaining is just to use the Working–Leser Engel curve extended by the term θ_j and additional household covariates (z) as demand shifters (to control for family preferences). As a result, the specification proposed by the authors takes the form:

$$w_i = \alpha_i + \beta_{1i}\ln(M) + \beta_{2i}\theta_j + \beta_{3i}z + \varepsilon_i \tag{VIII.8}$$

The less theoretical specification of Hoddinott and Haddad (1995) also facilitates the empirical work for three reasons. First, because it imposes from the assumption that individual characteristics only influence demand through the sharing rule. Therefore, Hoddinott and Haddad (1995) and Lancaster *et al*., 2006 in a similar application exploits individual characteristics such as differences in education as potential instruments. In this application a similar strategy is used, but the set of potential instruments is supplemented by those already used in Chapter 5 (value of the assets in the household, external shocks experienced by the family and housing conditions). However, since some of these exclusion restrictions seem to be controversial (basically in terms of the individual characteristics), the assumptions imposed are adequately tested for orthogonality and for relevance in a similar way to Chapter 5. Exogeneity for potentially endogenous regressors is still tested using the C-test.

Second, the author assumes the separability of labour supply and consumption in order to avoid dealing with several endogenous regressors. As these variables are not of particular interest in this chapter, this assumption is also undertaken during estimation. Similar restrictions have been imposed by Bourguignon *et al*. (1993), Browning *et al*. (1994), Browning and Chiappori (1998) and Lancaster *et al*. (2006). As discussed by Bourguignon *et al*. (1993), this not means that labour supply has no effect on demand. On the contrary, it only recognizes that its effect occurs mainly through the sharing rule. However, in order to be consistent with the discussion in previous chapters, and to control for direct influence of the labour supply, additional results are provided here using total household hours of work and its participation dummy as additional controls in the main equation[101] (as in Hayashi, 1995). When included, the exogeneity of labour supply is assumed.

Third, it makes hypothesis-testing and the interpretation of paramemters more transparent. Avoiding the non-linear nature of the theoretical model generates a simple test for bargaining just by verifying whether $\beta_{2i} = 0$. If rejected, then it is possible to conclude that household allocation will be affected not only by total income, but also by control over money by particular members.

[101] Several exercises were done to treat the six endogenous regressors (including hours of work), but similar problems to those already mentioned in Chapter 5 arise in terms of weak and non-orthogonal instruments.

At the same time, the sign of β_{2i} will be informative, since it will reveal which type of allocation (formal or informal) increases or decreases as the bargaining power (or control over money) of member j does. This last result is compared with $\partial w / \partial \theta > 0$ in order to analyse departures from the main conclusions.

VIII.5. The results

First empirical exercise: a non-linear model under the assumption of exogeneity

In Table VIII.5, the SUR–GLS estimations for equation (VIII.6) under the different specifications used in the application are presented. One important thing to note is that the fully usable sample (once the missing values of additional covariates are dropped) is 2,222 observations. The basic statistics of the variables included in the application and their definition using this final sample are presented in Appendix 8. In terms of the regressions presented, measures of global adjustment seem adequate, at least in terms of the discussion presented in Chapter 5. No interpretation of results for the covariates included is undertaken here, as these results are not of particular interest for the application. In contrast, the focus is on the bargaining structure imposed. Full regression results are presented in Appendix 9. However some individual characteristics deserve a comment, considering that they are controlling for possible measures of power not captured by incomes.

For example, age differentials seem to have no influence on market allocation. Neither male nor female age has an influence on informal consumption shares and both have a positive and statistically similar[102] impact on formal consumption. The case for human capital variables is different. Although improving education levels for both wives and husbands increase formal purchases with statistically similar values,[103] only the higher education of females reduces informal consumption. Similarly, in the case of labour supply, households with mothers working more hours increase their formal purchases. In contrast, the results for health variables (chronic diseases) go in a different direction. However, it is possible that these covariates are just revealing the consumption of a particular item (health, biased in favour of formal consumption).

Turning our attention to the parameters of main interest in the application, the effect of bargaining power on consumption ($\partial w / \partial \theta$), computed for different values of θ with total expenditure held constant at the sample mean, is presented at the bottom of the table. The results show that the bargaining structure imposed is jointly significant at conventional levels for all the specifications used. Also, it is important to note that $\partial w / \partial \theta$ is also significant at conventional levels for most

[102] Chi-squared results: 0.02 (1), 0.00 (2), 0.03 (3), 0.00 (4).
[103] Chi-squared results: 0.88 (1), 1.44 (2), 0.95 (3), 1.14 (4).

values of θ. This is considered evidence in favour of the collective model and basically the influence of bargaining power on demand allocation. This effect differs depending on the relative value of power enjoyed by each member.

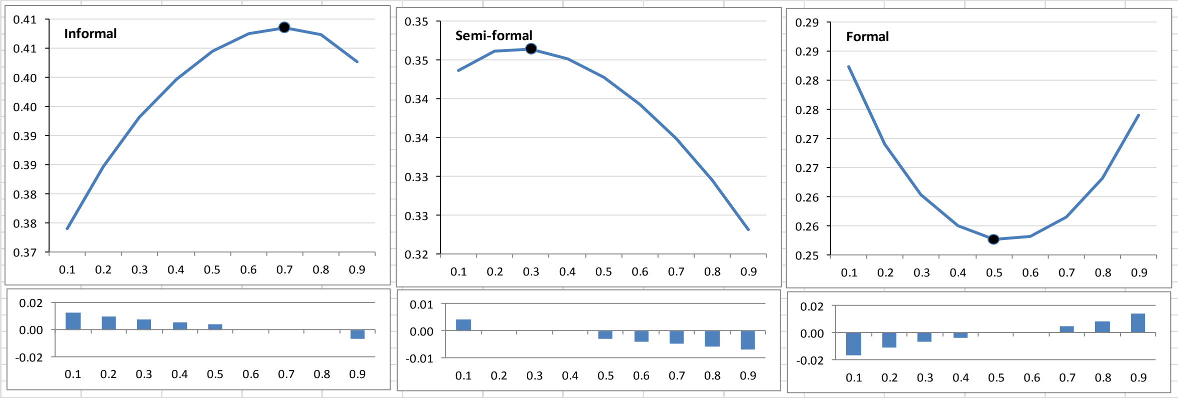
Table VIII.5: SUR–GLS regression results of the bargaining model (three-market definition and different specifications)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
θ	-0.18973	-0.07715	-0.04667	-0.08721	-0.16462	-0.07486	-0.18738	-0.07391
	(0.18485)	*(0.15310)*	*(0.17807)*	*(0.14998)*	*(0.18372)*	*(0.15417)*	*(0.18469)*	*(0.15166)*
$\theta \ln(\theta M)$	-0.03248***	0.07278***	-0.04710***	0.07239***	-0.03290***	0.07235***	-0.03169***	0.07152***
	(0.00906)	*(0.00791)*	*(0.00894)*	*(0.00777)*	*(0.00904)*	*(0.00796)*	*(0.00906)*	*(0.00789)*
$(1-\theta)\ln[(1-\theta)M]$	-0.05555***	0.06595***	-0.05635***	0.06484***	-0.05353***	0.06577***	-0.05444***	0.06507***
	(0.01569)	*(0.01313)*	*(0.01506)*	*(0.01299)*	*(0.01563)*	*(0.01320)*	*(0.01564)*	*(0.01299)*
$\partial w / \partial \theta$								
$\theta = 0.1$	0.12409***	-0.16532***	0.15396***	-0.16686***	0.12447***	-0.16468***	0.12138***	-0.16321***
	(0.02654)	*(0.02482)*	*(0.02613)*	*(0.02468)*	*(0.02656)*	*(0.02490)*	*(0.02634)*	*(0.02464)*
$\theta = 0.2$	0.09503***	-0.10711***	0.11467***	-0.10904***	0.09536***	-0.10678***	0.09300***	-0.10597***
	(0.02376)	*(0.02255)*	*(0.02330)*	*(0.02244)*	*(0.02376)*	*(0.02261)*	*(0.02358)*	*(0.02237)*
$\theta = 0.3$	0.07445***	-0.06879***	0.08805***	-0.07103***	0.07487***	-0.06866***	0.07288***	-0.06828***
	(0.02209)	*(0.02112)*	*(0.02164)*	*(0.02101)*	*(0.02209)*	*(0.02116)*	*(0.02195)*	*(0.02094)*
$\theta = 0.4$	0.05654***	-0.03769*	0.06581***	-0.04021**	0.05716***	-0.03771*	0.05537***	-0.03767*
	(0.02070)	*(0.01983)*	*(0.02025)*	*(0.01972)*	*(0.02070)*	*(0.01986)*	*(0.02058)*	*(0.01967)*
$\theta = 0.5$	0.03917**	-0.00942	0.04503**	-0.01224	0.04006**	-0.00957	0.03838**	-0.00985
	(0.01939)	*(0.01851)*	*(0.01897)*	*(0.01840)*	*(0.01939)*	*(0.01853)*	*(0.01930)*	*(0.01837)*
$\theta = 0.6$	0.02085	0.01856	0.0238673	0.01543	0.02212	0.01830	0.02045	0.01771
	(0.01821)	*(0.01713)*	*(0.01779)*	*(0.01703)*	*(0.01821)*	*(0.01714)*	*(0.01815)*	*(0.01702)*
$\theta = 0.7$	-0.00014	0.04876***	0.00040	0.04524***	0.00165	0.04837***	-0.00010	0.04745***
	(0.01744)	*(0.01581)*	*(0.01700)*	*(0.01571)*	*(0.01744)*	*(0.01582)*	*(0.01741)*	*(0.01574)*
$\theta = 0.8$	-0.02700	0.08522***	-0.0287391	0.08120***	-0.02445	0.08470***	-0.02640	0.08339***
	(0.01803)	*(0.01513)*	*(0.01748)*	*(0.01504)*	*(0.01803)*	*(0.01515)*	*(0.01802)*	*(0.01511)*
$\theta = 0.9$	-0.06933***	0.13950***	-0.07334***	0.13467***	-0.06543***	0.13881***	-0.06787***	0.13691***
	(0.02309)	*(0.01765)*	*(0.02218)*	*(0.01752)*	*(0.02305)*	*(0.01772)*	*(0.02308)*	*(0.01763)*
Observations	2,222		2,222		2,222		2,222	
LogL	2348.952		2429.140		2360.635		2359.235	
Chi-sq (overall)	1346.179***		1554.114***		1373.301***		1405.538***	
Chi-sq (bargain structure)	31.62***	135.93***	51.29***	133.95***	31.08***	133.19***	30.49***	131.50***

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

To gain a more adequate interpretation of the direction of the results, it is better to construct the predicted budget shares for a different realization of θ. In Figure VIII.3, these results are displayed for specification (1) for both informal and formal consumption shares and the semi-formal category, computed using the adding-up property. The horizontal axis corresponds to the male's power and the vertical to the consumption shares. At the bottom of each graph, the computed value of the marginal effect is presented, again for each realization of θ; however, to facilitate interpretation, only those significant (at least at 10%) are displayed.

Figure VIII.3: Predicted budget shares and male bargaining power (three-market definition and base specification)



Note: Bigger graphs refer to predicted budget shares (vertical axis: consumption share; horizontal axis: realization of θ); the bottom graphs refer to computed marginal effects (vertical axis: marginal effect; horizontal axis: realization of θ). Marginal effects for each realization of θ are multiplied by 0.1 to render them comparable to predicted shares.

It is possible to verify a non-linear, but positive, relation between male bargaining power and informal consumption: households with more-empowered women will have lower informal consumption shares than the reverse. Therefore, moving from the left to the right or increasing male bargaining power will increase participation of informal consumption in family baskets, but up to a maximum level. From the point estimates shown, this level is reached between θ=0.6 and θ=0.8. Thus, as the husband increases his control over money, he will move the resource allocation in favour of informal goods until he reaches a dominant situation in the house. From there, he will commit household consumption to a maximum level of informality and any additional effect of the distribution of power seems to vanish on the allocation of this particular market.

However, the graphs can produce much richer information. Following Basu (2006) and Lancaster *et al*. (2006), as the effect of power on household consumption depends on the relative power enjoyed by each family member, it is possible to define three situations. First, there will be a perfectly balanced household (at θ=0.5) when either of the two members has enough power to impose his or her preferences. The resultant allocation of resources defined at this level will be the *status quo*. Second, to the right of θ=0.5 is consistent with households with powerful husbands. Third, to the left of θ=0.5 is consistent with households with powerful wives. As explained by Basu (2006) and Lancaster *et al*. (2006, the effects of the bargaining power of each member will be different whether or not they are in a dominant situation.

Departing from the perfectly balanced situation (or *status quo*) and moving to the left (increasing the female's power when the female dominates) will be clearly associated with less-informal consumption baskets. The marginal effects estimates are also positive and significant at conventional levels, with point estimates higher,[104] as the mother increases her power within the house. For example, at around θ=0.5, holding everything else constant, an increase of 0.1 percentage points (pp.) in female power is consistent with informal shares between 0.004 and 0.005 pp. lower (depending on the specification) and around θ=0.1; an increase of 0.1 pp. in female power is consistent with informal shares between 0.012 and 0.015 pp. lower (three times larger). Alternatively, when the father dominates (to the right of the perfectly balanced situation), as explained, the relationship vanishes. As a result, it is true that increasing the female's bargaining power will reduce informal consumption shares. However, this is not true for all the distributional range. The effects of bargaining on informal consumption are only clear in households where the female enjoys enough power.

The estimated results for formal consumption are different. A non-linear relationship between male bargaining power and formal consumption share is again verified but, instead of the expected negative relationship, an almost perfect U is displayed. Therefore, from this perfect balanced situation (θ=0.5), increasing both the mother´s and the father´s power is consistent with higher formal shares. From the point value estimates, we can see that the average slopes on the left (when the bargaining power of the mother is higher) are slightly bigger than those on the right (when the bargaining power of the father is higher) but, when formally tested, they are not significantly different (in absolute terms).[105] As a result, it is better to conclude that both parents will increase their allocation to formal markets as their bargaining power increases and they are in a dominant situation.

To interpret this relationship it is necessary to remember the nature of formal goods: better quality, but more expensive. It is then reasonable to expect that each member will prefer (individually) to purchase high-quality goods, but if this is likely to cause financial difficulties for the family, he or she also needs to enjoy enough power to mould family preferences to his or her preferred allocation. In consequence, when the mother has a dominant position in the house, she will use her power to increase formal allocation (the slope of θ will be negative). Similarly, when the father has a dominant position, the slope of θ will be positive, revealing that he will also use his power to increase formal consumption. However, when the power is balanced, they will be committed to a minimum allocation for these goods. Interestingly, in both cases the computed

---

[104] When significant, chi-squared results show differences in computed marginal effects (base specification): θ=0.1 vs θ=0.2 (21.91***), θ=0.1 vs θ=0.3 (24.63***), θ=0.1 vs θ=0.4 (27.06***), θ=0.1 vs θ=0.5 (29.09***), θ=0.2 vs θ=0.3 (28.19***), θ=0.2 vs θ=0.4 (30.04***), θ=0.2 vs θ=0.5 (31.00***), θ=0.3 vs θ=0.4 (31.08***), θ=0.3 vs θ=0.5 (30.91***), θ=0.4 vs θ=0.5 (29.78***).

[105] Chi-squared for testing equality in marginal effects (in absolute terms) on both sides of the graph (base specification): 0.74, 0.44, 0.32, 0.27, 0.26.

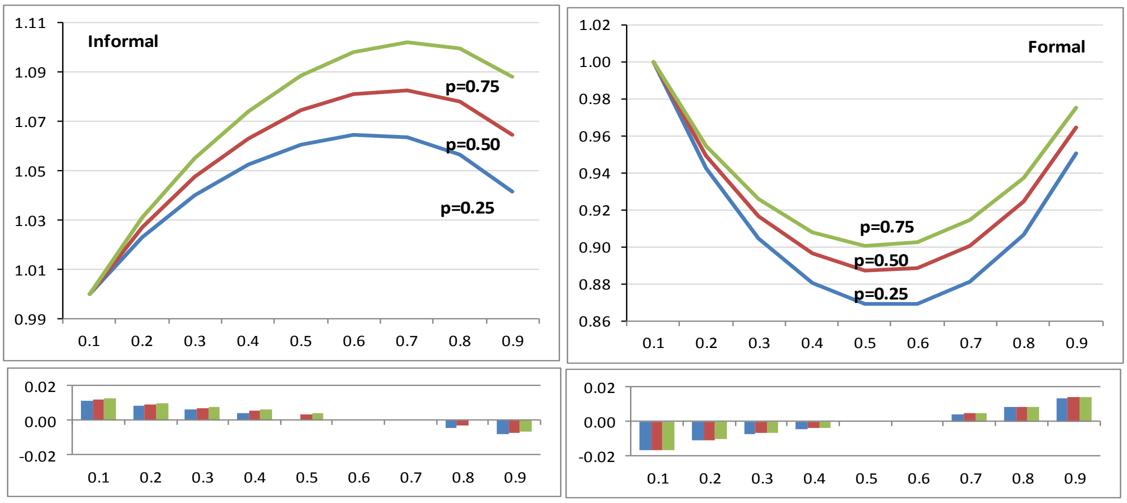
marginal effects are greater as the power of each member increases. A similar relationship is found by Lancaster *et al*. (2006) for luxuries.

To reconcile this and earlier evidence and provide a general interpretation for resource allocation, the estimates for the semi-formal category need to be considered. The perfect balanced situation is consistent with a minimum allocation on formal goods and (around) the maximum allocation on informal and semi-formal ones. Empowering either of the two members outperforms and increases formal consumption in the house. However, the composition of the consumption bundle in terms of informal markets will be different depending on the member empowered. In the case of husbands, higher formal consumption comes out with lower semi-formal purchases holding constant (at their maximum) informal shares. In the case of wives, higher formal consumption comes out with lower informal purchases holding constant (at their maximum) semi-formal shares. Therefore, both members in their dominant position will use their power to increase formal shares, but show a differentiated attitude towards informal consumption: only females will reduce informal shares, revealing a much more aggressive position towards low-quality goods when compared with husbands in a similar position.

The results of this exercise suggest that it is possible to think of some family members (possibly husbands) as being much more worried about the financial stability of the house. Thus, although they would prefer to increase quality (via formal consumption), their attitude will be more conservative in terms of informal goods in order to maintain a balanced budget. On the other hand, it is also possible that other members (possibly wives) may be relatively less worried about the financial stability of the house, but much more worried about quality. Thus, their higher formal consumption comes with a more aggressive position towards informality. This result is used as evidence to conclude in favour of the hypothesis mentioned at the beginning of the chapter: that increasing female power is correlated to less-informal consumption baskets. Unfortunately, this is true only for households where women are in a dominant situation, which is rare in developing countries. There, income within the household is normally biased in favour of the men. For example, in the Peruvian case, as discussed earlier, the mean value of θ is around 0.77. Thus, in order to fully exploit the effectiveness of distributional policies in terms of formalization, these need to be circumscribed to households where the female is empowered enough or sufficiently aggressive to pass the 0.5 barrier. If not, small changes in power distribution in favour of the female in households where the husband dominates will not produce the desirable outcome.

From expression (VIII.7) we can see that $\partial w/\partial\theta$ depends not only on the value of θ, but also on that of M (total expenditure). In Figure VIII.4, the previous results are replicated for different points of the distribution of M in order to assess the validity of the results in terms of the general living conditions of the households under analysis. As the absolute level of informal and formal consumption shares will vary depending on the absolute level of expenditure enjoyed by the family (richer families purchase more formal goods and less informal ones), consumption shares are normalized at 1 at the point θ =0.1 in order to make the analysis more transparent. As shown, once the predicted shares and the marginal effects of θ are computed for the different expenditure levels at percentiles 25, 50 and 75, the story behind the bargaining does not change dramatically. However, it does reveal that the influence of total expenditure on the results is crucial. In fact, as the family becomes richer, the effects of θ on informal consumption become stronger and the U relationship for formal consumption becomes steeper. These results might be indicating that women in richer families will be more aggressive towards informality, though financial concerns, as discussed earlier, may well then be less important.

Figure VIII.4: Predicted budget shares and male bargaining power for different income values (three-market definition and base specification)



Note: See footnote in Figure VIII.3 for details.

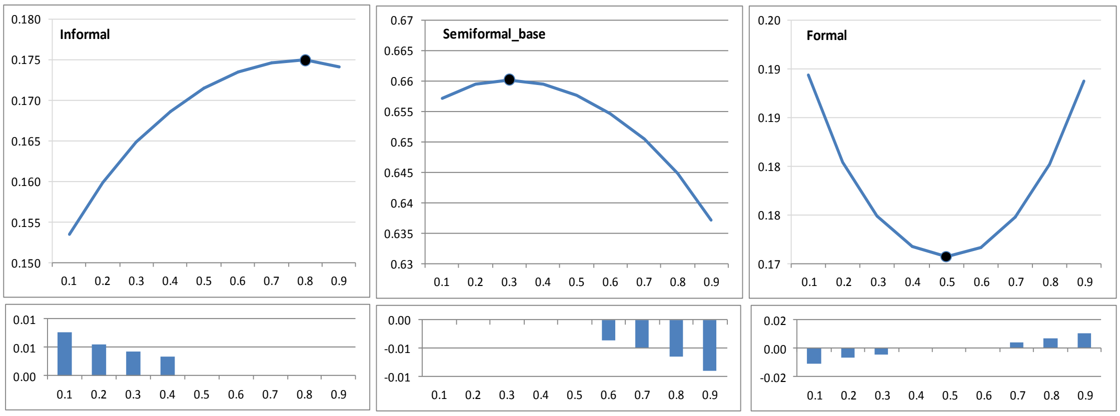
The conclusions are robust to changes in the definitions of informal and formal markets. The regression results for the five-market definition are presented in Table VIII.6 and the estimated budget shares are displayed in Figure VIII.5, again for different realizations of θ for the formal, semi-formal and informal categories, but only for specification (1). We can again observe a positive, but non-linear, relation between male power and informal consumption and a quasi-perfect U relationship in the case of formal shares. However, it is worth noting that point estimates for the computed marginal effects of θ under the five-market definition are lower (in absolute terms) than in the previous case. This defines a plainer relationship. As a result, the bargaining space (the range where $\partial w / \partial \theta$ is significant) is slightly smaller and the relationship for formal consumption is much more centred. Therefore, the results are robust to the change in the definition of informality used here, but when the definition is stricter, the average effect of the bargaining power on the consumption allocation seems to be lower.

Table VIII.6: SUR–GLS regression results of the bargaining model (five-market definition and different specifications)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.13513	-0.04445	0.14510	-0.04415	0.13882	-0.04584	0.13908	-0.04199
	(0.14335)	*(0.12500)*	*(0.14196)*	*(0.12352)*	*(0.13834)*	*(0.12589)*	*(0.14265)*	*(0.12414)*
$\theta \ln(\theta M)$	-0.02651***	0.05196***	-0.02738***	0.05013***	-0.02357***	0.05168***	-0.02836***	0.05069***
	(0.00663)	*(0.00630)*	*(0.00656)*	*(0.00629)*	*(0.00643)*	*(0.00639)*	*(0.00659)*	*(0.00630)*
$(1-\theta)\ln[(1-\theta)M]$	-0.01534	0.04750***	-0.01519	0.04591***	-0.01229	0.04706***	-0.01677	0.04652***
	(0.01265)	*(0.01114)*	*(0.01236)*	*(0.01097)*	*(0.01219)*	*(0.01120)*	*(0.01263)*	*(0.01106)*
Observations	2,222		2,222		2,222		2,222	
LogL	3270.640		3364.251		3311.910		3308.613	
Chi-sq (overall)	1090.769***		1295.196***		1165.259***		1192.609***	
Chi-sq (bargain structure)	21.17***	104.71***	23.51***	97.01***	19.15***	101.12***	24.28***	99.43***

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The specifications used are those of Table VIII.5.

Figure VIII.5: Predicted budget shares and male bargaining power (five-market definition and base specification)



Note: See footnote in Figure VIII.3 for details.

<u>Second empirical exercise: taking endogeneity into account</u>

Although previous estimates seem intuitive, they are calculated under the (possibly) unsatisfactory assumption of the exogeneity of total expenditure and bargaining power. This second exercise drops this assumption and calculates the results using an instrumental variable approximation. As explained above, to do this, the specification has been changed in order to avoid non-linearities in the endogenous variables that complicate the application of an IV. For this reason the linear approximation proposed in Hoddinott and Haddad (1995) is used here. The full set of potential instruments is used in the application – the ratio of educational attainment between the male and the female, the ratio of age between the male and the female, the linear and quadratic of education and age for each member, the dummy variables that identify chronic illness for the male and the female, the logarithm of the value of assets of the household, the characteristics of the house (a dummy variable for inadequate housing and a dummy variable that identifies overcrowded houses) and variables for external shocks experienced by the family in the last year (job loss, bankruptcy of the family business, the serious illness, accident or death of a member of the household, the abandonment of the family by the household head, and the family's experience of a fire in the house, a robbery or other crime, or a natural disaster).

As mentioned previously, the application is not free of problems. First, because the list of potential instruments is long and could cause finite sample problems, as suggested in Wooldridge (2002). Second, exclusion restrictions for some of the instruments used are controversial. The case for the value of assets, characteristics of the house and shocks experienced by the family is easier, as these are assumed proxies (or determinants) of permanent income or investment decisions of the household which, in line with the two-stage budgeting approach, could be conveniently assumed to be uncorrelated with short-term decisions on consumption (see Chapter 5). The case of individual characteristics is harder. We could argue that individual characteristics will affect the distribution of power within the household, as better-educated, older or healthy members will generally be associated with having a higher say in household decision-making. However, they could be also correlated with preferences.

The correct procedure for adequately handling the problems discussed here was to find better instruments, something that, as stated, was not possible with the datasets explored in this application. For this reason, the approximate solution proposed here is based on very rigorous statistical procedures for instrument selection based on their validity and relevance. Thus, at least in these terms, the econometric exercise used here can be considered adequate. The final set of instruments introduced comprised the logarithm of the value of assets, the ratio of male to female education, the educational attainment of the female and its square, one shock (if the family has experienced robbery or another crime against it) and one house characteristic[106] (the dummy variable that identifies overcrowded houses).

Before the IV results are presented, it is better to start revising the OLS approximation of the estimates. Although this is not the preferred estimation, it will help to more adequately assess the nature and the direction of the bias that IV procedures try to handle. In Table VIII.7 these estimates are displayed for the two variables of interest: θ and the logarithm of total expenditure. As mentioned previously, two versions of the model are presented – the first includes variables for the labour supply of the family as additional regressors and the second assumes separability

[106] The dummy variable of house characteristics is only used for the exercise under the second assumption of labour supply: separability.

between consumption and labour supply. In each case, the four specifications already explained are used for comparative purposes. Note that the sample size has been reduced to take into account the additional missing values introduced by the instrumental variables proposed in the application. Beyond that, the models show adequate goodness-of fit-measures (at least comparable with previous models, where prevailing noise in explaining informal consumption was already discussed).

Table VIII.7: OLS estimates for the linear approximation of the bargaining model (three-market definition and different specifications)

	(1)		(2)		(3)		(4)	
With labour supply								
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.00113	-0.00432	0.00618	-0.00462	0.00147	-0.00399	0.00141	-0.00420
	(0.01376)	(0.01072)	(0.01334)	(0.01066)	(0.01381)	(0.01071)	(0.01375)	(0.01068)
L(expenditure)	-0.05396***	0.10977***	-0.06624***	0.10511***	-0.04996***	0.10907***	-0.05356***	0.10860***
	(0.00632)	(0.00567)	(0.00626)	(0.00572)	(0.00643)	(0.00576)	(0.00633)	(0.00566)
Observations	2,160	2,160	2,160	2,160	2,160	2,160	2,160	2,160
R-squared	0.091	0.267	0.143	0.285	0.101	0.268	0.094	0.272
LogL	758.621	1275.637	822.280	1301.578	770.401	1276.560	762.252	1282.255
F (overall)	13.22***	36.61***	15.84***	31.98***	13.08***	31.55***	11.37***	31.63***
Without labour supply								
	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
θ	-0.00061	-0.00925	0.00482	-0.01018	0.00040	-0.00907	-0.00010	-0.00936
	(0.01342)	(0.01051)	(0.01300)	(0.01047)	(0.01346)	(0.01050)	(0.01340)	(0.01047)
L(expenditure)	-0.05365***	0.11065***	-0.06604***	0.10618***	-0.04978***	0.11004***	-0.05326***	0.10958***
	(0.00630)	(0.00558)	(0.00621)	(0.00562)	(0.00641)	(0.00567)	(0.00632)	(0.00558)
Observations	2,160	2,160	2,160	2,160	2,160	2,160	2,160	2,160
R-squared	0.091	0.265	0.143	0.282	0.101	0.266	0.094	0.269
LogL	758.034	1272.613	821.929	1298.135	769.942	1273.392	761.685	1279.041
F (overall)	14.72***	38.98***	17.27***	32.57***	14.35***	32.91***	12.40***	32.90***

Note: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. The specifications presented also include the following covariates: (1): L(fam_size), D(mem_son), D(mem_oth), P(men), D(gender_head), D(Lima), household composition in proportions; (2): (1) + L(pop_urb_dist), L(pop_urb_dist)-sq, pop_den_distx1000, D(not_slum), D(mid_city), D(border); (3): (1) + D(tongue), D(migrant), D(social); (4): (1) + Station_distx1000, Muni_persx1000, L(budget_dist_pc), D(sdp_dist). Labour supply (participation dummy and hours of work for the household) is included in all specifications when indicated.

The results for total expenditure seem to be well determined, statistically significant and with correct signs, though they are different for θ. The results show a possibly downward bias for this variable. Although in six of the eight regression models presented here the correct sign is shown (positive effect on informal consumption shares and negative on formal ones), it seems not to be well determined as it is not significant at conventional levels. Therefore, using standard statistical procedures it is difficult to assess the real direction of its effect. This is the departure point for the IV procedure proposed.

The diagnostic tests for the IV models are presented in Table VIII.8 for both of the assumptions made about labour supply and the different specifications used in the application. As shown, relevance (for a maximum relative bias of 10%) and orthogonality conditions are met for both assumptions about labour supply (with better results when hours of work are included). Also in both cases C-statistics found evidence of endogeneity in the instrumented variables, revealing the need to develop the IV procedure proposed here. The first-stage estimations are presented in Appendix 9. The additional instruments display the expected signs. For example, the effect of asset values has a strong and positive effect on total expenditure, as well as on the educational gap between males and females (possibly related to the return-to-education gap). Similarly, when the parameters associated with female education are inspected, they show the expected positive (and convex) effect on total expenditure. The bargaining power variable shows a negative correlation with total assets, revealing that the power in wealthier families is more evenly distributed. The education gap between males and females has a positive effect on male income shares. Interestingly, when the parameters of female education are inspected, they reveal a positive, but concave, relationship with power. This is possibly revealing of the access difficulties of, or wage discrimination against, women with low educational attainment (below secondary education), something that is possibly solved once higher education is reached. Finally, the variable attached to external shocks predicts a rebalance of power in favour of husbands.

Table VIII.8: Relevance and orthogonality tests of the IV application (three-market definition and different specifications)

	(1)		(2)		(3)		(4)	
With labour supply								
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
Under-identification Test	70.394***		70.53***		67.980***		70.280***	
F-value (1stage, theta)	25.61***		23.93***		26.46***		25.45***	
F-value (1stage, expend)	343.20***		309.79***		332.12***		341.91***	
F-value (1 stage, joint)	15.58		15.595		14.967		15.495	
Hansen J test	5.252	1.925	5.847	0.963	2.524	2.121	5.355	2.188
C (labour)	3.581	1.728	3.844	0.882	1.245	1.956	3.655	1.952
C (theta,expend)	20.467***	87.735***	14.067***	73.256***	16.793***	88.573***	21.084***	85.255***
Without labour supply								
	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
Under-identification Test	58.748***		58.737***		56.709***		58.869***	
F-value (1stage, theta)	20.51***		19.18***		21.64***		20.58***	
F-value (1stage, expend)	288.18 ***		260.45***		278.40***		287.30***	
F-value (1 stage, joint)	10.769		10.761		10.362		10.76	
Hansen J test	6.400	3.786	7.529	2.622	3.360	3.994	6.690	4.227
C (theta,expend)	19.951***	84.547***	13.240***	71.032***	16.300***	85.573***	20.324***	81.955***

Note: *** p<0.01, ** p<0.05, * p<0.1. Tests are performed under a 2SLS specification with small sample correction and robust standard errors. Individual F-statistics must be compared with the rule of thumb of 10 (Cameron and Trivedi, 2005), Joint F-statistics are compared with the critical value of 13.97 and 15.72 (5% of maximal IV relative bias) computed in Stock and Yogo (2005) and 8.78 and 9.48 (10% of maximal IV relative bias) as the minimum tolerable levels with and without labour supply, respectively. See footnote in Table V.7 for other details in the tests.

The IV regression results are presented in Table VIII.9 (including labour supply) and in Table VIII.10 (under the assumption of separability between consumption and labour). The estimations have been done in each case allowing for small sample adjustments and standard errors that are robust to the presence of arbitrary heteroskedasticity. The standard 2SLS estimates are also compared with other estimators such as the Limited Information Maximum Likelihood (LIML) and Generalized Method of Moments (GMM) estimators.[107] Although instruments are not weak under standard tests, as some relative bias is still present in the application,[108] it is useful to compare 2SLS (biased in the direction of OLS) with other methods (LIML and GMM) that are intended to correct this problem and are more robust to the presence of weak instruments (see Cameron and Trivedi, 2005).

Table VIII.9: IV regression results of the bargaining model including labour supply as additional regressor (three-market definition, different specifications and estimation methods)

(1)	2SLS		LIML		GMM	
	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.22711***	-0.08066	0.24249***	-0.08205	0.23632***	-0.07939
	(0.06979)	(0.05300)	(0.07523)	(0.05419)	(0.06928)	(0.05293)
L(expenditure)	-0.04847***	0.15254***	-0.04713***	0.15250***	-0.04810***	0.15334***
	(0.01184)	(0.00861)	(0.01225)	(0.00868)	(0.01181)	(0.00859)
Observations	2,160	2,160	2,160	2,160	2,160	2,160
LogL	631.231	1207.603	614.453	1206.746	621.252	1206.879
F-value (overall)	11.73***	39.77***	11.51***	39.71***	11.75***	40.32***
(2)	2SLS		LIML		GMM	
	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.18594***	-0.05413	0.19895***	-0.05451	0.19682***	-0.05441
	(0.06579)	(0.05193)	(0.07104)	(0.05251)	(0.06542)	(0.05186)
L(expenditure)	-0.06364***	0.15344***	-0.06250***	0.15346***	-0.06265***	0.15393***
	(0.01166)	(0.00886)	(0.01203)	(0.00889)	(0.01164)	(0.00884)
Observations	2,160	2,160	2,160	2,160	2,160	2,160
LogL	735.605	1242.857	723.181	1242.653	725.228	1241.792
F-value (overall)	13.00***	34.61***	12.83***	34.59***	13.10***	34.96***
(3)	2SLS		LIML		GMM	
	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.22732***	-0.08037	0.23442***	-0.08189	0.23618***	-0.07959
	(0.07109)	(0.05410)	(0.07362)	(0.05546)	(0.07052)	(0.05404)
L(expenditure)	-0.03895***	0.15310***	-0.03827***	0.15306***	-0.03819***	0.15404***
	(0.01238)	(0.00904)	(0.01259)	(0.00913)	(0.01236)	(0.00901)
Observations	2,160	2,160	2,160	2,160	2,160	2,160
LogL	642.179	1207.132	634.453	1206.201	632.501	1205.832
F-value (overall)	10.87***	34.29***	10.78***	34.24***	10.87***	34.86***
(4)	2SLS		LIML		GMM	
	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.23379***	-0.07844	0.24971***	-0.07999	0.24415***	-0.07797
	(0.06983)	(0.05258)	(0.07536)	(0.05394)	(0.06935)	(0.05251)
L(expenditure)	-0.04681***	0.15153***	-0.04541***	0.15149***	-0.04641***	0.15225***
	(0.01194)	(0.00858)	(0.01236)	(0.00866)	(0.01192)	(0.00855)
Observations	2,160	2,160	2,160	2,160	2,160	2,160
LogL	627.711	1215.471	609.902	1214.526	616.202	1214.369
F-value (overall)	10.24***	34.65***	10.05***	34.59***	10.25***	35.09***

Note: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. Note that the table is divided into four parts according to each of the specifications used in the present application.

[107] Diagnostic tests were re-calculated under each of these new estimators without changes to those already presented in Table VIII.8.
[108] Joint F-tests are below or just above the 5% relative bias critical value.

Table VIII.10: IV regression results of the bargaining model under the assumption of separability between labour supply and consumption decisions (three-market definition, different specifications and estimation methods)

(1)	2SLS		LIML		GMM	
	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.23606***	-0.06858	0.26042***	-0.07107	0.25454***	-0.06999
	(0.07738)	(0.05767)	(0.08667)	(0.06081)	(0.07656)	(0.05755)
L(expenditure)	-0.04523***	0.15363***	-0.04291***	0.15355***	-0.04392***	0.15470***
	(0.01267)	(0.00897)	(0.01345)	(0.00918)	(0.01262)	(0.00894)
Observations	2,160	2,160	2,160	2,160	2,160	2,160
LogL	613.243	1212.338	584.279	1211.043	591.452	1209.391
F-value (overall)	12.79***	44.17***	12.40***	44.06***	12.96***	45.06***
(2)	2SLS		LIML		GMM	
	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.18782***	-0.04190	0.20890**	-0.04248	0.20587***	-0.04363
	(0.07235)	(0.05661)	(0.08163)	(0.05860)	(0.07178)	(0.05648)
L(expenditure)	-0.06123***	0.15486***	-0.05919***	0.15493***	-0.05934***	0.15572***
	(0.01233)	(0.00921)	(0.01308)	(0.00935)	(0.01230)	(0.00919)
Observations	2,160	2,160	2,160	2,160	2,160	2,160
LogL	728.290	1244.112	706.645	1243.777	709.843	1241.785
F-value (overall)	14.03***	36.88***	13.72***	36.84***	14.22***	37.60***
(3)	2SLS		LIML		GMM	
	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.23761***	-0.06932	0.24991***	-0.07200	0.25472***	-0.07094
	(0.07875)	(0.05878)	(0.08345)	(0.06223)	(0.07789)	(0.05866)
L(expenditure)	-0.03541***	0.15416***	-0.03413**	0.15407***	-0.03362**	0.15546***
	(0.01334)	(0.00948)	(0.01376)	(0.00974)	(0.01328)	(0.00944)
Observations	2,160	2,160	2,160	2,160	2,160	2,160
LogL	623.415	1211.342	608.869	1209.948	603.014	1207.808
F-value (overall)	11.69***	37.42***	11.50***	37.32***	11.76***	38.30***
(4)	2SLS		LIML		GMM	
	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.24084***	-0.06518	0.26640***	-0.06775	0.26086***	-0.06806
	(0.07719)	(0.05718)	(0.08680)	(0.06064)	(0.07641)	(0.05702)
L(expenditure)	-0.04355***	0.15277***	-0.04110***	0.15271***	-0.04215***	0.15364***
	(0.01278)	(0.00894)	(0.01359)	(0.00918)	(0.01273)	(0.00891)
Observations	2,160	2,160	2,160	2,160	2,160	2,160
LogL	611.654	1220.488	580.740	1219.187	587.616	1217.148
F-value (overall)	10.92***	37.56***	10.58***	37.47***	11.05***	38.23***

Note: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. Note that the table is divided into four parts according to each of the specifications used in the present application.

In general, the regression results show adequate global adjustment and, individually, the core covariates are well determined (intuitive and significant at conventional levels). To save space, only the parameters of interest to the present application are displayed under the different specifications used throughout the chapter. In the case of total expenditure, the estimates are in line with previous results on this variable: a positive effect on formal consumption shares and negative effects on their informal counterparts. Also, the IV coefficients (under both assumptions of labour supply) are highly comparable with the OLS procedure presented above for the informal equation. In both cases, increasing the total expenditure of the family by 10% will generate a reduction of 0.5 of one percentage (using the mean of all specifications). For the formal equation, IV estimates are higher than OLS ones. In the first case, a similar shock will increase formal consumption by 1.5 pp. and, in the second, the estimated growth would be 1.1 pp.

Table VIII.11: Total expenditure elasticities for different regression models

	Unitary Model				Collective Model			
	OLS (3-market) 1/	SUR (3-market) 2/	SUR (5-market) 3/	IV- 2SLS (3-market) 4/	Linear OLS (3-market) 5/	NL-SUR (3-market) 6/	NL-SUR (5-market) 7/	Linear IV 2SLS (3-market) 8/
Informal	0.939***	0.932***	0.893***	0.951***	0.866***	0.908***	0.868***	0.880***
	(0.01405)	*(0.01051)*	*(0.01740)*	*(0.00788)*	(0.01570)	(0.01780)	(0.03092)	*(0.02940)*
Formal	1.282***	1.290***	1.323***	1.437***	1.418***	1.278***	1.292***	1.580***
	(0.01796)	*(0.01319)*	*(0.01804)*	*(0.02934)*	(0.02156)	(0.02415)	(0.02914)	*(0.03277)*

Note: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. Hypothesis-testing is presented for the null 'different to 1'. 1/ see Table V.9 specification (1); 2/ see Table V.1 specification (1); 3/ see Table V.5 specification (1); 4/ see Table V.13 specification (1); 5/ see Table VIII.7 specification (1) with labour supply; 6/ see Table VIII.5 specification (1); 7/ see Table VIII.6 specification (1); 8/ see Table VIII.9 specification (1).

Interestingly, and beyond point value differences attributable to marginal effects, with this set of results it is possible to argue that the different estimates for expenditure throughout the thesis do not necessarily produce a different story in terms of elasticities. In Table VIII.11, the different sets of expenditure elasticities are shown. Here the expenditure elasticities for the three-market definition already presented in Chapter 5 are replicated and compared with the implied values computed under the linear and non-linear approximations of the bargaining model. In each case, specification (1) is used for comparative purposes and, in the case of IV models, the most general 2SLS approximation is displayed. The other results, as discussed each time, are highly comparable and omitted for the sake of brevity. It is worth noting that the computation of total expenditure elasticities differs between linear and non-linear approximations. For the first, as already explained, this value is computed using: $\eta = \beta_{1i} / w_i + 1$ whereas, for the second, using expression (VIII.6), the correct formula becomes:

$$\frac{\partial C}{\partial M}\frac{M}{C} = \eta = \frac{\beta_a \theta + \beta_b (1-\theta)}{w} + 1 \qquad \text{(VIII.9)}$$

The results define a very clear situation already discussed in previous chapters: inferiority for informal consumption is rejected. In fact, both informal and formal purchases increase with expenditure and can be classified as normal goods, necessities in the case of informal and luxuries in the case of formal consumption. As a result, the impact of income will be greater for the former, which is in line with the formalization of consumption baskets as income growth or protection strategies developed by families (increasing informal consumption) as income decreases. Using the theoretical models in this chapter, we can argue that this conclusion is valid for both unitary and collective models, under the different regression methods used and definitions of informality.

For bargaining power effects, the estimates for θ are generally in line with the main conclusions of the present application: positive and significant for the informal equation and negative but non-significant for its formal counterpart. Therefore, we can conclude that family members actually negotiate when they decide on which markets they will purchase their goods and the relative bargaining power enjoyed by each member is a critical determinant of consumption allocation. However, considering the IV estimation presented here, this observation is better applied to the decision to purchase at informal markets than at formal ones. In the first case, increasing the bargaining power of the husband (or decreasing the bargaining power of the wife) will produce higher informal shares in the household. Thus, we can argue that a positive sloped line is a correct linear approximation of the mean for the relationship between θ and informal budget shares. This reveals no more than the differences in preferences between family members discussed earlier, with wives in a much stronger position against purchasing from informal markets when compared with husbands. For formal markets, the IV results reveal that bargaining power is less relevant in household decisions and it is total income that matters. Therefore, adjustment of family consumtion is done using semi-formal consumption.

When labour supply is included, the point estimates reveal that increasing the income share of the mother in the house by 0.10 pp. will produce a reduction in informal shares of around 0.023 pp. (using the mean of the estimations). These results are greater than those presented in the OLS version of the model and the average marginal effects computed using the non-linear SUR–GLS. On the assumption of separability between labour and consumption, the point estimates are slightly higher. A similar shock will produce a reduction in informal shares of around 0.024 pp. (using the mean of all the estimations). However, it is worth noting that estimates are generally lower under specification (2). In this case, the marginal effects reveal that, for a 0.10 pp. reduction in the income share, informal shares increase by around 0.019 pp. (on the first assumption of labour supply) and 0.020 pp. (on the second assumption of labour supply).

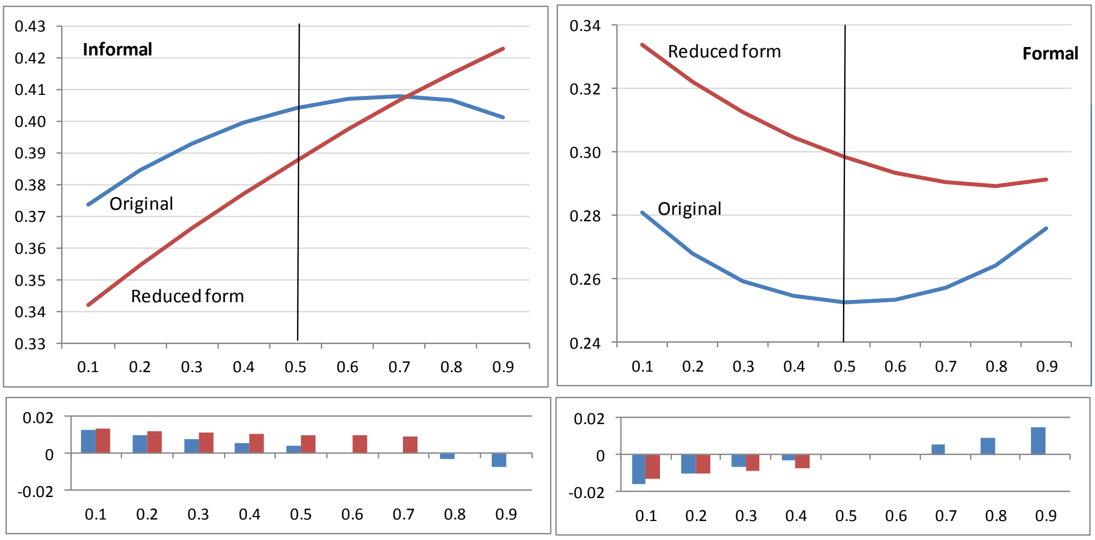
Although the results for θ seem, up to now, to be going in the same direction in the two exercises proposed in this chapter (at least for informal consumption), it is still difficult to compare the two estimates. To partially overcome this problem, the exercise proposed in Hoddinott and Haddad (1995) is also used here. As explained by the authors in one of their consistency checks, it is possible to estimate reduced forms where the potential endogenous variables are replaced directly by their instruments. This avoid the imposition (and testing) of exclusion restrictions in the non-linear model, but at the cost of theoretical difficulties in terms of the AIDS specification used here (mainly in terms of the property of aggregation).

Table VIII.12: SUR–GLS estimates using reduced form specification, three-market definition and specification (1)

	Original specification 1/		Reduced form specification 2/	
	Informal	Formal	Informal	Formal
θ	-0.25790	-0.05320	0.11724	0.09298
	(0.19113)	*(0.16360)*	*(0.19766)*	*(0.14304)*
$\theta \ln(\theta M)$	-0.03052***	0.07280***	-0.01505	0.03006***
	(0.00928)	*(0.00821)*	*(0.01276)*	*(0.00958)*
$(1-\theta)\ln[(1-\theta)M]$	-0.06037***	0.06799***	-0.01318	0.04715***
	(0.01596)	*(0.01389)*	*(0.01388)*	*(0.01048)*
Observations	2,160		2,160	
LogL	2295.909		2173.723	
Chi-sq (overall)	1280.475***		939.345***	
Chi-sq (bargain structure)	32.18***	132.88***	37.29***	324.30***

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. 1/ θ is defined as the income share and total expenditure is used as M ; 2/ θ is defined as the educational share, total value of assets is used as M and individual education levels are excluded from the regression.

Figure VIII.6: Predicted budget shares and male bargaining power (three-market definition and base specifaction), original model vs reduced form



Note: See footnote in Figure VIII.3 for details.

To try to preserve the basic insights of the non-linear model, specification (VIII.6) is changed in the following way: bargaining power (θ), defined as the income share of the husband, is replaced by the education share: number of years of education of the husband over the sum of the number of years of education of the husband and wife. The logarithm total expenditure is then replaced by the logarithm of the value of the assets owned by the household. Finally, educational attainment for both members, included as additional covariates, is dropped from the main specification. In Table VIII. 12, the regression results are presented for specification (1), where the original estimates of the non-linear model (but changing the estimation sample) are included for comparison purposes. In Figure VIII.6 the two results are compared in terms of the marginal effects and predicted shares.

As shown, changing the definition of power and total expenditure reveals that the positive, but non-linear, association between bargaining power and informal consumption shares is replaced by an almost linear approximation. However, it is worth noting that, although marginal effects have increased, they are still below the IV approximation. In any case, the general impression is that the results provide additional evidence in favour of the hypothesis maintained in this chapter: that empowering women in the household has a negative effect on informal shares. Similarly, it is still possible to verify that these effects are higher when the mother dominates than when father does.[109] In fact, increasing mothers' power by 0.1 pp. in the range θ=0.1 to θ=0.4 generates an increase in informal shares of between 0.013 pp. and 0.011 pp. higher than for similar effects in the range θ=0.6 to θ=0.9, where the estimated results yield impacts of between 0.010 pp. and 0.000 pp. (for mothers with very little power).

The results for formal consumption are also clearer using these new definitions of power and expenditure. Although point value estimates are highly comparable with the original estimation, the perfect U relationship could be replaced by a non-linear, but negative, relationship between male bargaining power and formal shares. Moreover, the results are well determined only when the mother dominates. This means that using education as the variable of power and asset values as expenditure, quantity substitution between formal and informal shares is still detected for sufficiently empowered women. In contrast, empowered husbands use their power only to increase informal consumption, maintaining formal purchases at their minimum.

<u>Third empirical exercise: the disaggregated version of the model</u>

A final exercise proposed is to explore what happens in a disaggregated version of the model in order to analyse possible departures from the hypothesis maintained when the estimations are restricted to particular expenditure groups. The rationale of this new version of the model was explained in Chapter 3. For the purposes of this chapter, it is enough to formalize it in terms of

---

[109] The computed marginal effects at different values of θ (when significant) where statistically different from zero. Chi-squared results: θ=0.1 vs θ=0.2 (2.64), θ=0.1 vs θ=0.3 (3.32*), θ=0.1 vs θ=0.4 (4.26**), θ=0.1 vs θ=0.5 (5.68**), θ=0.1 vs θ=0.6 (8.08***), θ=0.1 vs θ=0.7 (12.63**), θ=0.2 vs θ=0.3 (4.92**), θ=0.2 vs θ=0.4 (6.94**), θ=0.2 vs θ=0.5 (10.30***), θ=0.2 vs θ=0.6 (16.22***), θ=0.2 vs θ=0.7 (25.33***), θ=0.3 vs θ=0.4 (11.28***), θ=0.3 vs θ=0.5 (17.89***), θ=0.3 vs θ=0.6 (26.49***), θ=0.3 vs θ=0.7 (27.41***), θ=0.4 vs θ=0.5 (26.89***), θ=0.4 vs θ=0.6 (27.41***), θ=0.4 vs θ=0.7 (18.21***), θ=0.5 vs θ=0.6 (18.74***), θ=0.5 vs θ=0.7 (11.05***), θ=0.6 vs θ=0.7 (7.00***).

the bargaining structure imposed. In empirical terms, a disaggregated version of the model implies that the effects of bargaining on informal consumption will be different within each particular commodity group. To take this into account, the correct procedure is to estimate the first stage of the problem (the decision to purchase consumption groups) and the second stage (the decision to do so at formal or informal markets within them) using (VIII.6). Under the assumptions in Chapter 3 of weak separability, the specification becomes

$$\begin{aligned} w_k &= \alpha_{kb} + (\alpha_{ka} - \alpha_{kb})\theta + \beta_{ka}\theta \ln(\theta M) + \beta_{kb}(1-\theta)\ln[(1-\theta)M] + \ldots \\ &\ldots + \beta_{k1} z + \beta_{k2}\lambda + \varepsilon_k \end{aligned} \quad \text{(VIII.10)}$$

for the broad group allocation, where w_k is the consumption share for group k and

$$\begin{aligned} w_{kj} &= \alpha_{kjb} + (\alpha_{kja} - \alpha_{kjb})\theta + \beta_{kja}\theta \ln(\theta m_k) + \beta_{kjb}(1-\theta)\ln[(1-\theta)m_k] + \ldots \\ &\ldots + \beta_{kj1} z + \beta_{kj2}\lambda + \varepsilon_{kj} \end{aligned} \quad \text{(VIII.11)}$$

for the allocation of markets inside each broad group allocation, where w_{kj} is the consumption share for market j inside consumption group k and m_k is the total outlay of consumption group k. However, note that (VIII.11) imposes the assumption $\theta = \theta_k$. Given the limited data availability, it was not possible to identify control over resources for each particular consumption group, so income shares are used as a proxy. Using a similar reasoning to that of Hayashi (1995), as formal and informal consumption are normal goods (as demonstrated here and in previous chapters) θ and θ_k will at least be positively correlated, so an empirical exercise like (VIII.11) will still provide useful information.

To estimate the disaggregated model, given the levels of censoring commented on in the previous section, the estimator is no longer the SUR–GLS, but the S&Y of Chapters 6 and 7. The assumptions imposed in the choice model in order to identify the first stage (PROBIT) of the S&Y procedure are similar to those implemented in Monge (2007).

First, bargaining has no effect on the probability of purchasing; it only influences the quantity consumed. This means that PROBIT estimates are specified according to overall household conditions: the bargaining structure is replaced by the logarithm of total expenditure, education and age are replaced by the mean value of the male and the female, individual hours of work by the sum of the male and the female (with a unique participation dummy) and dummies for chronic illness by the combined result of the male and the female. In contrast, the specification of the second stage (the continuous part of the model) is similar to the SUR–GLS model already discussed, i.e., where the bargaining structure is imposed. This assumption could be controversial, as household members could also bargain over the decision to purchase or not to purchase an item. However, given that the focus of the research is mainly on the continuous part (where the sharing rule model and the AIDS specification are better suited) and PROBIT estimates are used only for correctional purposes, this assumption is considered adequate for the purposes of this analysis. At the same time, it helps to avoid non-linearities in the PROBIT (a model already non-linear) and facilitates the computation (and interpretation) of the marginal effects for the different values of θ.

Second, additional covariates are also introduced for identification purposes. The aim is to include on the PROBIT specification a number of variables that only influence the probability of purchasing a good (or a market, conditional on group expenditure), but not its quantity. These covariates are similar to those already used in Chapter 6 but, as already explained, only for the problematic categories. However, we should recognize that the performance of the identification strategy under this new framework is lower than results presented earlier.[110] The results for the first stage PROBIT estimates are shown in Appendix 9.

Under the structure imposed, the approximation of McDonald and Moffit (1980) – holding constant Φ and ϕ – is justified for the computation of marginal effects. Therefore, for non-problematic categories, the marginal effect of θ on w is still (VIII.7), considering only the notational change introduced in (VIII.10) and (VIII.11). However, for the problematic categories, the results have to take the first-step PROBIT estimates into account. Using a similar reasoning as in Chapter 6, the marginal effects now take the form for broad groups:

$$\frac{\partial w_k^{S\&Y}}{\partial\theta} = \begin{bmatrix} (\alpha_{ka} - \alpha_{kb}) + ... \\ ... + \beta_{ka}(1 + \ln(\theta M)) - \beta_{kb}(1 + \ln[(1-\theta)M]) \end{bmatrix} \Phi_k(\pi_{k0} Q_{k1}) \tag{VIII.12}$$

and for markets inside each group:

$$\frac{\partial w_{kj}^{S\&Y}}{\partial\theta} = \begin{bmatrix} (\alpha_{kja} - \alpha_{kjb}) + ... \\ ... + \beta_{kja}(1 + \ln(\theta m_k)) - \beta_{kb}(1 + \ln[(1-\theta)m_k]) \end{bmatrix} \Phi_{kj}(\pi_{kj0} Q_{kj1}) \tag{VIII.13}$$

---

[110] Similar experimentation on the identifying matrix was performed, as in Chapter 6, without improving the results or changing the basic conclusions presented here.

where Q_1 represents the different covariates assumed to influence the purchasing decision, π_0 is the set of parameters associated with them and $\Phi(\pi_0 Q_1)$ the cumulative distribution function.

Table VIII.13: Regression results of the bargaining model (consumption groups – base model)

	Fon	Foff	CC	HEA	TC	ED
θ	0.34594**	-0.03839	-0.08762	-0.03271	-0.17543**	0.00913
	(0.16393)	*(0.13432)*	*(0.07560)*	*(0.11351)*	*(0.08073)*	*(0.07233)*
$\theta\ln(\theta M)$	-0.06746***	-0.03438***	0.00788**	0.00232	0.03183***	0.01645***
	(0.00773)	*(0.00788)*	*(0.00328)*	*(0.00492)*	*(0.00461)*	*(0.00351)*
$(1-\theta)\ln[(1-\theta)M]$	-0.02930**	-0.04369***	-0.00127	0.00001	0.01221*	0.01945***
	(0.01388)	*(0.01143)*	*(0.00663)*	*(0.01015)*	*(0.00738)*	*(0.00699)*
ϕ		-0.08764***		-0.02098**		
		(0.02438)		*(0.00890)*		
$\partial w/\partial\theta$						
$\theta=0.1$	0.09353***	0.11498***	-0.00893	-0.01118	-0.03945***	-0.05847***
	(0.02304)	*(0.01895)*	*(0.01013)*	*(0.01533)*	*(0.01409)*	*(0.01322)*
$\theta=0.2$	0.04332**	0.090278***	-0.00362	-0.00986	-0.01595	-0.04478***
	(0.02026)	*(0.01581)*	*(0.00905)*	*(0.01379)*	*(0.01250)*	*(0.01229)*
$\theta=0.3$	0.01205	0.07342***	-0.00060	-0.00909	-0.00142	-0.03552***
	(0.01857)	*(0.01399)*	*(0.00835)*	*(0.01271)*	*(0.01157)*	*(0.01159)*
$\theta=0.4$	-0.01187	0.05925***	0.00147	-0.00854	0.00962	-0.02778**
	(0.01717)	*(0.01262)*	*(0.00774)*	*(0.01173)*	*(0.01081)*	*(0.01089)*
$\theta=0.5$	-0.03226**	0.04592***	0.00300	-0.00811	0.01895*	-0.02056**
	(0.01592)	*(0.01148)*	*(0.00716)*	*(0.01076)*	*(0.01013)*	*(0.01014)*
$\theta=0.6$	-0.05110***	0.03226***	0.00415	-0.00776	0.02748***	-0.01322
	(0.01484)	*(0.01057)*	*(0.00665)*	*(0.00980)*	*(0.00952)*	*(0.00931)*
$\theta=0.7$	-0.06993***	0.01703*	0.00500	-0.00747	0.03590***	-0.00509
	(0.01427)	*(0.01009)*	*(0.00639)*	*(0.00899)*	*(0.00908)*	*(0.00847)*
$\theta=0.8$	-0.09081***	-0.00198	0.00554	-0.00721	0.04510***	0.00499
	(0.01516)	*(0.01072)*	*(0.00687)*	*(0.00884)*	*(0.00922)*	*(0.00789)*
$\theta=0.9$	-0.11907***	-0.03125**	0.00559	-0.00698	0.05732***	0.02041**
	(0.02032)	*(0.01442)*	*(0.00946)*	*(0.01118)*	*(0.01130)*	*(0.00895)*
Observations	2,105	2,105	2,105	2,105	2,105	2,105
LogL	15183.850					
Chi-sq (overall)	6832.853***					
Chi-sq (bargain structure)	104.43***	38.61***	5.89	0.96	53.93***	35.05***

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

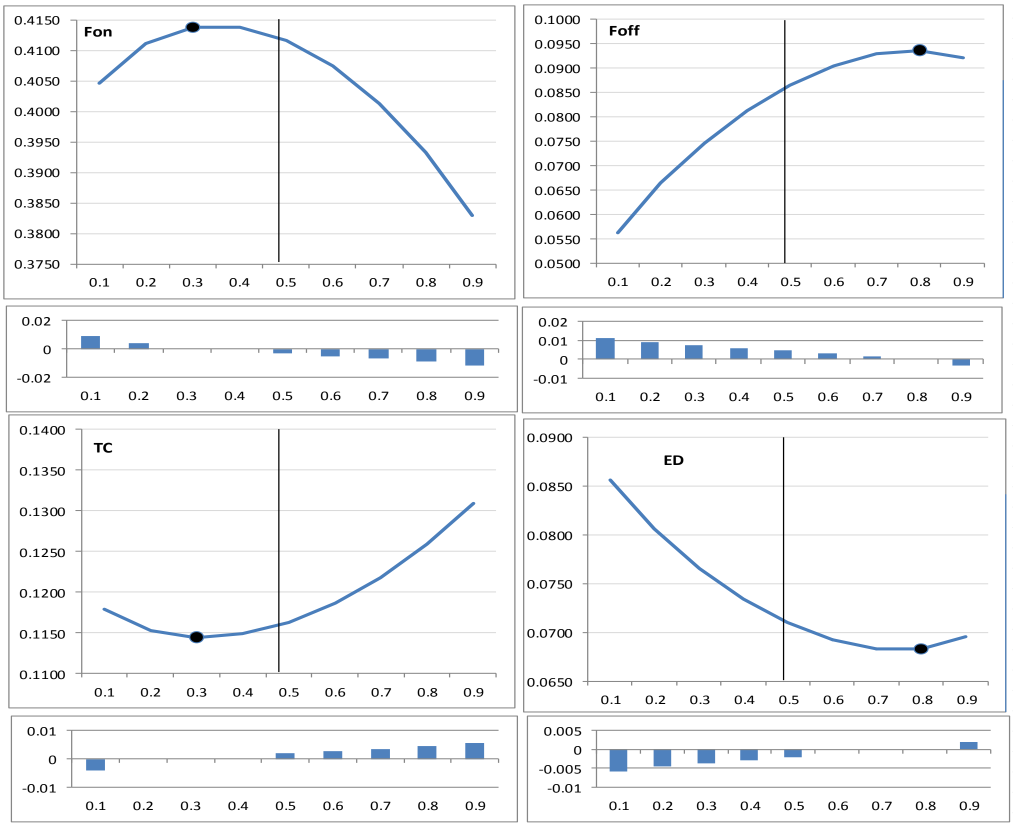
The SUR–GLS results for the allocation process among consumption groups are presented in Table VIII.13, using specification (1). Remember that the groups considered are Fon (Food to be consumed within the house), Foff (Food to be consumed outside the house), CC (clothing and personal care), HEA[111] (health goods and services), TC (transport and communication) and ED (education and culture). It is important to note that the sample size in this application is lower than the broad market application, as more missing values appear for the additional covariates used in the first stage PROBIT estimates. At the bottom of the table, the marginal effects, computed using the corrected formulas, are also provided for the different realizations of θ.

The bargaining structure imposed is significant in four out of six consumption groups: Fon, Foff, TC and ED. On both CC and HEA, joint significant tests are weak enough to cast doubt about the adequacy of the collective approach to these goods. The results for these commodities also reveal that the marginal effects of θ are not well determined for the different values computed. Therefore, it is better to conclude that the relationship between θ and these groups' allocation is a horizontal line – meaning there is no influence of bargaining in these cases. This is a normal result since, in most revised applications, it is usual to find that households bargain in some groups and not in others (see Monge, 2007). Although, Monge (2007) finds similar evidence for of health expenditure (not bargaining), the result is striking for CC given that the group is mainly made up of individual expenditure. The aggregation of the personal expenditure of both males and females in a unique group is probably obscuring the identification of negotiation patterns (in Monge, 2007 they are separated into child and adult expenditure). However, this further disaggregation of the group is not possible in this application since it will reduce the ability to identify markets.

The analysis for the other groups is more interesting, as most of the marginal effects are significant at conventional levels. To facilitate interpretation, the predicted shares for the different values of θ are presented in Figure VIII.7. The results are similar to those in the literature that argue in favour of women (or mothers) being closer to consumption on pro-household goods (or consumption groups) than men. For example, increasing the father's bargaining power in the house has, in general, a negative relation with the expenditure shares of food to be consumed in the house, and education and culture, while having a positive relation with food to be consumed outside the house, and transport and communication. Therefore, while the mother is more interested in using her power to spend on family goods (like food or education), the father will be more interested in using it for expenditure outside the house (like meals at work and transportation) which are possibly closer to his income-generating activities.

[111] For health expenditure, since chronic illness is implemented as an individual variable in the bargaining model, this variable has been replaced in the PROBIT model by the number of household members with an eventual health episode that needs attention, such as an accident or serious illness. This is the only difference with the specification in Chapter 6.

Figure VIII.7: Predicted budget shares and male bargaining power on broad groups



Note: See footnote in Figure VIII.3 for details.

The studies of Guyer (1980), Guyer and Peters (1987) and Bruce (1989) quoted in Hoddinott and Haddad (1995) are generally along these lines: gender roles in the household typically define mothers as responsible for household activities like acquiring and preparing food, while personal expenditure is more in the fathers' sphere. The survey by Garavito (1997) also suggests similar patterns for the Peruvian case, with women closer to 'within the household' activities while men are closer to 'outside the house' ones. This type of result is also encountered in Thomas and Chen (1994) for Taiwan, Hoddinott and Haddad (1995) and Duflo and Udry (2004) for the Côte d' Ivoire, Phipps and Burton (1998) for Canada, Thomas (1997) for Brazil, and Quisumbing and de la Brière (2000) and Quisumbing and Maluccio (2000) for various African countries.

However, it is important to stress two areas here which differ from previous applications. First, in terms of the interpretation of results, it is not necessarily true that fathers prefer less pro-welfare goods than mothers. Given the limited availability of the data used here, it is only possible to conclude that the two family members have different approximations to household welfare. On the one hand, the mother is interested in the children's consumption and, on the other, the father concentrates on income-generating activities. This is closer to the conclusions of Iguiñiz (1996), where gender roles are understood using the notion of time allocation in the house, with females responsible for activities related to the provision of capabilities and men responsible for production activities.

Second, as already discussed in the model for broad markets, this literature has overlooked non-linearities in the relationship between power and consumption and the differentiated influence they have when members are in a dominant position. For example, increasing the wife's control over money (moving from the right to the left) will increase food consumption within the household, but this is true only until she reaches a dominant position in the household. A maximum level of food allocation is reached at around $(1-\theta)$ =0.6 or $(1-\theta)$ =0.7, after which it will start to decrease for sufficiently empowered women. The implication is that reinforcing female power in the house in order to increase food consumption is a reasonable but limited policy. In contrast, when the female dominates, increasing her power will lead to clearer effects on the budget shares of education and culture. This result is reasonable if a change in preferences between short-term consumption and human capital investment is allowed for females once they are sufficiently empowered and have already committed the household to a relatively high food allocation.

Similar results are found for the father's response to bargaining power. Increasing his power will lead to increases in meals at work until he has enough power. At this point, he will keep this consumption at this maximum level and, instead, will use his power to increase consumption on transport and communication. The evidence seems reasonable considering that, from the estimates of the expenditure elasticities computed in Chapter 6, transport and communication, in contrast to food outside the house, is classified as a luxury. Therefore, it would be relatively easy for the man to increase his workplace meals when he is not sufficiently empowered, but more difficult if he wants to purchase a luxury – he would need more power to do this.

Once the first stage is solved, it is possible to estimate the second stage: the allocation across markets conditional on the total outlay of consumption groups. The results – only for the base specification (1) – are presented in Table VIII.14. The exercises carried out for the other specifications reveal minimal departures from the estimates presented here and are omitted for the sake of brevity. Note that samples sizes differ between the estimated equations, as each is estimated just for the subsample where broad group expenditure is positive. The global adjustment seems adequate in each regression, but the joint test for the bargaining structure and significance of the marginal effect of θ on W is not achieved in all cases. From the results that the effect of bargaining power on market allocation is dubious in the case of Foff and HEA. On the other hand, the results for Fon, CC, TC and ED are better determined.

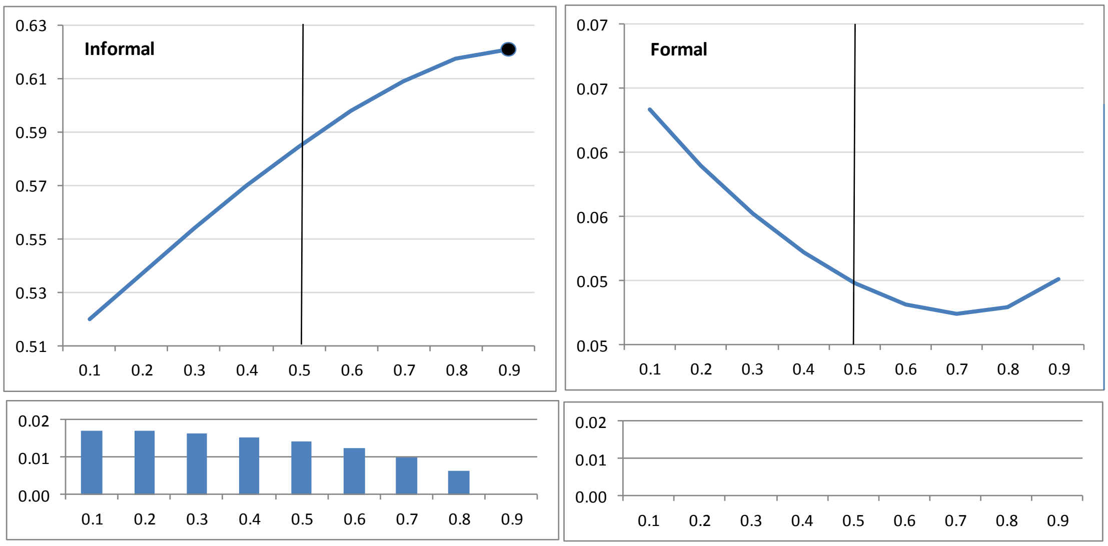
Table VIII.14: Regression results of the bargaining model (markets within each consumption group – base model)

	Fon		Foff		CC		HEA		TC		ED	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
θ	-0.81048***	0.46787	0.20671	0.45732*	0.34401	-0.00971	0.24175	-0.13556	0.04227	-0.08267	0.05886	0.10023
	(0.27644)	*(0.54355)*	*(0.27672)*	*(0.23479)*	*(0.21437)*	*(0.27672)*	*(0.54383)*	*(0.17860)*	*(0.23010)*	*(0.24588)*	(0.14459)	(0.10630)
$\theta \ln(\theta M)$	0.01511	0.00988	-0.15215***	-0.07700***	-0.07074***	-0.06351***	-0.07523***	0.01480	-0.07286***	0.12045***	-0.10708***	0.11529***
	(0.01402)	*(0.02599)*	*(0.01686)*	*(0.01520)*	*(0.01292)*	*(0.01909)*	*(0.02808)*	*(0.01120)*	*(0.01553)*	*(0.01583)*	(0.00997)	(0.00801)
$(1-\theta)\ln[(1-\theta)M]$	-0.09206***	0.06727	-0.12480***	-0.01675	-0.03765	-0.07855**	-0.03100	-0.01219	-0.06602***	0.10945***	-0.10556***	0.13507***
	(0.02689)	*(0.05189)*	*(0.03224)*	*(0.02698)*	*(0.02520)*	*(0.03243)*	*(0.06548)*	*(0.02212)*	*(0.02514)*	*(0.02573)*	(0.01634)	(0.01245)
ϕ		-0.03192	0.05354	0.05364	-0.36065***	-0.00037	0.21562***	0.20992**	0.50701***	0.50485***	-0.00341	
		(0.03285)	*(0.08386)*	*(0.08351)*	*(0.06138)*	*(0.05201)*	*(0.06948)*	*(0.08770)*	*(0.03797)*	*(0.03435)*	(0.05390)	
$\partial w / \partial \theta$												
$\theta = 0.1$	0.17078***	-0.04738	0.17380***	0.09591*	0.19663***	0.08164***	0.01242	0.03629	0.12621***	-0.20873***	0.21312***	-0.31388***
	(0.04462)	*(0.03413)*	*(0.05110)*	*(0.05266)*	*(0.04260)*	*(0.02100)*	*(0.02341)*	*(0.04266)*	*(0.04491)*	*(0.04088)*	*(0.02839)*	*(0.03178)*
$\theta = 0.2$	0.17041***	-0.04017	0.10766**	0.05521	0.15091***	0.0644***	0.00168	0.04410	0.07501*	-0.12918***	0.14758***	-0.21806***
	(0.04124)	*(0.03007)*	*(0.05019)*	*(0.05110)*	*(0.03977)*	*(0.01783)*	*(0.02195)*	*(0.04068)*	*(0.04107)*	*(0.03711)*	*(0.02672)*	*(0.03107)*
$\theta = 0.3$	0.16424***	-0.03383	0.06453	0.03061	0.12209***	0.05267***	-0.00499	0.04798	0.04131	-0.07681**	0.10408***	-0.15328***
	(0.03897)	*(0.02721)*	*(0.04932)*	*(0.04995)*	*(0.03792)*	*(0.01594)*	*(0.02106)*	*(0.03933)*	*(0.03875)*	*(0.03491)*	*(0.02589)*	*(0.03061)*
$\theta = 0.4$	0.15440***	-0.02738	0.02985	0.01242	0.09973***	0.04284***	-0.01008	0.05008	0.01395	-0.03429	0.06847***	-0.09929***
	(0.03679)	*(0.02441)*	*(0.04830)*	*(0.04874)*	*(0.03625)*	*(0.01454)*	*(0.02039)*	*(0.03808)*	*(0.03677)*	*(0.03307)*	*(0.02532)*	*(0.03010)*
$\theta = 0.5$	0.14099***	-0.02032	-0.00137	-0.00246	0.08036**	0.03362**	-0.01440	0.05104	-0.01090	0.00436	0.03584	-0.04894*
	(0.03453)	*(0.02141)*	*(0.04706)*	*(0.04736)*	*(0.03463)*	*(0.01346)*	*(0.01987)*	*(0.03681)*	*(0.03484)*	*(0.03134)*	*(0.02489)*	*(0.02947)*
$\theta = 0.6$	0.12320***	-0.01211	-0.03196	-0.01553	0.06215*	0.02420*	-0.01837	0.05102	-0.03552	0.04264	0.00326	0.00222
	(0.03224)	*(0.01824)*	*(0.04559)*	*(0.04581)*	*(0.03308)*	*(0.01273)*	*(0.01959)*	*(0.03556)*	*(0.03293)*	*(0.02969)*	*(0.02462)*	*(0.02873)*
$\theta = 0.7$	0.09905***	-0.00192	-0.06463	-0.02780	0.0436	0.01372	-0.02233	0.04993	-0.06207**	0.08395***	-0.03220	0.05885**
	(0.03035)	*(0.01564)*	*(0.04391)*	*(0.04413)*	*(0.03190)*	*(0.01261)*	*(0.01979)*	*(0.03453)*	*(0.03126)*	*(0.02841)*	*(0.02467)*	*(0.02791)*
$\theta = 0.8$	0.06374**	0.01203	-0.10367**	-0.04036	0.0224	0.00067	-0.02668	0.04730	-0.09414***	0.13385***	-0.07539***	0.12901***
	(0.03036)	*(0.01655)*	*(0.04218)*	*(0.04264)*	*(0.03200)*	*(0.01377)*	*(0.02110)*	*(0.03438)*	*(0.03068)*	*(0.02845)*	*(0.02547)*	*(0.02722)*
$\theta = 0.9$	0.00171	0.03536	-0.16114***	-0.05557	-0.0070	-0.01937	-0.03252	0.04136	-0.14188***	0.20817***	-0.14027***	0.23621***
	(0.03770)	*(0.02781)*	*(0.04157)*	*(0.04306)*	*(0.03693)*	*(0.01813)*	*(0.02558)*	*(0.03784)*	*(0.03489)*	*(0.03356)*	*(0.02871)*	*(0.02774)*
Observations	2,090		1,794		2,097		1,732		2,011		1,997	
LogL	1251.457		-1059.761		921.617		21.343		803.544		151.769	
Chi-sq (overall)	1009.947***		22722.731***		6878.425***		28690.601***		307233.119***		3672.620***	
Chi-sq (bargain structure)	18.53***	2.33	129.20***	29.20***	41.29***	17.42***	9.22**	3.18	34.41***	95.51***	171.11***	46.40***

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. M refers to group expenditure, different in each subsystem. See footnote in Table VIII.7 for other details.

However, the most interesting case is food within the household (Fon). The predicted budget shares for formal and informal purchases within this consumption group are presented in Figure VIII.8. As shown, the bargaining power of the male is positively related with informal food consumption and describes an almost linear relationship. Therefore, the estimates show that increasing the mother's power in the house will reduce the family's informal food consumption whether or not her position is dominant.

Figure VIII.8: Predicted budget shares and male bargaining power on market allocation (formal and informal) within Fon (food consumed within the household)



Note: See footnote in Figure VIII.3 for details.

Point estimates reveal that, as the female's control over money increases, the marginal effects[112] are greater. When she is in a dominant position (θ is between 0.1 and 0.4), increasing her income share by 0.1 pp. will lead (on average) to a reduction of 0.002 in informal allocation. By contrast, when she is not in a dominant position (θ is between 0.6 and 0.9), a similar increase leads to a reduction of around 0.0007 pp. in the informal consumption of food within the house. The case is different for formal consumption, where bargaining power has no effect on resource allocation. Although the relationship between formal food shares and male bargaining power depicts a

[112] Chi-squared results: θ=0.1 vs θ=0.2 (0.00), θ=0.1 vs θ=0.3 (0.18), θ=0.1 vs θ=0.4 (0.68), θ=0.1 vs θ=0.5 (1.51), θ=0.1 vs θ=0.6 (2.69), θ=0.1 vs θ=0.7 (4.19**), θ=0.1 vs θ=0.8 (5.98**), θ=0.1 vs θ=0.9 (8.02***), θ=0.2 vs θ=0.3 (1.07), θ=0.2 vs θ=0.4 (2.17), θ=0.2 vs θ=0.5 (3.51*), θ=0.2 vs θ=0.6 (4.99**), θ=0.2 vs θ=0.7 (6.52**), θ=0.2 vs θ=0.8 (8.02***), θ=0.2 vs θ=0.9 (9.48***), θ=0.3 vs θ=0.4 (3.82*), θ=0.3 vs θ=0.5 (5.31**), θ=0.3 vs θ=0.6 (6.73***), θ=0.3 vs θ=0.7 (8.02***), θ=0.3 vs θ=0.8 (9.18***), θ=0.3 vs θ=0.9 (10.21***), θ=0.4 vs θ=0.5 (6.81***), θ=0.4 vs θ=0.6 (8.02***), θ=0.4 vs θ=0.7 (9.05***), θ=0.4 vs θ=0.8 (9.92***), θ=0.4 vs θ=0.9 (10.67***), θ=0.5 vs θ=0.6 (9.00***), θ=0.5 vs θ=0.7 (9.79***), θ=0.5 vs θ=0.8 (10.43***), θ=0.5 vs θ=0.9 (10.97***), θ=0.6 vs θ=0.7 (10.34***), θ=0.6 vs θ=0.8 (10.80***), θ=0.6 vs θ=0.9 (11.19***) θ=0.7 vs θ=0.8 (11.08***), θ=0.7 vs θ=0.9 (11.35***), θ=0.8 vs θ=0.9 (11.48***).

negative correlation, and considering the hypothesis-testing for the marginal effects at different levels of θ (where significance is not achieved for any value), it is better to think in terms of a horizontal line around the mean. Therefore, it seems that, on average, the family is committed to a fixed level of formal food consumption and neither of the parents will use their power to deviate from that situation.

In sum, estimates for overall group consumption reveal that increasing the power of the mother within the house will result in an increase in the allocation of resources in favour of food to be consumed within the household up to a maximum level reached when the mother has enough power. This result is in line with the anthropological tradition of the greater pro-welfare or pro-household expenditure of mothers. At the same time, the results for market allocation within this consumption group reveal that increasing the mother’s bargaining power will also cause a reallocation of food purchases away from informal markets, increasing the average quality of this particular item. This result is again in line with the anthropological tradition of the greater pro-welfare expenditure of mothers, but goes one step further since it reveals that it is not only the quantity of pro-household items which increases, but also the quality.

Interestingly, we can also see that the quantity substitution of informal food promoted by the mother in the house is clearer with semi-formal categories than formal ones. These last seem to be less affected by the bargaining power of the mother and much more affected by the average level of income of the household. This pattern of results is different to that found for overall market consumption, where quantity substitution promoted by the mother was between informal and formal markets. The difference can be explained if we consider the characteristics of food consumption: perishable, with higher purchasing frequencies. Under these conditions, access to the markets will be crucial for substitution possibilities in the short term. As mentioned in previous chapters, access to formal outlets can be complicated if there are supply limitations, as noted in Arellano and Burgos (2010). Thus, if this is not true for semi-formal outlets (and we presume that it is not), it will be easier for the mother to change an informal commercialization channel for a semi-formal than for a formal one. This last substitution will possibly require a higher investment (in time and money) so the effect of power (holding constant total income) should reasonably be lower.

We can therefore conclude that, although increasing the bargaining power of the mother will generate less-informal consumption bundles, the effect is not equal across all consumption groups. It is more reliable, for example, for food within the household than for other categories. This is a reasonable result if, as stated previously, obtaining and preparing food is within the mother’s sphere, as she will be much more interested in preserving the quality of this type of consumption than of others. This evidence is in line with anthropological tradition and comparable with results obtained in Thomas (1997), where the positive effects of women's bargaining power on the nutritional status of children are found (see also Xu, 2007 for other examples). Among other positive decisions that more-empowered mothers can make (like better health-care practice) in order to improve nutrition, the results presented here suggest the existence of at least one additional channel: a better-quality source of food.

Given the relevance of this result, a final exercise proposed is to inspect the effect of bargaining power on informal food consumption shares, conditional on total food expenditure explicitly handling the endogeneity of expenditure and bargaining power. Considering the low level of censoring of this purchasing category, the IV strategy proposed earlier can be replicated in this new context in order to inspect the robustness of the conclusions arrived at. In Table VIII.15, the diagnosis tests are presented. The information provided by these tests is similar to that found previously: the instruments included pass both the relevance and the orthogonality tests, and the exogeneity tests of potential endogenous variables reveal the need to instrument them. The results also perform better under a specification that includes labour supply in terms of the correlation between excluded instruments and endogenous variables. Similarly, as we are able to identify some relative bias in the estimates, it is better to perform the regression results under different IV techniques. The first-stage results are presented in Appendix 9 and most of the additional instruments display similar results to earlier ones.

Table VIII.15: Relevance and orthogonality tests of the IV application (informal food shares)

	With labour supply				Without labour supply			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Under-identification Test	71.238***	70.693***	68.882***	71.444***	55.201***	53.924***	53.203***	55.467***
F-value (1stage, theta)	26.02***	24.24***	26.73***	25.86***	20.62***	19.24***	21.66***	20.69***
F-value (1stage, expend)	79.50***	61.05***	77.81***	78.79***	64.12***	47.87***	62.65***	63.35***
F-value (1 stage, joint)	16.021	15.911	15.416	16.03	10.244	10.017	9.860	10.280
Hansen J test	1.791	2.394	0.774	1.641	2.197	3.752	1.127	2.137
C (labour)	0.731	1.627	0.064	0.569	.	.	.	.
C (theta,expend)	15.704***	9.804***	18.391***	16.918***	18.484***	11.467***	21.247***	19.661***

Note: *** p<0.01, ** p<0.05, * p<0.1. Tests are performed under a 2SLS specification with small sample correction and robust standard errors. Individual F-statistics must be compared with the rule of thumb of 10 (Cameron and Trivedi, 2005), Joint F-statistics are compared with the critical value of 13.97 and 15.72 (5% of maximal IV relative bias) computed in Stock and Yogo (2005) and 8.78 and 9.48 (10% of maximal IV relative bias) as the minimum tolerable levels with and without labour supply, respectively. See footnote in Table V.7 for other details in the tests.

In Table VIII.16, these results are presented for the OLS version, the standard 2SLS and the additional LIML and GMM under the different specifications. Note that the sample size has change compared to Table VIII.8, since the missing values for additional regressors included to identify PROBIT models on other consumption categories are not dropped, and the missing values for instruments are taken into account. The results are in line with previous conclusions. First, in terms of the group expenditure covariate, this has a positive effect on informal shares – a result already found in previous chapters and not deeply explored here. Second, the bargaining power variable shows a positive effect on informal consumption once endogeneity is taken into account. This result confirms the negative impact on informal food consumption of increasing the bargaining power of the mother and the possible downward bias of OLS estimates.

Table VIII.16: IV results of the bargaining model for informal food consumption share (three-market definition, different specifications and estimation methods)

	With labour supply				Without labour supply			
(1)	OLS	2SLS	LIML	GMM	OLS	2SLS	LIML	GMM
θ	0.00786	0.26964**	0.27536**	0.27767**	-0.00658	0.32146***	0.33365***	0.33881***
	(0.02245)	(0.10839)	(0.11060)	(0.10804)	(0.02192)	(0.12323)	(0.12776)	(0.12227)
L(expend_Fon)	0.00900	0.11017***	0.11136***	0.11088***	0.00628	0.12870***	0.13131***	0.13052***
	(0.01141)	(0.02952)	(0.02986)	(0.02951)	(0.01145)	(0.03305)	(0.03384)	(0.03294)
Observations	2,142	2,142	2,142	2,142	2,142	2,142	2,142	2,142
LogL	-273.311	-391.546	-395.423	-396.032	-278.260	-461.121	-471.408	-473.564
F-value (overall)	8.65***	8.42***	8.39***	8.40***	8.90***	8.40***	8.33***	8.45***
(2)	OLS	2SLS	LIML	GMM	OLS	2SLS	LIML	GMM
θ	0.01928	0.26080**	0.26812**	0.27026***	0.00640	0.29603**	0.31477**	0.31355***
	(0.02183)	(0.10449)	(0.10747)	(0.10428)	(0.02138)	(0.11722)	(0.12467)	(0.11672)
L(expend_Fon)	-0.01245	0.06604**	0.06760**	0.06687**	-0.01500	0.08324**	0.08753**	0.08537**
	(0.01138)	(0.03096)	(0.03146)	(0.03095)	(0.01141)	(0.03443)	(0.03586)	(0.03435)
Observations	2,142	2,142	2,142	2,142	2,142	2,142	2,142	2,142
LogL	-210.055	-302.536	-307.291	-307.731	-214.005	-355.022	-370.356	-367.416
F-value (overall)	12.47***	11.36***	11.30***	11.38***	12.99***	11.05***	10.89***	11.16***
(3)	OLS	2SLS	LIML	GMM	OLS	2SLS	LIML	GMM
θ	0.00724	0.25840**	0.26097**	0.25814**	-0.00667	0.31028**	0.31653**	0.31574**
	(0.02245)	(0.11021)	(0.11122)	(0.10971)	(0.02190)	(0.12534)	(0.12769)	(0.12415)
L(expend_Fon)	0.01151	0.12884***	0.12944***	0.12875***	0.00892	0.14936***	0.15083***	0.14964***
	(0.01148)	(0.03084)	(0.03100)	(0.03080)	(0.01151)	(0.03488)	(0.03532)	(0.03471)
Observations	2,142	2,142	2,142	2,142	2,142	2,142	2,142	2,142
LogL	-263.974	-394.581	-396.401	-394.368	-268.538	-467.245	-472.691	-470.667
F-value (overall)	8.40***	8.09***	8.07***	8.08***	8.72***	8.04***	8.00***	8.05***
(4)	OLS	2SLS	LIML	GMM	OLS	2SLS	LIML	GMM
θ	0.00941	0.26260**	0.26770**	0.26993**	-0.00472	0.31076**	0.32224**	0.32661***
	(0.02227)	(0.10729)	(0.10928)	(0.10697)	(0.02177)	(0.12163)	(0.12594)	(0.12067)
L(expend_Fon)	0.00991	0.11542***	0.11651***	0.11604***	0.00730	0.13386***	0.13638***	0.13544***
	(0.01128)	(0.02938)	(0.02969)	(0.02937)	(0.01132)	(0.03292)	(0.03368)	(0.03280)
Observations	2,142	2,142	2,142	2,142	2,142	2,142	2,142	2,142
LogL	-264.031	-383.647	-387.124	-387.652	-268.742	-450.639	-460.325	-461.713
F-value (overall)	8.73***	8.83***	8.81***	8.83***	9.00***	8.77***	8.69***	8.81***

Note: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. Note that the table is divided into four parts according to each of the specifications used in the present application.

VIII.6. Concluding remarks

The empirical application presented in this chapter results in four additional conclusions regarding informal consumption. First, family members in Peruvian households bargain and they do it both to decide which goods they will purchase and where these goods will be bought (formal or informal markets). Therefore, demand-side incentives that reinforce the control over money (or the bargaining power) of particular household members can be exploited not only to achieve particular (socially desirable) consumption allocations, but also as formalization policies. This conclusion is robust for the different assumptions of preference formation or budgeting approaches, the specifications of the demand equations implemented, and the definitions of informality and once the endogeneity of income shares and total expenditure is taken into account in the models.

Second, the direction of the effects of bargaining on consumption is in line with the anthropological tradition that emphasizes that households with more-empowered wives achieve greater pro-welfare consumption. This type of result is found in this application in the form of less-informal (i.e. better-quality) consumption baskets associated with the mother's higher income shares. According to the results for broad markets, the models predict that informal consumption shares are higher when the husbands are in a dominant position and minimized when the wife is in a similar situation. This result was also confirmed using the IV model under the linear approximation of the bargaining structure, revealing that a 0.10 pp. increase in the mother's income share is associated with reductions between 0.025 pp. and 0.027 pp. on informal budget shares. However, according to more theoretical models, the relationship is far from linear. In fact, the effect of bargaining power on consumption critically depends on the relative level of power enjoyed by each household member and on whether or not they are in a dominant position. In this application, it was found that the effect of the mother'' power is stronger when she are in a dominant position. Interestingly, under alternative definitions of power (like the education share) the results were stronger (in terms of point estimates) and clearer (in terms of significance at conventional levels).

Third, the results for formal consumption are less clear. On the one hand, the linear models that control for endogeneity reveal that household members do not bargain in terms of formal consumption and only the absolute level of income is relevant for increasing or decreasing this consumption. On the other hand, non-linear models reveal that, when each member is adequately empowered, increasing his or her power will lead to higher formal consumption, with the marginal effects statistically equal. As a result, an almost perfect U-shaped relationship between bargaining power and formal budget shares was depicted. Therefore, in these models, bargaining cannot be rejected, but it was not possible to draw conclusions on the existence of differentiated preferences. In fact, it is better to say that each member will be interested in using his or her power to increase formal consumption within the household. This evidence was reconciled with previous evidence using the predicted behaviour of semi-formal markets. As a result, the models predict that the two family members in a dominant position will increase formal consumption as their power increase, but only the mothers will do so via a reduction in informal consumption. The husbands will keep informality at its maximum level and will adjust semi-formal consumption.

Fourth, once the models were disaggregated in terms of consumption groups, we were able to verify that the negative correlation between the female’s bargaining power and informal consumption was more plausible for some goods than for others. This is an intuitive result, since it is natural to think that the effects of bargaining will be different across goods and mothers will be much more interested in preserving the quality only of some types. In this application, the clearer evidence in favour of this hypothesis was for food within the house. This is an understandable result given that food is a typical type of consumption in the mother’s sphere. In fact, when the models were solved for particular consumption groups, a positive correlation between the mother´s bargaining power and food shares was found. Therefore, as a result it is possible to argue that increasing wives’ control over money will not only result in higher food consumption, but also in better-quality sources (less-informal) for these goods. This result was also confirmed when IV models were used for estimation. From a public-policy perspective, this evidence suggests the possibility of exploiting targeted income transfers to the mother in order to attain better nutritional outcomes for the children.

IX. CONCLUSIONS

The thesis presented here is one of the first attempts to study informality from a demand side perspective. In the past this issue has received scant scholarly attention, limiting our understanding of informality as an economic phenomenon, the public policies available to reduce it in developing countries and the possible consequences of those policies. For this reason the approach taken was to investigate the main demand properties of informal consumption in order to test specific hypotheses on its responses to income, prices, labour supply and the bargaining power of household members. The main database used in the research was the Peruvian National Household Survey (ENAHO, in Spanish) conducted by the National Office of Statistics (INEI, in Spanish) during 2006, and restricted to urban households. The attractive property of this survey is that it permits identification of where households make their purchases, the working conditions of household members, the unit values at the district level for food consumption and individual labour and non-labour earnings for household members in order to obtain detailed bargaining-power measures.

The main specification used in the research was the Almost Ideal Demand System (AIDS), given its desirable theoretical and empirical properties. However, several econometric problems arose during estimation: heteroskedasticity, cross-equation correlations, endogeneity and censoring bias. To partially control for these issues, different econometric techniques were implemented: Seemingly Unrelated Regression methods with robust standard errors (SUR–GLS), Instrumental Variables estimations (IV) with enough caution when weak instruments were present and Shonkwiller and Yen's (1999) procedures with within- and between-regression robust standard errors (S&Y).

In general, the conclusions of the application can be organized in four groups. First, in the case of total expenditure, increasing the total available resources of the family increases formal shares and reduces informal ones. This result is consistent with the quality differentials attached to each market. The results also allow us to argue in favour of the existence of family protective attitudes, since families will bias their consumption baskets towards informal outlets as income diminishes, or, similarly, in favour of the formalization power of economic growth, as long as it will generate more-formal purchasing baskets at the household level. However, the inferiority of informal consumption was decisively rejected, even for the richest individuals in the sample, for the most informal markets and for every consumption group where the hypothesis was tested. This means that both informal and formal consumption are normal, with the former classified as necessity and the latter as luxury. Therefore, increasing the expenditure levels of the family in highly informal countries will increase both formal and informal consumption, but the former at a higher speed than the latter. Using the disaggregated version of the model, this result is replicated in four out of six consumption groups: food to be consumed within the household, food to be consumed outside the household, transport and communication, and education and culture. In the other two groups – clothing and personal care, and health goods and services – total expenditure seems to have the same influence on both formal and informal segments of the market.

Second, the linkage between working and purchasing from informal markets under different estimation methods, specifications and definitions of informal markets is a strong result for a developing country like Peru. Therefore, working more hours in the informal (formal) sector is highly correlated with buying more goods in informal (formal) markets. The intuition behind the linkage hypothesis is that previous experience in a sector (as a worker) helps the consumer to reduce transactional costs (network linkage) and reveals the worker's preference to remain there as a consumer (preference linkage or non-separability effect). However, a strong result also found in the application is that not all formal or informal hours of work have the same effect on consumption. In fact, the evidence suggests that the main channel for the positive linkage with informal consumption occurs through the informal self-employment sector. This is an intuitive result, since it was expected that stylized facts that support linkages (like informational advantages or the preference to keep hidden) are stronger for those informal workers who are also producers. At the same time, the positive linkage with formal consumption occurs mainly through the formal self-employment sector. This is also an intuitive result, considering that only those in this formal labour segment will receive tax advantages if they bias household consumption through formal markets. These results were accompanied by the negative linkage of the informal wage-earning sector on formal consumption, possibly driven by greater difficulties acceding to formal markets by this segmented or refugee labour option.

Using the disaggregated version of the model, when particular consumption groups are modelled the evidence in favour of linkages is found in only three out of six consumption groups. The strongest evidence is found in food to be consumed outside the household – an intuitive result considering that this group is mainly composed of workplace meals, a type of consumption where informational advantages are probably better grounded and the location decisions of vendors could generate stronger linkages between market participants. At the same time, the hypothesis is verified in its weaker version for the two consumption groups transport and communication, and education and culture. In the other consumption groups no conclusive evidence was found. Interestingly enough, two out of the three cases where the linkage hypothesis is verified (meals and travel to work) correspond to consumption groups previously defined as quantity substitutes for leisure, so we could at least suggest that, for in this type of goods, the linkage hypothesis seems to be more plausible.

Third, the inclusion of prices in the analysis confirms the expected demand patterns for formal and imformal markets: negative compensated and uncompensated own-price elasticities and positive income effects. It was also possible to generally confirm elastic observed demand curves for both types of consumption. Point value estimates for uncompensated own-price elasticities are around or greater than 1, which is probably evidence of high substitution possibilities between sectors. Thus, it is possible at least to suggest some degree of integration of formal and informal food markets in highly informal countries. Cross-price elasticities also confirm that formal and informal consumption can be treated as imperfect substitutes, with higher substitution effects in the case of formal consumption than for informal. In fact, when compensated elasticities are analysed, the data support formal markets being more sensitive to price shocks (both own- and cross-) than informal ones. This is an intuitive result, which considers that informal channels of commercialization could be more flexible (in terms of adaptive strategies to consumers) and diverse (in the type and presentation of products sold there), so informal to formal substitution possibilities could be greater than the other way around. Also, if confidence links between market participants in the informal sector explains at least some of consumption patterns there, it seems obvious to presume that informal markets are less responsive to demand shocks.

Fourth, family members in Peruvian households bargain, and they do so both when deciding which goods they will purchase and where these goods will be bought (formal or informal markets). The direction of the effects of bargaining on consumption is also clear and in line with the anthropological tradition that emphasizes that households with more-empowered wives reach more pro-household consumption levels than they would otherwise. This effect is detected in this application both in the form of higher proportion of “family goods” associated with more empowered wifes and less less-informal (better quality) consumption baskets. However, the relationship is not necessarily linear. We can see that the effects of bargaining power on consumption critically depends on the relative level of power enjoyed by each member and whether or not they are in a dominant position. Stronger results against informal consumption are found when the mother is in a dominant position. The results for formal consumption are less clear. Using non-linear models, although bargaining cannot be rejected, it was not possible to conclude that differentiated preferences across household members exist. Using linear models, it seems that income shares have no effect on formal consumption and only the absolute level of income matters.

Once collective models were disaggregated in terms of consumption groups, it was possible to verify that the negative correlation between wives’ bargaining power and informal consumption was more plausible in the case of food consumption in the house. This is an understandable result given that food is a typical type of consumption in the mothers' sphere and she will be more interested in preserving quality here than in other items.

With this evidence, several policy suggestions emerge. First, economic growth will have a positive effect on formalizing the economy from the demand side, but this a slow process that will occur with limitations. In fact, the evidence presented here is consistent with informal markets being less responsive to income than formal ones, with high levels of informal consumption (around 30% and 15% of their budget allocated there), even for the richest individuals in the sample. At the same time, additional experimentation with the data reveals that income effects differ across households. Formalization driven by positive income shcks tend to be higher for the richest households, those living in the more developed cities of the country and those working in the modern sector of the economy. For the poorest households, small targeted income transferences to the poor could come even with the short-term cost of increasing informality. Therefore, any pro-poor growth policies promoted by governments need to be strong enough to bring about structural changes in the living conditions of the most deprived and also to procure structural changes in their market allocation. Most of these policies are closer to development plans that reinforce income-generating capacity (for example, through asset access or human capital investments) or improve access to better-quality jobs.

Second, as long as working and purchasing from the informal sector are linked activities, policy-makers are possibly underestimating the effects of their public policies on informality in the supply side of the shadow economy. For example, policies that reduce hours of work in the informal sector and increase those for formal jobs will have second-round effects through the expected re-composition of formal and informal demand (in favour of more formal consumption baskets). The highest effectiveness of a supply-side policy such as this is presumed to occur if the focus is on the self-employment sector. At the same time, higher recompesition effects are expected in consumption groups such as meals at work, travel to work and educational expenditure. A lower level of effectiveness is presumed on food within the household, clothing and personal care and health expenditures.

Third, changing the relative prices could also be exploited to formalize the economy. The simulations implemented reveal that combining informal price increases with formal price reductions will naturally be the most effective alternative, and such a combination will help to minimize the welfare impacts on households. However, these policies are not free of costs. In the cases presented in this application, the main cost is the unequal distribution of welfare effects, with the winners being at the top of the income distribution chain and the losers at the bottom. A natural response for a result like this is to suggest targeted income transferences in order to compensate for differentiated welfares impacts. Nevertheless, considering the differentiated effects of income on informal consumption, a decision along these lines could off-set the expected formalization results. For this reason, although the application does not deny the existence of an optimal combination of price and income policies that would fulfil formalization objectives without welfare losses, the results presented here show that the space is narrow and the design of these policies needs to be extremely carefully undertaken.

Fourth, targeted income transferences or other policies that reinforce bargaining power in the household could be used to deal with formalization from a consumption side. As mentioned previously, as long as women control a greater portion of the family budget, the share of consumption allocated to informal markets will be smaller. The results suggest that these effects are stronger for food consumption (the most important group for Peruvian families). Therefore, if this evidence is interpreted as better-quality consumption, the estimations done here also provide an explanation of the recurrent positive effects found in the literature of targeted income transfers to the mother on better nutritional outcomes for the children. From the perspective followed in this application, this could be considered a positive side-effect of formalization policies.

Although the results encountered are intuitive and line with most of the stylized facts of informal consumption, the analysis presented here is not free from limitations, which they could be used to define a future research agenda to confirm some of the results found or to tease out additional results not investigated here.

First, on theoretical grounds, the assumption made here in order to facilitate the empirical investigation is to use a two-step demand model with two extreme possibilities. In the first one, the household first decides on its allocation across markets (the budget is distributed between formal and informal options) and then decides which goods to purchase. In the second one, they decide first which goods will be purchased and then where these will be bought. Both extreme solutions imply aggregation and weak separability assumptions that could be unsatisfactory on theoretical grounds. The consequences (mainly in terms of restrictions on the substitution possibilities between goods in different groups) generate the necessity to emphasize that more than exact solutions to the consumption problem, these models provide good approximations conditional on the assumptions made. The question as to which of these models is superior or whether there is a third (possibly intermediate or more general) solution have not been explored here and would probably make a good research question to be solved in the future in order to obtain adequate representations of preferences.

On theoretical grounds, too, the demand equations estimated in this thesis were conditional (on labour allocation), as originally proposed by Pollak (1969). As noted by Browning and Meghir (1991), although this specification facilitates empirical analysis of the quantity effects, it also leaves important research questions aside, such as the substitution possibilities (in terms of price effects) between labour and consumption. To produce a result like this, more-structural models are needed where the full demand profile (specifying leisure consumption and not assuming it as predetermined) is studied. This will certainly complicate the analysis, since most simplifying assumptions on labour allocation cannot be implemented, but will produce more general results for the effects of formal and informal labour allocation on market consumption. Along the same lines, another extension of the model could come from a deeper exploration of the transaction cost function that links labour supply and market consumption. This will require a closer analysis of such costs which, instead of being implicitly assumed in the model, need to be explicitly modelled. One way of advancing future research on this is shown in Gardes and Starzec (2009), who use an analysis of price differentials; however, a more formal way to do this will require specification of transactional cost functions in the demand setting.

Second, in empirical terms, there are also several improvements that could be implemented. For example, in terms of the estimation methods, it would be worthwhile exploring maximum likelihood alternatives in order to estimate the complete systems proposed here. Similarly, in the case of the price models, more research is needed on the correct way to impose theoretical restrictions in censored systems. In terms of the data issues, some additional research is also needed. In this application, the construction of formal and informal markets is based on a classification of the different outlets where the household acquires consumption. This classification has been based on observed characteristics of the outlet, experience in the particular context and anecdotal evidence obtained during the research. This would be improved by a more in-depth exploration of the outlets in the sample, basically in terms of their registration situation with the different authorities (central and municipal). This information was not available in the sample used here, but could easily either be incorporated into a joint survey that investigates the formality conditions of outlets or could increase the number of questions in the actual survey with a module that investigates the purchasing experience of the household. With this intermediate solution, it should be possible to check (at least indirectly) whether the transaction made was more or less formal in a way similar to how the same survey checks the formality conditions of the workplace.

The need for better data will also be important for the information on prices and the set of instruments used for identification purposes, in both IV and PROBIT models. In the first case, a joint survey focusing on the different outlets where the families purchase their goods could be used to also provide market prices for both formal and informal outlets and thus avoid the complication of using unit values. In the second case, the household survey could exploit the purchasing experience module proposed previously to inspect attitudinal factors behind formal and informal consumption. Issues like stigma, risk perception or tax morale, now absent from the survey, could improve identification strategies in the PROBIT models. The case of the IV is more complicated, but most of the recent literature on impact evaluation and randomization trials could provide additional paths to follow in this case.

Third, it is worth mentioning here the additional research questions that emerged during the writing of this thesis. For example, it would be useful to re-test the price effects for particular products instead of aggregating expenditure or consumption groups. This would help to place the hypothesis within a more homogenous demand context, ruling out the possible influence of quality differentials or remaining aggregation problems. It would also be interesting to confirm the results for income and hours worked in a different context – for example, comparing recession and expansion periods of the economy since they will define different strategies of households in terms of their participation in the formal and informal markets both as consumers and as workers. Along the same lines of thought, replicating the results in other highly informal countries in Latin America and Africa, where the market structure (on both the supply and the demand side) may be systematically different, would help to maintain the construction of certain empirical regularities around informal consumption.

X. REFERENCES

Akaichi, F. and C. Revoredo (2013) *Demand for Dairy Products in Malawi*. Aberdeen: Scotland's Rural College, Land Economy and Environment Research Group Working Paper.

Akerlof, G. (1970) 'The market for lemons: quality uncertainty and the market mechanism', *The Quarterly Journal of Economics*, 84(3): 488–500.

Alderman, H., P. Chiappori, L. Haddad, J. Hoddinott and R. Kanbur (1995) 'Unitary versus collective models of the household: is it time to shift the burden of proof?' *The World Bank Research Observer*, 10(1): 1–19.

Alviola, P., O. Capps and X. Wu (2010) *Micro-Demand Systems Analysis of Non-Alcoholic Beverages in the United States: An Application of Econometric Techniques Dealing With Censorship*. Tamu: Agribusiness, Food, and Consumer Economics Research Center, International Market Research Report No. CP–01–09.

Amaral, P. and E. Quintin (2006) 'A competitive model of the informal sector', *Journal of Monetary Economics*, 53(7): 1541–1553.

Amemiya, T. (1984) 'Tobit models: a survey', *Journal of Econometrics*, 24(1–2): 3–61.

Anbarci, N., P. Gomis-Porqueras and M. Pivato (2012) *Formal and Informal Markets: A Strategic and Evolutionary Perspective*. Munich: University of Munich, MPRA Paper No. 42513.

Andrews, D.W.K. and J. Stock (2005) *Inference with Weak Instruments*. New Haven, CT: Cowles Foundation, Discussion Paper No. 1530.

Andrews, D.W.K., M. Moreyra and J. Stock (2007) 'Performance of conditional Wald tests in IV regression with weak instruments', *Journal of Econometrics*, 139(1): 116–132.

Arellano, R. (2010) *Al Medio hay Sitio: Crecimiento Social Según los Estilos de Vida*. Lima: Editorial Planeta and Arellano Marketing.

Arellano, R. and D. Burgos (2010) *Cuidad de los Reyes, de los Quispe, de los Chávez: Una Visión Social y de Mercado de la Nueva Urbe Latinoamericana*. Lima: Editorial Planeta and Arellano Marketing.

Arias, J., O. Azuara, P. Bernal, J. Heckman and C. Villareal (2010) *Policies to Promote Growth and Economic Efficiency in Mexico*. Bonn: Institute for the Study of Labor, IZA Discussion Paper No. 4740.

Arias, O. and L. Lucchetti (2007) *Informal Employment and Self–Rated Welfare: Measuring Compensating Differentials*. Washington: World Bank, photocopy.

Bailón, J. and A. Nicoli (2009) *Chicha Power: El Marketing se Reinventa*. Lima: Lima University.

Balassa, B. (1964) 'The purchasing power parity doctrine: a reappraisal', *Journal of Political Economy*, 72(6): 584–596.

Banks, J., R. Blundell and A. Lewbel (1997) 'Quadratic Engel Curves and Consumer Demand', *The Review of Economics and Statistics*, 79(4): 527-539.

Barslund, M. (2011) *Estimation of Tobit Type Censored Demand Systems: A Comparison of Estimators*. Copenhagen: University of Copenhagen, Department of Economics Discussion Paper No. 07–16.

Basu, K. (2006) ‘Gender and say: a model of household behaviour with endogenously–determined balance of power’, *The Economic Journal*, 116(511): 558–580.

Baum, C.F., M.E. Schaffer and S. Stillman (2003) ‘Instrumental variables and GMM: estimation and testing’, *Stata Journal*, 3(1): 1–31.

Becker, G. (1981) *A Treatise on the Family*. Cambridge, MA and London: Harvard University Press.

Beuran, M. and E. Kalugina (2005) *Subjective Welfare and the Informal Sector: The Case of Russia*. Paris: Paris 1 University/Moscow: Higher School of Economics (photocopy).

Bielich, C. (2010) ‘¿Cómo influye el sistema laboral de las empresas de transporte público en la problemática del tránsito?’, *Economía y Sociedad*, 75: 13–19.

Blisard, N. and J. Blaylock (1993) ‘Distinguishing between market participation and infrequency of purchase models of butter demand’, *American Journal of Agricultural Economics*, 75(3): 698–705.

Böhme, M. and R. Thiele (2011) *Is the informal sector constrained from the demand side? Evidence for six West African capitals*, Kiel: Kiel Institute for the World Economy, Working Paper No. 1683.

Böhme, M. and R. Thiele (2012a) *Informal–Formal Linkages and Informal Enterprise Performance in Urban West Africa*. Kiel: Kiel Institute for the World Economy, Working Paper No. 1751.

Böhme, M. and R. Thiele (2012b) ‘Is the informal sector constrained from the demand side? Evidence for six West African capitals’, *World Development*, 40(7): 1369–1381.

Bourguignon, F., M. Browning, P. Chiappori and V. Lechene (1993) ‘Intra-household allocation of consumption: a model and some evidence from French data’, *Annales d’Economie et de Statistique*, 29: 137–156.

Boysen, O. (2012) *A Food Demand System Estimation for Uganda*. Dublin: Institute for International Integration Studies IIIS Discussion Paper No. 396.

Browning, M. and C. Meghir (1991) ‘The effects of male and female labour supply on commodity demands’, *Econometrica*, 59(4): 925–951.

Browning, M. and P. Chiappori (1998) ‘Efficient intra-household allocations: a general characterization and empirical tests’, *Econometrica*, 66(6): 1241–1278.

Browning, M., F. Bourguignon, P. Chiappori and V. Lechene (1994) ‘Income and outcomes: a structural model of intrahousehold allocation’, *Journal of Political Economy*, 102(6): 1067–1096.

Bruce, J. (1989) ‘Homes divided’, *World Development*, 17(7): 979–992.

Caillavet, F. (2005) ‘Poverty and healthy diet: education poor or income poor? A two-step model for fat consumption in France’. Reading: University of Reading, paper given to the Economics and Policy of Diet and Health, 97th EAAE Seminar, 21–22 April.

Calderon, U. (2010) Análisis de la Industria Farmacéutica Peruana – 2010. Lima: Pontificia Universidad Católica del Perú and Tulane University, Global MBA II thesis.

Cameron, A.C. and P.K. Trivedi (2005) *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Cappecchi, V. (1989) ‘The informal economy and the development of flexible specialization in Emilia Romagna’, in Portes, A., M. Castells and L.A. Benton (eds) *The Informal Economy: Studies in Advanced and Less Developed Countries*. Baltimore: The Johns Hopkins University Press, pp. 189–215.

Carpentier, A. and H. Guyomard (2001) ‘Unconditional elasticities in two-stage demand systems: an approximate solution’, *American Journal of Agricultural Economics*, 83(1): 222–229.

Cassel, D. (1984) ‘The growing shadow economy: implications for stabilization policy’, *Intereconomics*, 19(5): 219–225.

Castells, M. and A. Portes (1989) ‘World underneath: the origins, dynamics and effects of the informal economy’, in Portes A., M. Castells and L.A. Benton (eds) *The Informal Economy: Studies in Advanced and Less Developed Countries*. Baltimore: The Johns Hopkins University Press, pp. 11–40.

Centeno, M.A. and A. Portes (2006) ‘The informal economy in the shadow state’, in Fernandez–Kelly, P. and J. Sheffner (eds) *Out of the Shadows: Political Action and Informal Economy in Latin America*. University Park: Penn State University Press, pp. 23–48.

Cermeño, R. (1987) ‘Caída del ingreso real, recesión del sector moderno y expansión del sector informal: un enfoque microeconómico’, *Revista Economía*, 10(20): 73–99.

Chacaltana, J. and R. Infante (2014) *Hacia un Desarrollo Inclusivo: El Caso de Perú*. Santiago: Economic Commission for Latin America and the Caribbean (ECLAC)/Geneva: International Labour Organization (ILO).

Chalfant, J. (1987) ‘A globally flexible Almost Ideal Demand System’, *Journal of Business and Economic Statistics*, 5(2): 233–242.

Charlot, O., F. Malherbet and C. Terra (2013) ‘Informality in developing economies: regulation and fiscal policies’, Rede de Economia Aplicada (REAP), Working Paper 052.

Chen K. and C. Chen (2002) *Cross-Product Censoring in Demand System with Limited Dependent Variables: A Multivariate Probit Model Approach*. Alberta: University of Alberta, Department of Rural Economy, Staff Paper 00–02.

Chiappori, P. (1988) ‘Rational household labour supply’, *Econometrica*, 56(1): 63–90.

Chiappori, P. (1992) ‘Collective labour supply and welfare’, *Journal of Political Economy*, 100(3): 437–467.

Chiappori, P. (1997) ‘Collective models of household behaviour: the sharing rule approach’, in Haddad, L., J. Hoddinott and H. Alderman (eds) *Intrahousehold Resource Allocation in Developing Countries: Models, Methods, and Policy*. Baltimore: Johns Hopkins University Press for the International Food Policy Research Institute, pp. 39–52.

Chiarini, B. and E. Marzano (2006) ‘Market consumption and hidden consumption: a test for substitutability’, *Applied Economics*, 38(6): 707–716.

Cowell, F.A. (1985) ‘Tax evasion with labour income’, *Journal of Public Economics*, 26(1): 19–35.

Cox, T. and M. Wohlgenant (1986) ‘Prices and quality effects in cross-sectional demand analysis’, *American Journal of Agricultural Economics*, 68(4): 908–919.

Dasgupta, N. (1992) *Petty Trading in the Third World: The Case of Calcutta*. Avebury: Gower.

De Paula, A. and J.A. Scheinkman (2006) *The Informal Sector*. Cambridge Mass: National Bureau for Economic Research, NBER Working Paper No. 13486.

De Soto, H. (1986) *El Otro Sendero*. Peru: Editorial El Barranco.

Deaton, A. (1975) *Models and Projections of Demand in Post-War Britain*. London: Chapman & Hall.

Deaton, A. (1981) 'Theoretical and empirical approaches to consumer demand under rationing', in Deaton, A. (ed.) *Essays on the Theory and Measurement of Consumer Behaviour in Honour of Sir Richard Stone*. New York: Cambridge University Press, pp. 55–72.

Deaton, A. (1986) 'Demand analysis', in Griliches, Z. and M.D. Intriligator (eds) *Handbook of Econometrics Vol. 3*. Amsterdam: North-Holland, pp. 1767–1839.

Deaton, A. (1988) 'Quality, quantity and spatial variation of price', *American Economic Review*, 78(3): 418–430.

Deaton, A. (1990) 'Price elasticities from survey data: extensions and Indonesian results', *Journal of Econometrics*, 44(3): 281–309.

Deaton, A. (1997) *The Analysis of Household Surveys: A Microeconometric Approach tTo Development Policy*. Baltimore: The Johns Hopkins University Press for The World Bank.

Deaton, A. and J. Muellbauer (1980a) 'An almost ideal demand system', *The American Economic Review*, 70(3): 312–326.

Deaton, A. and J. Muellbauer (1980b) *Economics and Consumer Behaviour*. New York: Cambridge University Press.

Díaz, J. (2014) 'Formalización empresarial y laboral', in Chacalatana, J. and R. Infante (eds) *Hacia un Desarrollo Inclusivo: El Caso de Perú*. Santiago: Economic Commission for Latin America and the Caribbean (ECLAC); Geneva: International Labour Organization (ILO).

Dong, D., B.W. Gould and H.M. Kaiser (2004) 'Food demand in Mexico: an application of the Amemiya–Tobin approach to the estimation of a censored food system', *American Journal of Agricultural Economics*, 86(4): 1094–1107.

Doss, C. (2011) *Intrahousehold Bargaining and Resource Allocation in Developing Countries: Background Paper to the World Development Report 2012 (Gender Equality and Development)*. Washington DC: The World Bank.

Duflo, E. and C. Udry (2004) *Intrahousehold Resource Allocation in Côte d'Ivoire: Social Norms, Separate Accounts and Consumption Choices*. Cambridge Mass: National Bureau for Economic Research, NBER Working Paper No. 10498.

Eisend, M. and P. Schuchert-Güler (2006) 'Explaining counterfeit purchases: a review and a preview', *Academy of Marketing Science Review*, 2006(12): 1–22.

Espinoza, V. (1992) Networks of Informal Economy: Work and Community among Santiago's Urban Poor. Toronto: University of Toronto, Graduate Department of Sociology, unpublished PhD thesis.

Farrell, D. (2004) 'The hidden dangers of the informal economy', *The McKinsey Quarterly*, 3: 121–133.

Feige, E. (1990) 'Defining and estimating underground informal economies: the new institutional economics approach', *World Development*, 18(7): 989–1002.

Ferrer-i-Carbonell, A. and K. Gërxhani (2004) *Subjective Poverty and the (In)Formal Sector in a Transition Country*. Amsterdam: Amsterdam Institute for Advanced Labour Studies (AIAS).

Fields, G.S. (1990) 'Labour market modelling and the urban informal sector: theory and evidence', in Turnham, D., B. Salomé and A. Schwarz (eds) *The Informal Sector Revisited*. Paris: Development Centre of the Organization for Economic Co-Operation and Development, pp. 49–69.

Fields, G.S. (2007) *Modelling Labour Market Policy in Developing Countries: A Selective Review of the Literature and Needs for the Future*. Washington DC: The World Bank, Policy Research Working Paper No. 4362.

Finlay, K. and L.M. Manguson (2009) 'Implementing weak-instrument-robust tests for a general class of instrumental-variables models', *The Stata Journal*, 9(3): 398–421.

Finlay, K., L.M. Magnusson and M.E. Schaffer (2013) *WEAKIV: Stata Module to Perform Weak-Instrument-Robust Tests and Confidence Intervals for Instrumental-Variable (IV) Estimation of Linear, Probit and Tobit Models*. Boston: Boston College, Department of Economics.

Flores-Lagunes, A. (2007) 'Finite sample evidence of iv estimators under weak instruments', *Journal of Applied Econometrics*, 22(3): 677–694.

Fomba, B. and C. Mvolo (2010) 'Taxation of goods and services of the informal sector in Cameroun'. Oxford: University of Oxford, St Catherine's College, paper presented at the CSAE Conference 2010: Economic Development in Africa, 21-23 March.

Fortin, B., G. Lacroix and C. Montmarquette (2000) 'Are underground workers more likely to be underground consumers?', *The Economic Journal*, 110(466): 838–860.

Fortin, B., N. Marceau and L. Savard (1997) 'Taxation, wage controls and the informal sector', *Journal of Public Economics*, 66(2): 293–312.

Friedman, J. and J. Levinsohn (2002) 'The distributional impacts of Indonesia's financial crisis on household welfare: a rapid response methodology', *World Bank Economic Review*, 16(3): 397–423.

Fukuchi, T. (1998) 'A simulation analysis of the urban informal sector', *The Developing Economies*, 36(3): 225–256.

Galiani, S. and F. Weinschelbaum (2007) *Modelling Informality Formally: Households and Firms*. La Plata: Universidad Nacional de La Plata, CEDLAS Working Paper No. 0047.

Garavito, C. (1997) *Balance y Agenda de la Investigación Económica sobre Género en el Sector Urbano Peruano*. Lima: Consorcio de Investigación Económica y Social.

García, M. (1990) Resource Allocation and Household Welfare: A Study of Personal Sources of Income on Food Consumption, Nutrition and Health in the Philippines. The Hague: Institute of Social Studies, unpublished Ph.D. thesis.

Gardes, F. and C. Starzec (2002) 'Polish households between transition and informal markets' (mimeo). Université Paris I–Pantheon–Sorbonne, CNRS–TEAM, INSEE

Gardes, F. and C. Starzec (2009) *Polish Households' Behaviour in the Regular and Informal Economies*. Paris: Centre d'Economie de la Sorbonne, CES Working Paper No. 2009–17.

Gasparini, L. (2001) *Microeconometric Decomposition of Aggregate Variables. An Application to Labor Informality in Argentina*. Buenos Aires: Fundación de Investigaciones Económicas Latinoamericanas (FIEL), Working Paper No. 68.

Ghersi, E. (1997) ‘The informal economy in Latin America’, *The Cato Journal*, 17(1): 99–108.

Gould, B.W. (1992) ‘At-home consumption of cheese: a purchase infrequency model’, American *Journal of Agricultural Economics*, 74(2): 444–452.

Gould, W., J. Pitblado and W. Sribney (2006) *Maximun Likelihood Estimation with Stata*. College Station Tx: StataCorp LP, Stata Press (3rd edition).

Granovetter, M. (1985) ‘Economic action and social structure: the problem of embeddedness’, *American Journal of Sociology*, 91(3): 481–510.

Gravelle, H. and R. Rees (2004) *Microeconomics*. New Jersey: Prentice Hall (3rd edition).

Green, R. and J. Alston (1990) ‘Elasticities in AIDS models’, *American Journal of Agricultural Economics*, 72(2): 442–445.

Greene, W. (2003) *Econometric Analysis*. New Jersey: Prentice Hall (5th edition).

Guarnizo, L. (1992) One Century in Two: Dominican-Owned Firms in New York and the Dominican Republic. Baltimore: The Johns Hopkins University, Department of Sociology, Ph.D thesis.

Guyer, J. (1980) *Household Budgets and Women’s Incomes*. Boston: University of Boston, African Studies Center, Working Paper No. 28.

Guyer, J. and P. Peters (1987) ‘Introduction’, *Development and Change*, 18(2): 197–224.

Haddad, L., J. Hoddinott and H. Alderman (1997) ‘Introduction: the scope of intrahousehold resource allocation issues’, in Haddad, L., J. Hoddinott and H. Alderman (eds) *Intrahousehold Resource Allocation in Developing Countries: Models, Methods, and Policy*. Baltimore: John Hopkins University Press/London: International Food Policy Research Institute, pp. 1–16.

Harris, J.R. and M. Todaro (1970) ‘Migration, unemployment and development: a two-sector analysis’, *American Economic Review*, 60(1): 126–142.

Hart, K. (1970) ‘Small-scale entrepreneurs in Ghana and development planning’, *Journal of Development Studies*, 10(4): 104–120.

Hayashi, F. (1995) *Is the Japanese Extended Family Altruistically Linked? A Test Based on Engel Curves*. Cambridge MA: National Bureau of Economic Research Working Paper No. 5033.

Hayashi, F. (2000) *Econometrics*. Princeton NJ: Princeton University Press.

Heckman, J. (1979) ‘Sample selection bias as a specification error’, *Econometrica*, 47(1): 153–161.

Herrera, J. (2003) *La Pobreza en el Perú*. Lima: Instituto Nacional de Estadística e Informática. IRD–INEI (photocopy).

Hoddinott, J. and L. Haddad (1995) ‘Does female income share influence household expenditures? Evidence from Côte d’Ivoire’, *Oxford Bulletin of Economics and Statistics*, 57(1): 77–96.

Houthakker, H.S. (1952–1953) ‘Compensated changes in quantities and qualities consumed’, *The Review of Economic Studies*, 19(3): 155–164.

Huber, P. (1967) ‘The behaviour of maximum likelihood estimates under nonstandard conditions’, *Proceedings of the 5th Berkeley Symposium in Mathematical Statistics and Probability*, 1: 221–233.

Iguiñiz, J. (1996) 'Definiciones de desarrollo y experiencias de género: apuntes desde la perspectiva de Sen', in Henriquez, N. (ed.) *Encrucijadas del Saber: Los Estudios de Género en las Ciencias Sociales*. Lima: Pontificia Universidad Católica del Perú, Facultad de Ciencias Sociales, Programa de Estudios de Género, pp. 157-188.

ILO (1972) *Employment, Incomes and Equality: A Strategy for Increasing Productive Employment in Kenya*. Geneva: International Labour Organization.

INEI (1992) *Perú: Estadísiticas del Sector Informal*. Lima: Instituto Nacional de Estadísitica e Informatica.

INEI (2001) *Metodología de Cálculo del Indice de Precios al Consumidor de Lima Metropolitana*. Lima: Instituto Nacional de Estadística e Informática.

INEI (2008) *Indicadores de Precios de la Economía: Boletín Mensual de Febrero de 2008*. Lima: Instituto Nacional de Estadística e Informática.

Instituto Cuanto (2007) *Perú en Números*. Lima: Instituto Cuanto, Anuario Estadístico del Instituto Cuanto.

Jorgenson, D.W., L.J. Lau and T.M. Stocker (1982) 'The transcendental logarithmic model of aggregate consumer behaviour', in Basmann, R. and G. Rhodes (eds) *Advances in Econometrics, Vol. 1*. Greenwich, Ct: JAI Press, pp. 97–238.

Kannappan, S. (1985) 'Urban employment and the labor market in developing nations', *Economic Development and Cultural Change*, 33(4): 699–730.

Kesselman, J. (1989) 'Income tax evasion: an intersectoral analysis', *Journal of Public Economics*, 38(2): 137–182.

Kleibergen, F. (2002) 'Pivotal statistics for testing structural parameters in instrumental variables regression', *Econometrica*, 70(5): 1781–1803.

Kleibergen, F. and R. Paap (2006) 'Generalized reduced rank tests using the singular value decomposition', *Journal of Econometrics*, 133(1): 97–126.

Krstic, G. (1998) *Analysis of the Hidden Economy in FR Yugoslavia with Estimates for 1997 and Recommendations for its Legalization*. Belgrade: Economics Institute of Belgrade, Economics Discussion Paper No. 504.

La Porta, R. and A. Shleifer (2011) *The Unofficial Economy in Africa*. Cambridge Mass: National Bureau for Economic Research, NBER Working Paper No. 16821.

Lacroix, G. and B. Fortin (1992) 'Utility-based estimation of labour supply functions in the regular and irregular sectors', *The Economic Journal*, 102(415): 1047–1422.

Lamb, C. (1989) 'Zero expenditure and corner solutions', in Pudney, S. (ed.) *Modelling Individual Choice*. Oxford: Basil Blackwell, pp. 138–185.

Lancaster, G., P. Maitra and R. Ray (2003) *Endogenous Power, Household Expenditure Patterns and New Tests of Gender Bias: Evidence from India*. Hobart: University of Tasmania/Melbourne: Monash University, photocopy.

Lancaster, G., P. Maitra and R. Ray (2006) 'Endogenous intra-household balance of power and its impact on expenditure patterns: evidence from India', *Economica*, 73(291): 435–460.

Lazaridis, P. (2003) 'A microeconometric analysis of food quality choice by Greek households', *Agricoltura Mediterranea*, 133(2): 149–156.

Lehmann, H. and N. Pignatti (2007) *Informal Employment Relationships and Labor Market Segmentation in Transition Economies: Evidence from Ukraine*. Bonn: Institute for the Study of Labor, IZA Discussion Paper No. 3269.

Lemieux, T., B. Fortin and P. Fréchette (1994) 'The effect of taxes on labor supply in the underground economy', *The American Economic Review*, 84(1): 231–254.

Leser, C.E.V. (1963) 'Forms of Engel functions', *Econometrica*, 31(4): 694–703.

Levenson, A. and W. Maloney (1998) *The Informal Sector, Firm Dynamics and Institutional Participation*. Washington DC: The World Bank, Policy Research Working Paper No. 1988.

Livingstone, I. (1991) 'A reassessment of Kenya's rural and urban informal sector', *World Development*, 19(6): 651–670.

Loayza, N. (1997) *The Economics of the Informal Sector: A Simple Model and some Empirical Evidence from Latin America*. Washington DC: The World Bank, Policy Research Working Paper No. 1727.

Loayza, N. (2007) *The Causes and Consequences of Informality in Peru*. Lima: Banco Central de Reserva de Perú, Working Paper No. 2007–018.

Lucas, R. (1978) 'On the size and distribution of business firms', *Bell Journal of Economics*, 9(2): 508–523.

Lundberg, S. and R. Pollak (1993) 'Separate spheres bargaining and the marriage market', *The Journal of Political Economy*, 101(6): 988–1010.

Lundberg, S. and R. Pollak (1996) 'Bargaining and distribution in marriage', *Journal of Economic Perspective*, 10(4): 139–158.

Macroconsult (2010) *Bases del Plan Nacional de Regionalización*. Lima: Informe de Consultoría elaborado para la Presidencia del Consejo de Ministros, photocopy.

Magnac, T. (1991) 'Segmented or competitive labor markets', *Econometrica*, 59(1): 165–187.

Magnusson, L.M. (2010) 'Inference in limited dependent variable models robust to weak identification', *Econometrics Journal*, 13(3): S56–S79.

Maloney, W.F. (1999) 'Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico', *The World Bank Economic Review*, 13(2): 275–302.

Maloney, W.F. (2004) 'Informality revisited', *World Development*, 32(7): 1159–1178.

Manser, M. and M. Brown (1980) 'Marriage and household decision-making: a bargaining analysis', *International Economic Review*, 21(1): 31–44.

McDonald, J.F. and R.A. Moffitt (1980) 'The uses of Tobit analysis', *Review of Economics and Statistics*, 62(2): 318–321.

McElroy, M. (1997) 'The Policy implications of family bargaining and marriage markets', in Haddad, L., J. Hoddinott and H. Alderman (eds) *Intrahousehold Resource Allocation in Developing Countries: Models, Methods, and Policy*. Washington DC: International Food Policy Research Institute/Baltimore: Johns Hopkins University Press, pp. 53–74.

McElroy, M. and M. Horney (1981) 'Nash-bargained household decisions: toward a generalization of the theory of demand', *International Economic Review*, 22(2): 333–349.

McGee, T., K. Salih and M.L. Young (1989) ‘Industrial development, ethnic cleavages and employment patterns: Penang State, Malaysia’, in Portes, A., M. Castells and L.A. Benton (eds) *The Informal Economy: Studies in Advanced and Less Developed Countries*. Baltimore: The Johns Hopkins University Press, pp. 265–278.

Mencher, J. (1988) ‘Women's work and poverty: women's contribution to household maintenance in south India’, in Dwyer, D. and J. Bruce (eds) *A Home Divided*. Redwood City: Stanford University Press, pp. 99–119.

Mikusheva, A. and B. Poi (2006) ‘Tests and confidence sets with correct size when instruments are potentially weak’, *The Stata Journal*, 6(3): 335–347.

Monge, A. (2007) Intra-Household Bargaining and Consumption Decisions: A Case Study Applied to Peru. Brighton: University of Sussex, unpublished M.A. dissertation in Development Economics.

Moreira, M. (2003) ‘A conditional likelihood ratio test for structural models’, *Econometrica*, 71(4): 1027–1048.

Moschini, G. (1995) ‘Units of measurement and the Stone Index in demand system estimation’, *American Journal of Agricultural Economics*, 76(1): 63–68.

Muellbauer, J. (1975) ‘Aggregation, Income distribution and consumer demand’, *The Review of Economic Studies*, 42(4): 525–543.

Muellbauer, J. (1976) ‘Community preferences and the representative consumer’, *Econometrica*, 44(5): 979–999.

Muellbauer, J. (1981) ‘Testing neoclassical models of the demand of consumers’, in Deaton, A. (ed.) *Essays on the Theory and Measurement of Consumer Behaviour in Honour of Sir Richard Stone*. New York: Cambridge University Press, pp. 213–235.

Murphy, K.M. and R.H. Topel (1985) ‘Estimation and inference in two-step econometric models’, *Journal of Business and Economic Statistics*, 3(4): 370–379.

Nelson, C. and R. Startz (1990) ‘Some further results on the exact small sample properties of the instrumental variable estimator’, *Econometrica*, 58(4): 967–976.

Nelson, P. (1970) ‘Information and consumer behavior’, *Journal of Political Economy*, 78(2): 311–329.

Niimi, Y. (2007) Household Welfare and Price Changes in Vietnam in an Era of Trade Liberalization. Brighton: University of Sussex, unpublished D.Phil. thesis in Economics.

Ortner, S. (1974) ‘Is female to male as nature is to culture?’, in Rosaldo, M. and L. Lamphere (eds) *Woman, Culture and Society.* Redwood City: Stanford University Press, pp. 68–87.

Pasco-Font, A. and J. Saavedra (2001) *Reformas Estructurales y Bienestar: Una Mirada al Perú de los Noventa*. Lima: Grupo de Análisis para el Desarrollo.

Peace, G., D. Galleta and J. Thong (2003) ‘Software piracy in the workplace: a model and empirical test’, *Journal of Management Information Systems*, 20(1): 153–177.

Perry, G., O. Arias, H. López, W. Maloney and L. Servén (2006) *Poverty Reduction and Growth: Virtuous and Vicious Circles*. Washington DC: The World Bank.

Perry, G., W. Maloney, O. Arias, P. Fajnzylver, A. Mason and J. Saavedra-Chanduví (2007) *Informality: Exit and Exclusion*. Washington DC: The World Bank.

Phipps, S. and P. Burton (1998) ‘What’s mine is yours? The influence of male and female incomes on patterns of household expenditure’, *Economica*, 65(260): 599–613.

Phumpiu, P. and L. Triveño (2007) *The Puzzles of One-Night-Stand Economies*. Lima: Universidad Católica del Perú, photocopy.

Pisani, M. (2013a) *Consumption, Informal Markets, and the Underground Economy: Hispanic Consumption in South Texas*. New York: Palgrave Macmillan.

Pisani, M. (2013b) ‘Cross-border consumption on informal and underground goods: a case study of alternative consumerism in South Texas’, *Social Science Quarterly*, 4(1): 242–262.

Pollak, R. (1969) ‘Conditional demand functions and consumption theory’, *The Quarterly Journal of Economics*, 83(1): 60–78.

Pollak, R. (1971) ‘Conditional demand functions and the implications of separable utility’, *Southern Economic Journal*, 37(4): 423–433.

Portes, A. (2000) ‘La economía informal y sus paradojas’, in Carpio, J., E. Klein and I. Novacovsky (eds) *Informalidad y Exclusión Social*. Lima: Fondo de Cultura Económica, Secretaría de Desarrollo Social de la Nación and Organización Internacional del Trabajo (OIT), pp. 25–49.

Pradhan, M. and A. van Soest (1997) ‘Household labor supply in urban areas of Bolivia’, *The Review of Economics and Statistics*, 79(2): 300–310.

Quisumbing, A. and B. de la Brière (2000) *Women's Assets and Intra-Household Allocation in Rural Bangladesh: Testing Measures of Bargaining Power*. Washington DC: International Food Policy Research Institute, Food Consumption and Nutrition Division Discussion Paper No. 86.

Quisumbing, A. and J. Maluccio (2000) *Intra-Household Allocation and Gender Relations: New Empirical Evidence from Four Developing Countries*. Washington DC: International Food Policy Research Institute, Food Consumption and Nutrition Division Discussion Paper No. 84.

Rauch, J. (1991) ‘Modelling the informal sector formally’, *Journal of Development Economics*, 35(1): 33–47.

Regalado, O., C. Fuentes, G. Aguirre, N. García, R. Miu and R. Vallejo (2009) *Factores Críticos de Exito en los Centros Comerciales de Lima Metropolitana y el Callao*. Lima: Universidad ESAN, Serie Gerencia Global No. 13.

Reilly, B., G. Krstić and M. Cominetta (2006) ‘Purchasing from informal markets: evidence from Serbia’, *Applied Economics Quarterly*, 52(3): 239–264.

Restrepo-Echevarría, P. (2011) *Macroeconomic Volatility: The Role of the Informal Economy*. Athens: University of Ohio, photocopy.

Riley, N. (1997) ‘Gender, power and population change’, *Population Bulletin*, 52(1): 1–46.

Rogers, C.A. and K. Swinnerton (2004) *A Model of Informal Sector Labor Markets*. Washington DC: U.S. Department of Labor, photocopy.

Rosaldo, M. (1974) ‘Woman, culture and society: a theoretical overview’, in Rosaldo, M. and L. Lamphere (eds) *Woman, Culture and Society*. Redwood: Stanford University Press, pp. 17–42.

Rosenzweig, M.R. (1988) ‘Labor markets in low-income countries’, in Chenery, H. and T.N. Srinivasan (eds) *Handbook of Development Economics Vol. 1*. Amsterdam: North-Holland, pp. 713–762.

Samuelson, P. (1964) ‘Theoretical notes on trade problems’, *Review of Economics and Statistics,* 46(2): 145–154.

Sanchis, E. (1982) ‘Economía subterránea y descentralización productiva en la industria manufacturera’, *Boletín de Estudios Económicos,* 37(117): 461–481.

Schneider, F. and A. Buehn (2007) ‘Shadow economies and corruption all over the world: revised estimates for 120 countries’, *Economics,* 1(2007–9): 1–53.

Schneider, F. and D. Enste (2000) ‘Shadow economies: size, causes and consequences’, *Journal of Economic Literature,* 30(1): 77–114.

Sen, A. (1985). “Women, Technology and Sexual Divisions” Trade and Development, United Nations, New York (6).

Sethuraman, J.V. (1981) *The Urban Informal Sector in Developing Countries*. Geneva: International Labour Organization.

Shonkwiler, J. and T. Yen (1999) ‘A two-step estimation of a censored system of equations’, *American Journal of Agricultural Economics,* 81(4): 972–982.

Short, P. (1996) ‘Kinship, reciprocity and vulnerability: social relations in the informal economy’, *Australian Journal of Social Issues,* 31(2): 127–145.

Stock, J.H. and M. Yogo (2005) ‘Testing for weak instruments in linear IV regression’, in Andrews, K. and J.H. Stock (eds) *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press, pp. 80–108.

Stock, J.H., J.H. Wright and M. Yogo (2002) ‘A survey of weak instruments and weak identification in generalized method of moments’, *Journal of Business and Economic Statistics,* 20(4): 518–529.

Stone, J.R.N. (1953) *The Measurement of Consumers’ Expenditure Behaviour in the United Kingdom 1920–1938, Vol.1*. Cambridge: Cambridge University Press.

Straub, S. (2005) *Informal Sector: The Credit Market Channel*. Edinburgh: University of Edinburgh, School of Economics.

Su, S.J.B. and S.T. Yen (2000) ‘A censored system of cigarette and alcohol consumption’, *Applied Economics,* 32(6): 729–737.

SUNAT (2011) *Estimación del Contrabando y Defraudación de Rentas de Aduanas en el Perú Durante el 2010*. Lima: Oficina de Estadística de la Intendencia Nacional de Estudios Tributarios y Planeamiento, Superintendencia Nacional de Administración Tributaria.

Swaminathan, M. (1991) *Understanding the Informal Sector: A Survey*. Helsinki: United Nations University, WIDER Working Paper No. 94.

Tafere, K., A. Seyoum, S. Tamru, N. Tafera and Z. Paulos (2010) *Food Demand Elasticities in Ethiopia: Estimates Using Household Income Consumption Expenditure (HICE) Survey Data*. Washington DC: International Food Policy Research Institute, ESSP II Working Paper 11.

Theil, H. (1952–1953) ‘Qualities, prices and budget enquiries’, *The Review of Economic Studies,* 19(3): 129–147.

Theil, H. (1971) ‘An economic theory of the second moments of disturbance of behavioural equations’, *American Economic Review,* 61(1): 190–194.

Theil, H. (1976) *Theory and Measurement of Consumer Demand, Vol. II*. Amsterdam: North-Holland.

Thomas, D. (1997) 'Incomes, expenditures and health outcomes: evidence on intahousehold resource allocation', in Haddad, L., J. Hoddinott and H. Alderman (eds) *Intrahousehold Resource Allocation in Developing Countries: Models, Methods, and Policy*. Baltimore: International Food Policy Research Institute and Johns Hopkins University Press, pp.142–164.

Thomas, D. and C.L. Chen (1994) *Income Shares and Shares of Income: Empirical Tests of Models of Household Resource Allocations*. Santa Monica: RAND Corporation, Labor and Population Program Working Paper Series 94–08.

Thomas, J.J. (1992) *Informal Economic Activity*. Hertfordshire: Harvester Wheatsheaf.

Tobin, J. (1958) 'Estimation of relationships for limited dependant variables', *Econometrica*, 26(1): 24–36.

Tokman, V. (1978) 'Competition between the informal and formal sectors in retailing: the case of Santiago', *World Development*, 6(9–10): 1187–1198.

Tokman, V. (2000) 'El sector informal posreforma económica', in Carpio, J., E. Klein and I. Novacovsky (eds) *Informalidad y Exclusión Social*. Mexico: Fondo de Cultura Económica, pp. 65–74.

Ulph, D. (1988) *A General Non-Cooperative Nash Model of Household Consumption Behavior*. Bristol: University of Bristol, Discussion Paper No. 88/205.

Vassilopoulos, A., S. Klonaris, A. Drichoutis and P. Lazardis (2009) 'Quality demand for meat and fish products: a microeconometric analysis'. Valetta: Paper presented at the 5th edition of the International Conference on Applied Business and Research, 21-25 September.

von Braun, J. (1988) 'Effects of technological change in agriculture on food consumption and nutrition: rice in a West African setting', *World Development*, 16(9): 1083–1098.

Wan, G. (1996) 'Income elasticities of household demand in rural China: estimates from cross-sectional survey data', *Journal of Economic Studies*, 23(3): 18–33.

Weesie, J. (1999) 'Seemingly unrelated estimation and the cluster-adjusted sandwich estimator'. College Station, Tx: Stata, Technical Bulletin No. 52: 34–47.

White, H. (1980) 'A heteroscedasticity-consistent covariance matrix estimator and a direct test of heteroscedasticity', *Econometrica*, 48(4): 817–838.

Williams, C. (2002) 'Why do people use alternative retain channels? Some case-study evidence from two English cities', *Urban Studies*, 39(10): 1897–1910.

Williams, C. (2006) 'How much for cash? Tackling the cash-in-hand ethos in the household services sector', *Services Industries Journal*, 26(5): 479–492.

Williams, C. (2008) 'Consumers' motives for buying goods and services on an off-the-books basis', *The International Review of Retail, Distribution and Consumer Research*, 18(4): 405–421.

Williams, C. and C. Paddock (2003) 'The meaning of alternative consumption practices', *Cities*, 20(5): 311–319.

Williams, C. and J. Windebank (2002) 'The excluded consumer: a neglected aspect of social exclusion', *Policy & Politics*, 30(4): 501–513.

Williams, C. and J. Windebank (2005) ‘Why do households use alternative consumption practices? Some lessons from Leicester’, *Community, Work and Family,* 8(3): 301–320.

Williamson, J. (1988) ‘Migration and urbanization’, in Chenery, H. and T.N. Srinivasan (eds) *Handbook of Development Economics, Vol. 1*. Amsterdam: North-Holland, pp. 425–465.

Wooldridge, J. (2002) *Econometric Analysis of Cross Section and Panel Data*. London: MIT Press.

Working, H. (1943) ‘Statistical laws of family expenditure’, *Journal of the American Statistical Association*, 38(221): 43–56.

Xu, Z. (2007) ‘A survey on intra–household models and evidence’. Munich: Munich Personal RePEc Archive, Paper No. 3763.

Yamada, G. (1996) ‘Urban informal employment and self–employment in developing countries: theory and evidence’, *Economic Development and Cultural Change*, 44(2): 289–314.

Yen, S. and B.-H. Lin (2006) ‘A sample selection approach to censored demand systems’, *American Journal of Agricultural Economics*, 88(3): 742–749.

Yen, S. and K. Chen (2002) *Potato Consumption in Canada: Is It Becoming a Normal Good?* Edmonton: University of Alberta, Department of Rural Economy, Staff Paper No. 02–05.

Yen, S., B.-H. Lin and D. Smallwood (2003) ‘Quasi- and simulated-likelihood approaches to censored demand systems: food consumption by food stamp recipients in the United States’, *American Journal of Agricultural Economics*, 85(2): 458–478.

XI. APPENDICES

List of Appendices

Appendix 1: Mathematical appendix

1. Expressions (III.32a), (V.1), (VI.1), (VI.4) and (VII.1a)

Using the most general theoretical demand (III.3) and a standard AIDS formulation, the demand function takes the form:

$$w_\rho = \alpha_\rho + \sum_\omega \gamma_{\rho\omega} \log[p_\omega(1+a_\omega)] + \beta_\rho \log\left(\frac{M}{P^*}\right) + \sum_r \delta_{\rho r} h_r + \psi_\rho(\lambda) + e_\rho$$

$$w_\rho = \alpha_\rho + \sum_\omega \gamma_{\rho\omega} \log(p_\omega) + \sum_\omega \gamma_{\rho\omega} \log(1+a_\omega) + \beta_\rho \log\left(\frac{M}{P^*}\right) + \sum_r \delta_{\rho r} h_r + \psi_\rho(\lambda) + e_\rho$$

Then if $\log(1+a_\omega) \cong \sum_r \chi_{\omega r} h_r$ (any log-linear functional form is assumed just to explore how the theory works in empirical terms):

$$w_\rho = \alpha_\rho + \sum_\omega \gamma_{\rho\omega} \log(p_\omega) + \sum_\omega \gamma_{\rho\omega} (\sum_r \chi_{\omega r} h_r) + \beta_\rho \log\left(\frac{M}{P^*}\right) + \sum_r \delta_{\rho r} h_r + \psi_\rho(\lambda) + e_\rho$$

$$w_\rho = \alpha_\rho + \sum_\omega \gamma_{\rho\omega} \log(p_\omega) + \beta_\rho \log\left(\frac{M}{P^*}\right) + \sum_r h_r[\delta_{\rho r} + \sum_\omega \gamma_{\rho\omega} \chi_{\omega r}] + \psi_\rho(\lambda) + e_\rho$$

$$w_\rho = \alpha_\rho + \sum_\omega \gamma_{\rho\omega} \log(p_\omega) + \beta_\rho \log\left(\frac{M}{P^*}\right) + \sum_r \tau_{\rho r} h_r + \psi_\rho(\lambda) + e_\rho$$

which is expression (III.36a) with $\tau_{\rho r} = \delta_{\rho r} + \sum_{\rho\omega} \gamma_{\rho\omega} \chi_{\omega r}$ for the non-separability (first term on the right) and network effects (second term on the right). Once prices are set to unity, Engel curves can be obtained:

$$w_\rho = \alpha^*_\rho + \beta_\rho \log(M) + \sum_r \tau_{\rho r} h_r + \psi_\rho(\lambda) + e_\rho$$

which is expression (V.1). A similar procedure can be followed to derive expressions (VI.1) and (VI.4), but only changing the notation for each step and considering in (VI.1) that $\chi_{gr} = 0$ and in

(VI.4) $\rho_{jr} = 0$ to be coherent with the theoretical model. Finally, expression (VII.1a) is the expression previous to the price normalization.

2. Expressions (V.2), (V1.2), (VI.5) and (VII.2)

Focusing on (V.2), as the rest of expression will follow the same procedure, we can consider the Working–Leser function (V.1):

$$w_j = \alpha^*_j + \beta_j \log(M) + \sum_r \tau_{jr} h_r + \psi_j(\lambda) + e_j$$

and take the partial derivative of w_j with respect to $\log(M)$. Then,

$$\frac{\partial w_j}{\partial \log(M)} = \beta_j$$

Note that the left-hand expression can be also expressed as

$$\frac{\partial w_j}{\partial \log(M)} \cong \frac{d\left(\frac{p_j c_j}{M}\right)}{d\log(M)} = \frac{p_j\left(\frac{Mdc_j - c_j dM}{M^2}\right)}{\frac{dM}{M}} = p_j \frac{dc_j}{dM} - w_j$$

so it is possible to write

$$p_j \frac{dc_j}{dM} - w_j = \beta_j$$

$$\frac{dc_j}{dM} = \frac{\beta_j + w_j}{p_j}$$

which, in elasticity form, is

$$\eta_j = \frac{dc_j}{dM}\frac{M}{c_j} = \frac{\beta_j + w_j}{p_j}\frac{M}{c_j} = \frac{\beta_j}{w_j} + 1$$

3. Expressions (V.3), (VI.3) and (VI.6)

Focusing on (V.3), as the rest of the expression will follow the same procedure, we can consider the Working–Leser function (V.1):

$$w_j = \alpha_j^* + \beta_j \log(M) + \sum_r \tau_{jr} h_r + \psi_j(\lambda) + e_j$$

and take the partial derivative with respect to h_r; as a result it follows that

$$\frac{\partial w_j}{\partial h_r} = \tau_{jr}$$

The right-hand expression can be written

$$\frac{\partial w_j}{\partial h_r} \cong \frac{d\left(\frac{p_j c_j}{M}\right)}{dh_r} = \frac{p_j}{M}\frac{dc_j}{dh_r}$$

so, in general, it follows that

$$\frac{p_j}{M}\frac{dc_j}{dh_r} = \tau_{kjr}$$

$$\frac{dc_j}{dh_r} = \tau_{jr}\frac{M}{p_j}$$

and in elasticity form

$$\mu_{jr} = \frac{dc_j}{dh_r}\frac{h_r}{c_j} = \tau_{jr}\frac{M}{p_j}\frac{h_r}{c_j} = \tau_{jr}\left(\frac{h_r}{w_j}\right)$$

4. Expression (V.4)

For the Working–Leser function (V.1):

$$w_j = \alpha^*_j + \beta_j \log(M) + \sum_r \tau_{jr} h_r + \psi_j(\lambda) + e_j$$

and, taking the partial derivative with respect to h_r as in the previous case, but allowing M to change as a result, it follows that

$$\frac{\partial w_j}{\partial h_r} = \tau_{jr} + \beta_j \frac{\partial \log(M)}{\partial h_r}$$

solving

$$\frac{\partial \log(M)}{\partial h_r} = \frac{\partial M}{M \partial h_r} = \frac{s_r}{M}$$

As before, the right-hand expression can be

$$\frac{\partial w_j}{\partial h_r} \cong \frac{d\left(\frac{p_j c_j}{M}\right)}{dh_r} = \frac{M p_j dc_j - p_j c_j dM}{M^2 dh_r} = \frac{p_j dc_j}{M dh_r} - \frac{p_j c_j dM}{M^2 dh_r} = \frac{p_j}{M}\left(\frac{dc_j}{dh_r} - \frac{c_j dM}{M dh_r}\right)$$

so,

$$\frac{p_j}{M}\left(\frac{dc_j}{dh_r} - \frac{c_j dM}{M dh_r}\right) = \tau_{jr} + \beta_j \frac{s_r}{M}$$

$$\frac{dc_j}{dh_r} = \left(\tau_{jr} + \beta_j \frac{s_r}{M}\right)\frac{M}{p_j} + \frac{c_j}{M} s_r$$

which, in elasticity form, becomes:

$$\mu^T_{jr} = \frac{dc_j}{dh_r}\frac{h_r}{c_j} = \tau_{jr}\frac{h_r}{w_j} + \frac{h_r s_r}{M}\left(\frac{\beta_j}{w_j} + 1\right)$$

5. <u>Expression (VI.7)</u>

Using (III.28), it was established that

$$\frac{\partial c_{kj}^{U}}{\partial M} = \frac{\partial c_{kj}}{\partial m_k}\frac{\partial C_k}{\partial M}P_k$$

Then, from demonstration (2) it is possible to show that

$$\frac{dc_{kj}}{dm_k} = \frac{\beta_{kj} + w_{kj}}{p_{kj}}$$

$$\frac{dC_k}{dM} = \frac{\beta_k + w_k}{P_k}$$

so

$$\frac{\partial c_{kj}^{U}}{\partial M} = \frac{\beta_{kj} + w_{kj}}{p_{kj}}\frac{\beta_k + w_k}{P_k}P_k$$

and, in elasticity form

$$\eta_{kj}^{U} = \frac{\partial c_{kj}^{U}}{\partial M}\frac{M}{c_{kj}} = (\beta_{kj} + w_{kj})(\beta_k + w_k)\frac{M}{p_{kj}c_{kj}} = \left(\frac{\beta_{kj}}{w_{kj}} + 1\right)\left(\frac{\beta_k}{w_k} + 1\right)$$

6. <u>Expression (VI.8)</u>

For expression (III.27)

$$\frac{\partial c_{kj}^{U}}{\partial h_r} = \frac{\partial c_{kj}}{\partial h_r} + \frac{\partial c_{kj}}{\partial m_k}P_k\left(\frac{\partial C_k}{\partial h_r}\right)$$

which in elasticity terms is

$$\mu_{kjr}^{U} = \frac{\partial c_{kj}^{U}}{\partial h_r}\frac{h_r}{c_{kj}} = \frac{\partial c_{kj}}{\partial h_r}\frac{h_r}{c_{kj}} + \frac{\partial c_{kj}}{\partial m_k}P_k\left(\frac{\partial C_k}{\partial h_r}\right)\frac{h_r}{c_{kj}}$$

$$\mu_{kjr}^{U} = \mu_{kjr}^{C} + \frac{\partial c_{kj}}{\partial m_k}P_k\left(\frac{\partial C_k}{\partial h_r}\right)\frac{h_r}{c_{kj}}$$

$$\mu_{kjr}^{U} = \mu_{kjr}^{C} + \eta_{kj}^{C}\frac{\partial C_k}{\partial h_r}\frac{h_r}{C_k}$$

where μ_{kjr}^{C} follows from demonstration (3) and η_{kj}^{C} from demonstration (2). Then to identify $\frac{\partial C_k}{\partial h_r}\frac{h_r}{C_k}$ it is possible to use (VI.1)

$$w_k = \alpha_k^{*} + \beta_k \log(M) + \sum_r \tau_{kr} h_r + \psi_k(Z) + e_k$$

taking partial derivatives with respect to h_r

$$\frac{\partial w_k}{\partial h_r} = \tau_{kr}$$

and considering that

$$\frac{\partial w_k}{\partial h_r} \cong \frac{d\left(\frac{P_k C_k}{m}\right)}{dh_r} = \frac{P_k}{m}\frac{dC_k}{dh_r}$$

so

$$\frac{P_k}{m}\frac{dC_k}{dh_r} = \tau_{kr}$$

$$\frac{dC_k}{dh_r} = \tau_{kr}\frac{m}{P_k}$$

$$\frac{dC_k}{dh_r}\frac{h_r}{C_k} = \tau_{kr}\frac{m}{P_k}\frac{h_r}{C_k}$$

$$\frac{dC_k}{dh_r}\frac{h_r}{C_k} = \tau_{kr}\left(\frac{h_r}{w_k}\right)$$

Therefore,

$$\mu_{kjr}^{U} = \mu_{kjr}^{C} + \eta_{kj}^{C}\tau_{kr}\left(\frac{h_r}{w_k}\right)$$

7. <u>Expression (VI.9)</u>

Using expression (III.32)

$$\frac{\partial c_{kj}^{UT}}{\partial h_r} = \frac{\partial c_{kj}}{\partial h_r} + \frac{\partial c_{kj}}{\partial m_k}P_k\left(\frac{\partial C_k}{\partial h_r} + s_r\frac{\partial C_k}{\partial M}\right)$$

and expressed in elasticity terms

$$\mu_{kjr}^{UT} = \frac{\partial c_{kj}^{UT}}{\partial h_r}\frac{h_r}{c_{kj}} = \frac{\partial c_{kj}}{\partial h_r}\frac{h_r}{c_{kj}} + \frac{\partial c_{kj}}{\partial m_k}P_k\left(\frac{\partial C_k}{\partial h_r} + s_r\frac{\partial C_k}{\partial M}\right)\frac{h_r}{c_{kj}}$$

$$\mu_{kjr}^{UT} = \mu_{kjr}^{C} + \left(\frac{\partial c_{kj}}{\partial m_k}P_k\frac{\partial C_k}{\partial h_r}\right)\frac{h_r}{c_{kj}} + s_r\frac{\partial C_k}{\partial M}\frac{h_r}{c_{kj}}$$

$$\mu_{kjr}^{UT} = \mu_{kjr}^{C} + \eta_{kj}^{C}\theta_{kr}\left(\frac{h_r}{w_k}\right) + s_r\frac{\partial C_k}{\partial M}\frac{h_r}{c_{kj}}$$

$$\mu_{kjr}^{UT} = \mu_{kjr}^{U} + s_r\frac{\partial C_k}{\partial M}\frac{h_r}{c_{kj}}$$

$$\mu_{kjr}^{UT} = \mu_{kjr}^{U} + \eta_{kj}^{C}\eta_k\left(\frac{s_r h_r}{M}\right)$$

which is simply

$$\mu_{kjr}^{UT} = \mu_{kjr}^{U} + \eta_{kj}^{U}\left(\frac{s_r h_r}{M}\right)$$

8. Expressions (VI.16a) and (VI.16b)

Using expression (V1.11)

$$w = \Phi(\pi_0 Q_1)G(\pi_1 Q_2) + \pi_2 \phi(\pi_0 Q_1) + u$$

and take partial derivatives of w with respect to x, considering that $\pi_0 Q_1$ is a linear expression, $G(\pi_1 Q_2) = \pi_1 Q_2$ is also linear in parameters, Q_1 contains x with parameter $\pi_{0,x}$ and Q_2 contains x with parameter $\pi_{1,x}$. Then

$$\frac{\partial w}{\partial x} = \Phi(\pi_0 Q_1)\frac{\partial G(\pi_1 Q_2)}{\partial x} + G(\pi_1 Q_2)\frac{\partial \Phi(\pi_0 Q_1)}{\partial x} + \pi_2 \frac{\partial \phi(\pi_0 Q_1)}{\partial x}$$

$$\frac{\partial w}{\partial x} = \Phi \pi_{1,x} + G(\pi_1 Q_2)\phi\pi_{0,x} - \pi_2 \pi_0 Q_1 \phi\pi_{0,x}$$

$$\frac{\partial w}{\partial x} = \Phi \pi_{1,x} + \phi\pi_{0,x}[G(\pi_1 Q_2) - \pi_2(\pi_0 Q_1)]$$

or

$$\frac{\partial w}{\partial x} = \pi_{1,x}\Theta = \pi_{1,x}\left[\Phi + \phi(\pi_{0,x}/\pi_{1,x})[G(\pi_1 Q_2) - \pi_2(\pi_0 Q_1)]\right]$$

9. Expression (VII.3)

$$\varepsilon_{jg}^{O} = \frac{\partial c_j^O}{\partial p_g}\frac{p_g}{c_j} = -\delta_{jg} + \frac{\gamma_{jg}}{\hat{w}_j} - \beta_j\left(\frac{\hat{w}_g}{\hat{w}_j}\right)$$

Consider the formulation of the AIDS model (VII.1a) – (VII.1b)

$$w_j = \alpha_j + \sum_g \gamma_{jg} \log(p_g) + \beta_j \log\left(\frac{m}{P*}\right) + \tau_j(A) + e_j$$

where

$$\log P* = \sum_g \bar{w}_g \log(p_g)$$

and take the partial derivative of w_j with respect to $\log(p_g)$. Then,

$$\frac{\partial w_j}{\partial \log(p_g)} = \gamma_{jg} - \beta_j \frac{\partial \log(P*)}{\partial \log(p_g)}$$

From (34b), the last term of the right-hand expression is w_g, so

$$\frac{\partial w_j}{\partial \log(p_g)} = \gamma_{jg} - \beta_j w_g$$

and the left-hand side can be expressed as

$$\frac{\partial w_j}{\partial \log(p_g)} \cong \frac{d\left(\frac{p_j c_j}{m_k}\right)}{d \log(p_g)}$$

Therefore when $g \neq j$

$$\frac{\partial w_j}{\partial \log(p_g)} \cong \frac{d\left(\frac{p_j c_j}{m}\right)}{d \log(p_g)} = \frac{p_g p_j}{m} \frac{dc_j}{dp_g}$$

and

$$\frac{p_g p_j}{m} \frac{dc_j}{dp_g} = \gamma_{jg} - \beta_j w_g$$

$$\frac{dc_j}{dp_g} = (\gamma_{jg} - \beta_j w_g)\frac{m}{p_g p_j}$$

which, in elasticity form, is equal to

$$\frac{dc_j}{dp_g}\frac{p_g}{c_j} = (\gamma_{jg} - \beta_j w_g)\frac{m}{p_g p_j}\frac{p_g}{c_j} = \frac{\gamma_{jg}}{w_j} - \beta_j\left(\frac{w_g}{w_j}\right)$$

However, when $g = j$

$$\frac{\partial w_j}{\partial \log(p_j)} \cong \frac{d\left(\frac{p_j c_j}{m}\right)}{d\log(p_j)} = \frac{p_j(p_j dc_j + c_j dp_j)}{mdp_j} = \frac{p_j p_j}{m}\frac{dc_j}{dp_j} + w_j$$

and

$$\frac{p_j p_j}{m}\frac{dc_j}{dp_j} + w_j = \gamma_{jj} - \beta_j w_j$$

$$\frac{dc_j}{dp_g} = (\gamma_{jj} - \beta_j w_j - w_j)\frac{m}{p_j p_j}$$

which, in elasticity form, is equal to

$$\frac{dc_j}{dp_j}\frac{p_j}{c_j} = (\gamma_{jj} - \beta_j w_j - w_j)\frac{m}{p_j p_j}\frac{p_j}{c_j} = -1 + \frac{\gamma_{jj}}{w_j} - \beta_j$$

Considering that these derivatives are computed from an ordinary demand and using the definition of the Kroneker delta $\delta_{jg} = 1$ when $j = g$ and $\delta_{jg} = 0$ when $g \neq j$, both cases (when $g \neq j$ and $g = j$) can be adequately summarized in a unique expression by:

$$\varepsilon_{jg}^{O} = \frac{\partial c_j^O}{\partial p_g}\frac{p_g}{c_j} = -\delta_{jg} + \frac{\gamma_{jg}}{w_j} - \beta_j\left(\frac{w_g}{w_j}\right)$$

10. <u>Expression (VIII.9)</u>

Begin with equation (VIII.6)

$$w = \alpha_b + (\alpha_a - \alpha_b)\theta + \beta_a \theta \ln(\theta M) + \beta_b (1-\theta) \ln\left[(1-\theta)M\right] + \beta_1 z + \beta_2 \lambda + \varepsilon$$

and take partial derivatives with respect to $\ln(M)$

$$\frac{\partial w}{\partial \ln(M)} = \beta_a \theta + \beta_b (1-\theta)$$

and note that

$$\frac{\partial w}{\partial \log(M)} \cong \frac{d\left(\frac{PC}{M}\right)}{d \log(M)} = \frac{P\left(\frac{MdC - cdM}{M^2}\right)}{\frac{dM}{M}} = P\frac{dC}{dM} - w$$

then

$$\frac{dC}{dM} = \frac{\beta_a \theta + \beta_b (1-\theta) + w}{P}$$

which in elasticity form becomes

$$\frac{\partial C}{\partial M}\frac{M}{C} = \eta = \frac{\beta_a \theta + \beta_b (1-\theta)}{w} + 1$$

Appendix 2: Definition of variables and basic tabulations (models in Chapter 5)

Variable	Definition	SUR–GLS		IV	
		8,007 obs.		4,874 obs.	
		Mean	s.d.	Mean	s.d.
Dependant variables					
Informal	Informal consumption as a share of total expenditure (purchased), three-market definition (see Chapter 4)	0.393	0.17873	0.390	0.18351
Semi-Formal	Semi-formal consumption as a share of total expenditure (purchased), three-market definition (see Chapter 4)	0.316	0.18033	0.317	0.18356
Formal	Formal consumption as a share of total expenditure (purchased), three-market definition (see Chapter 4)	0.291	0.17045	0.293	0.17630
Market_1	Consumption done in Market 1 as a share of total expenditure (purchased), five-market definition (see Chapter 4)	0.177	0.13145		
Market_2	Consumption done in Market 2 as a share of total expenditure (purchased), five-market definition (see Chapter 4)	0.216	0.16717		
Market_3	Consumption done in Market 3 as a share of total expenditure (purchased), five-market definition (see Chapter 4)	0.316	0.18033		
Market_4	Consumption done in Market 4 as a share of total expenditure (purchased), five-market definition (see Chapter 4)	0.099	0.10590		
Market_5	Consumption done in Market 5 as a share of total expenditure (purchased), five-market definition (see Chapter 4)	0.192	0.13738		
Independent variables					
L(expenditure)	Natural logarithm of total *per-capita* expenditure (purchased) in annual new *soles*	8.091	0.72461	8.152	0.73187
L(fam_size)	Natural logarithm of the number of family members	1.295	0.57086	1.187	0.60389
D(mem_spo)	Dummy variable: 1 = household has a spouse, 0 = otherwise	0.645	0.47849	0.585	0.49275
D(mem_son)	Dummy variable: 1 = household has a son or daughter, 0 = otherwise	0.793	0.40496	0.731	0.44326
D(mem_oth)	Dummy variable: 1 = household has other member, 0 = otherwise	0.311	0.46313	0.300	0.45811
P(men)	Proportion of men in the household	0.503	0.24510	0.509	0.26633
age_head	Age of the head of the household (in years)	49.200	14.52858	49.310	15.67740
age_head-sq	Age of the head of the household (in years) squared	2631.725	1532.74500	2677.188	1655.47900
D(gender_head)	Dummy variable: 1 = household head in male, 0 = otherwise	0.739	0.43908	0.716	0.45106
year_educ_head	Number of years of education of the head of the household	10.299	4.84059	10.460	4.83813

Hrs_for	Sum of the weekly hours of work of all family members, primary activity, sector = extended definition of formality (see Chapter 4)			26.533	37.20317
Hrs_inf	Sum of the weekly hours of work of all family members, primary activity, sector = extended definition of informality (see Chapter 4)			46.379	53.49863
Hrs_formal-Hrs_informal	Difference in hours worked between the formal and informal sectors			-19.846	73.33665
Hrs_ff	Sum of the weekly hours of work of all family members, primary activity, sector = formal workers in formal firms (see Chapter 4)	21.163	34.15949	21.874	34.77438
Hrs_if	Sum of the weekly hours of work of all family members, primary activity, sector = informal workers in formal firms (see Chapter 4)	5.546	18.75729	5.901	19.67698
Hrs_ii	Sum of the weekly hours of work of all family members, primary activity, sector = informal workers in informal firms (see Chapter 4)	18.628	35.99963	19.087	37.25073
Hrs_sf	Sum of the weekly hours of work of all family members, primary activity, sector = informal self-employment (see Chapter 4)	7.995	22.11749	4.660	17.03040
Hrs_si	Sum of the weekly hours of work of all family members, primary activity, sector = formal self-employment (see Chapter 4)	28.448	40.49600	21.391	35.60842
Hrs_fam	Sum of the weekly hours of work of all family members, primary activity, sector = family work (see Chapter 4)	8.040	22.05925		
Hrs_second	Sum of the weekly hours of work of all family members, secondary activity (see Chapter 4)	6.294	14.30344		
D(Lima)	Dummy variable: 1 = household lives in Lima, 0 = otherwise	0.168	0.37409	0.202	0.40130
L(pop_urb_dist)	Natural logarithm of the population size of the district where the household lives	11.081	1.28319	11.187	1.23384
L(pop_urb_dist)-sq	Natural logarithm of the population size of the district where the household lives squared	124.429	27.60059	126.676	26.87404
pop_den_distx1000	Population density: inhabitants per 1,000 km2	2.946	5.36708	3.351	5.66370
D(not_slum)	Dummy variable: 1 = household doesn´t live in a slum, 0 = otherwise	0.470	0.49911	0.488	0.49990
D(mid_city)	Dummy variable: 1 = household lives in a departmental capital (except Lima), 0 = otherwise	0.509	0.49996	0.504	0.50003
D(border)	Dummy variable: 1 = household lives in a border city (Tumbes, Tacna or Puno), 0 = otherwise	0.113	0.31619	0.102	0.30237
D(tongue)	Dummy variable: 1 = any of the two parents of the household has indigenous mother tongue , 0 = otherwise	0.242	0.42827	0.220	0.41439
D(migrant)	Dummy variable: 1 = any of the two parents of the household members is a migrant, 0 = otherwise	0.645	0.47868	0.649	0.47722
D(social)	Dummy variable: 1 = any of the household members pertain to social organizations (unions, social clubs, social programs, etc.) , 0 = otherwise	0.403	0.49051	0.372	0.48349
Station_distx1000	Number of police stations in the district per 1,000 inhabitants	0.059	0.10642	0.058	0.11427
Muni_persx1000	Number of workers in the local government of the district per 1,000 inhabitants	3.594	4.74743	3.732	5.56388
L(budget_dist_pc)	Natural logarithm of the *per-capita* budget of the local government, in thousands of new *soles*	5.117	0.66009	5.127	0.68336

D(sdp_dist)	Dummy variable: 1 = the district has a slum development plan, 0 = otherwise	0.051	0.21915	0.045	0.20763
D(Hrs_for)	Dummy variable: 1 = any household member is formal worker using the extended definition of formality (see Chapter 4), 0 = otherwise			0.422	0.49387
D(Hrs_inf)	Dummy variable: 1 = any household member is informal worker using the extended definition of informality (see Chapter 4), 0 = otherwise			0.633	0.48194
D(Hrs_ff)	Dummy variable: 1 = any household member is formal worker in formal firm in his/her primary activity (see Chapter 4), 0 = otherwise	0.347	0.47591	0.352	0.47754
D(Hrs_if)	Dummy variable: 1 = any household member is informal worker in formal firm in his/her primary activity (see Chapter 4), 0 = otherwise	0.105	0.30662	0.106	0.30823
D(Hrs_ii)	Dummy variable: 1 = any household member is informal worker in informal firm in his/her primary activity (see Chapter 4), 0 = otherwise	0.307	0.46127	0.300	0.45811
D(Hrs_sf)	Dummy variable: 1 = any household member is formal self-employer in his/her primary activity (see Chapter 4), 0 = otherwise	0.145	0.35200	0.086	0.28095
D(Hrs_si)	Dummy variable: 1 = any household member is informal self-employer in his/her primary activity (see Chapter 4), 0 = otherwise	0.490	0.49992	0.390	0.48776
D(Hrs_fam)	Dummy variable: 1 = any household member is family worker in his/her primary activity (see Chapter 4), 0 = otherwise	0.199	0.39924		
D(Hrs_second)	Dummy variable: 1 = any household member has a second job (see Chapter 4), 0 = otherwise	0.274	0.44623		
Instruments					
L(asset)	Natural logarithm of the value of assets owned by the household			7.154	1.53504
L(asset)-sq	Natural logarithm of the value of assets owned by the household squared			53.529	21.34006
D(risk1)	Dummy variable 1 = if a member of the family has lost his/her job in the last year, 0 = otherwise			0.045	0.20808
D(risk2)	Dummy variable 1 = if the family has experienced bankruptcy of the family business in the last year, 0 = otherwise			0.009	0.09459
D(risk5)	Dummy variable 1 = if the head of the house has abandoned the family in the last year, 0 = otherwise			0.010	0.10177
D(risk7)	Dummy variable 1 = if the family has experienced robbery or other crime against it in the last year, 0 = otherwise			0.030	0.17161
D(risk8)	Dummy variable 1 = if the family has experienced a natural disaster in the last year, 0 = otherwise			0.011	0.10275
L(wage_for)	Natural logarithm of salaries in the formal sector (district mean)			1.665	0.40636
L(wage_inf)	Natural logarithm of salaries in the informal sector (district mean)			1.047	0.37472
D(house)	Dummy variable 1 = if the family house shows inadequate living conditions (INEI, methodology), 0 = otherwise			0.067	0.24985
non_labour	Non-labour income (income from property rents and extraordinary income)			841.330	3272.11600
wage_diff	Difference in the logarithm of wages (formal–informal)			0.618	0.45861

Appendix 3: Full regression results (Chapter 5)

a) SUR–GLS (three-market definition)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expenditure)	-0.02679***	0.08424***	-0.03744***	0.08121***	-0.02526***	0.08400***	-0.02597***	0.08291***
	(0.00413)	*(0.00384)*	*(0.00425)*	*(0.00386)*	*(0.00415)*	*(0.00389)*	*(0.00412)*	*(0.00382)*
Hrs_ff	-0.00027***	0.00022***	-0.00027***	0.00022***	-0.00027***	0.00022**	-0.00028***	0.00022***
	(0.00009)	*(0.00008)*	*(0.00009)*	*(0.00008)*	*(0.00009)*	*(0.00008)*	*(0.00009)*	*(0.00008)*
Hrs_if	0.00017	0.00011	0.00013	0.00013	0.00017	0.00010	0.00016	0.00012
	(0.00019)	*(0.00015)*	*(0.00019)*	*(0.00015)*	*(0.00020)*	*(0.00015)*	*(0.00019)*	*(0.00015)*
Hrs_ii	-0.00012	-0.00032***	-0.00013	-0.00028***	-0.00013	-0.00032***	-0.00012	-0.00032***
	(0.00008)	*(0.00006)*	*(0.00008)*	*(0.00006)*	*(0.00008)*	*(0.00006)*	*(0.00008)*	*(0.00006)*
Hrs_sf	0.00006	0.00054***	-0.00007	0.00056***	0.00002	0.00054***	0.00009	0.00053***
	(0.00017)	*(0.00018)*	*(0.00017)*	*(0.00018)*	*(0.00017)*	*(0.00018)*	*(0.00017)*	*(0.00018)*
Hrs_si	0.00029***	0.00006	0.00023***	0.00007	0.00027***	0.00006	0.00028***	0.00006
	(0.00007)	*(0.00006)*	*(0.00007)*	*(0.00006)*	*(0.00007)*	*(0.00006)*	*(0.00007)*	*(0.00006)*
Hrs_fam	0.00002	0.00009	0.00004	0.00009	-0.00001	0.00009	0.00002	0.00008
	(0.00012)	*(0.00010)*	*(0.00012)*	*(0.00010)*	*(0.00012)*	*(0.00010)*	*(0.00012)*	*(0.00010)*
Hrs_second	0.00022	0.00004	0.00017	0.00005	0.00024	0.00004	0.00023	0.00003
	(0.00019)	*(0.00015)*	*(0.00019)*	*(0.00015)*	*(0.00019)*	*(0.00015)*	*(0.00019)*	*(0.00015)*
D(Hrs_ff)	0.01063	-0.00110	0.01395**	-0.00065	0.00893	-0.00057	0.01174*	-0.00249
	(0.00684)	*(0.00629)*	*(0.00672)*	*(0.00622)*	*(0.00683)*	*(0.00630)*	*(0.00683)*	*(0.00626)*
D(Hrs_if)	-0.01265	-0.01452	-0.01211	-0.01483	-0.01214	-0.01433	-0.01315	-0.01468
	(0.01158)	*(0.00930)*	*(0.01143)*	*(0.00923)*	*(0.01167)*	*(0.00931)*	*(0.01159)*	*(0.00931)*
D(Hrs_ii)	0.00449	-0.00970*	0.00401	-0.01051**	0.00379	-0.00987*	0.00419	-0.00914*
	(0.00650)	*(0.00520)*	*(0.00643)*	*(0.00516)*	*(0.00650)*	*(0.00521)*	*(0.00647)*	*(0.00519)*
D(Hrs_sf)	-0.02336**	-0.00404	-0.00658	-0.00657	-0.02294**	-0.00378	-0.02578**	-0.00320
	(0.01138)	*(0.01084)*	*(0.01096)*	*(0.01074)*	*(0.01140)*	*(0.01085)*	*(0.01137)*	*(0.01080)*
D(Hrs_si)	0.01158**	-0.00846*	0.01020*	-0.00773*	0.01163**	-0.00858*	0.01066*	-0.00733
	(0.00563)	*(0.00472)*	*(0.00555)*	*(0.00469)*	*(0.00561)*	*(0.00471)*	*(0.00561)*	*(0.00471)*
D(Hrs_fam)	-0.00595	0.00387	-0.00319	0.00361	-0.00658	0.00395	-0.00577	0.00388
	(0.00717)	*(0.00576)*	*(0.00702)*	*(0.00575)*	*(0.00715)*	*(0.00576)*	*(0.00717)*	*(0.00576)*
D(Hrs_second)	-0.01374**	0.01084**	-0.01154**	0.01121**	-0.01345**	0.01112**	-0.01387**	0.01138**
	(0.00596)	*(0.00493)*	*(0.00588)*	*(0.00491)*	*(0.00593)*	*(0.00494)*	*(0.00595)*	*(0.00493)*
L(fam_size)	0.00833	0.00849	0.00357	0.00720	0.00857	0.00886	0.00850	0.00954
	(0.00745)	*(0.00633)*	*(0.00741)*	*(0.00628)*	*(0.00741)*	*(0.00633)*	*(0.00743)*	*(0.00633)*
D(mem_spo)	0.03871***	-0.03319***	0.03880***	-0.03125***	0.03757***	-0.03308***	0.03814***	-0.03234***
	(0.00733)	*(0.00634)*	*(0.00721)*	*(0.00628)*	*(0.00735)*	*(0.00635)*	*(0.00731)*	*(0.00632)*
D(mem_son)	0.01038	-0.01074	0.00695	-0.00985	0.01165	-0.01077	0.00950	-0.01027
	(0.00783)	*(0.00671)*	*(0.00770)*	*(0.00665)*	*(0.00781)*	*(0.00671)*	*(0.00781)*	*(0.00670)*
D(mem_oth)	0.00876*	0.00819*	0.01009*	0.00809*	0.01078**	0.00788*	0.00894*	0.00828*
	(0.00531)	*(0.00426)*	*(0.00523)*	*(0.00423)*	*(0.00530)*	*(0.00426)*	*(0.00530)*	*(0.00426)*
P(men)	-0.02414**	0.00993	-0.02076**	0.01055	-0.02346**	0.01003	-0.02286**	0.00897
	(0.00962)	*(0.00847)*	*(0.00947)*	*(0.00843)*	*(0.00960)*	*(0.00847)*	*(0.00961)*	*(0.00846)*
age_head	0.00177**	0.00371***	0.00194**	0.00370***	0.00158*	0.00377***	0.00198**	0.00360***
	(0.00088)	*(0.00078)*	*(0.00086)*	*(0.00077)*	*(0.00088)*	*(0.00078)*	*(0.00088)*	*(0.00078)*
age_head-sq	-0.00002*	-0.00002**	-0.00002**	-0.00002***	-0.00001	-0.00002**	-0.00002**	-0.00002**
	(0.00001)	*(0.00001)*	*(0.00001)*	*(0.00001)*	*(0.00001)*	*(0.00001)*	*(0.00001)*	*(0.00001)*
D(gender_head)	-0.01548**	0.01333**	-0.01395*	0.01430**	-0.01740**	0.01372**	-0.01471*	0.01217*
	(0.00772)	*(0.00655)*	*(0.00757)*	*(0.00651)*	*(0.00771)*	*(0.00656)*	*(0.00770)*	*(0.00655)*
year_educ_head	-0.00375***	0.00569***	-0.00380***	0.00497***	-0.00320***	0.00569***	-0.00375***	0.00571***
	(0.00053)	*(0.00044)*	*(0.00053)*	*(0.00045)*	*(0.00054)*	*(0.00045)*	*(0.00053)*	*(0.00044)*
D(Lima)	0.07501***	0.02212***	0.06152***	0.02208***	0.07249***	0.02184***	0.06864***	0.02401***
	(0.00516)	*(0.00458)*	*(0.00946)*	*(0.00794)*	*(0.00521)*	*(0.00464)*	*(0.00527)*	*(0.00466)*

L(pop_urb_dist)			0.13562***	0.00021				
			(0.02253)	*(0.01966)*				
L(pop_urb_dist)-sq			-0.00534***	-0.00014				
			(0.00105)	*(0.00092)*				
pop_den_distx1000			-0.00068	0.00071				
			(0.00049)	*(0.00043)*				
D(not_slum)			-0.00348	0.03371***				
			(0.00401)	*(0.00357)*				
D(mid_city)			-0.01707***	0.00407				
			(0.00466)	*(0.00396)*				
D(border)			0.05811***	-0.00079				
			(0.00598)	*(0.00527)*				
D(tongue)					0.02744***	-0.00186		
					(0.00473)	*(0.00403)*		
D(migrant)					0.01238***	-0.00157		
					(0.00414)	*(0.00354)*		
D(social)					-0.00134	-0.00432		
					(0.00398)	*(0.00334)*		
Station_distx1000							-0.09539***	0.02647*
							(0.01875)	*(0.01502)*
Muni_persx1000							-0.00108***	0.00241***
							(0.00039)	*(0.00053)*
L(budget_dist_pc)							0.00192	-0.00133
							(0.00338)	*(0.00286)*
D(sdp_dist)							-0.02618***	-0.01483**
							(0.00832)	*(0.00730)*
Constant	0.56248***	-0.58452***	-0.18142	-0.55361***	0.53677***	-0.58185***	0.55327***	-0.57542***
	(0.03870)	*(0.03486)*	*(0.12129)*	*(0.10958)*	*(0.03883)*	*(0.03526)*	*(0.04208)*	*(0.03744)*
Observations	8,007		8,007		8,007		8,007	
Pseudo LogL	7926.212		8154.336		7955.881		7973.908	

b) SUR–GLS (quintiles)

	Quintile 1		Quintile 2		Quintile 3		Quintile 4		Quintile 5	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expenditure)	0.04337***	0.00736	-0.00206	0.08551***	0.00728	0.15719***	-0.07818**	0.10435***	-0.10473***	0.10236***
	(0.01359)	(0.01348)	(0.04019)	(0.03100)	(0.04296)	(0.03855)	(0.03413)	(0.03196)	(0.00956)	(0.01196)
L(fam_size)	0.03730**	-0.00438	0.01311	-0.02046	-0.00238	0.03884***	-0.00825	0.02158	-0.01699	0.02417
	(0.01616)	(0.01189)	(0.01651)	(0.01343)	(0.01635)	(0.01462)	(0.01581)	(0.01400)	(0.01528)	(0.01770)
D(mem_spo)	0.00786	-0.02693*	0.04761***	-0.03650**	0.03122**	-0.02944**	0.04783***	-0.04258***	0.04467***	-0.01399
	(0.02082)	(0.01567)	(0.01751)	(0.01420)	(0.01584)	(0.01441)	(0.01435)	(0.01313)	(0.01356)	(0.01581)
D(mem_son)	0.00083	-0.01682	0.00709	0.00217	0.01160	-0.02448	0.00653	-0.00175	-0.00158	0.00090
	(0.02205)	(0.01630)	(0.01988)	(0.01503)	(0.01732)	(0.01564)	(0.01660)	(0.01449)	(0.01351)	(0.01543)
D(mem_oth)	0.01386	0.00185	0.00974	0.01001	0.01149	0.00812	0.01041	0.00944	0.01139	0.00010
	(0.01263)	(0.00831)	(0.01171)	(0.00892)	(0.01147)	(0.00944)	(0.01066)	(0.00970)	(0.01024)	(0.01191)
P(men)	-0.03451	-0.01668	0.00022	0.00198	-0.05145**	0.01512	-0.01889	0.00563	-0.02978*	0.05148***
	(0.02672)	(0.01964)	(0.02302)	(0.02062)	(0.02150)	(0.01909)	(0.01987)	(0.01798)	(0.01726)	(0.01953)
age_head	0.00511**	0.00236	0.00096	0.00423**	0.00003	0.00345**	0.00056	0.00602***	0.00298*	0.00145
	(0.00224)	(0.00171)	(0.00216)	(0.00178)	(0.00189)	(0.00166)	(0.00185)	(0.00170)	(0.00157)	(0.00180)
age_head-sq	-0.00005**	-0.00001	-0.00001	-0.00003	0.00000	-0.00002	-0.00000	-0.00004***	-0.00002	0.00000
	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
D(gender_head)	-0.00036	0.01445	-0.03055*	0.01995	-0.00777	0.00719	-0.02011	0.01988	-0.00877	-0.00953
	(0.02147)	(0.01557)	(0.01788)	(0.01395)	(0.01625)	(0.01487)	(0.01550)	(0.01360)	(0.01444)	(0.01670)
year_educ_head	-0.00369***	0.00566***	-0.00236*	0.00433***	-0.00322***	0.00397***	-0.00443***	0.00515***	-0.00218**	0.00717***
	(0.00129)	(0.00090)	(0.00124)	(0.00100)	(0.00108)	(0.00087)	(0.00109)	(0.00096)	(0.00107)	(0.00128)
Hrs_ff	-0.00031	0.00014	-0.00046*	0.00051*	0.00008	-0.00010	-0.00028*	0.00012	-0.00015	0.00012

	(0.00036)	(0.00028)	(0.00027)	(0.00027)	(0.00019)	(0.00017)	(0.00017)	(0.00016)	(0.00014)	(0.00015)
Hrs_if	0.00019	0.00016	0.00039	0.00043	-0.00011	0.00008	0.00027	-0.00026	-0.00003	0.00050
	(0.00040)	(0.00026)	(0.00051)	(0.00043)	(0.00033)	(0.00023)	(0.00047)	(0.00036)	(0.00043)	(0.00036)
Hrs_ii	-0.00029	-0.00011	-0.00035**	-0.00012	-0.00015	-0.00022*	-0.00005	-0.00059***	0.00016	-0.00089***
	(0.00018)	(0.00011)	(0.00017)	(0.00011)	(0.00016)	(0.00013)	(0.00018)	(0.00017)	(0.00028)	(0.00032)
Hrs_sf	0.00013	0.00158***	0.00083*	0.00025	-0.00030	0.00055	-0.00043	0.00099***	-0.00004	-0.00005
	(0.00063)	(0.00054)	(0.00049)	(0.00042)	(0.00033)	(0.00034)	(0.00034)	(0.00036)	(0.00027)	(0.00038)
Hrs_si	0.00024*	0.00032***	-0.00001	0.00033***	0.00029**	-0.00003	0.00033*	-0.00025	0.00063***	-0.00027
	(0.00014)	(0.00011)	(0.00014)	(0.00011)	(0.00013)	(0.00011)	(0.00017)	(0.00016)	(0.00023)	(0.00022)
Hrs_fam	-0.00015	0.00026	-0.00011	0.00008	0.00023	0.00007	0.00024	-0.00044*	-0.00017	0.00050
	(0.00026)	(0.00019)	(0.00027)	(0.00021)	(0.00021)	(0.00018)	(0.00027)	(0.00025)	(0.00033)	(0.00035)
Hrs_second	0.00068*	0.00047*	-0.00045	0.00030	0.00001	0.00008	0.00017	-0.00044	0.00087**	-0.00064*
	(0.00038)	(0.00025)	(0.00048)	(0.00032)	(0.00039)	(0.00036)	(0.00042)	(0.00039)	(0.00035)	(0.00036)
D(Hrs_ff)	0.00810	0.00461	0.01823	0.00807	-0.00600	0.00738	0.01120	-0.01050	0.00003	-0.00851
	(0.02372)	(0.01737)	(0.01976)	(0.01693)	(0.01395)	(0.01216)	(0.01338)	(0.01288)	(0.01184)	(0.01378)
D(Hrs_if)	-0.03404	-0.01617	-0.03723	-0.01610	-0.00091	-0.01264	-0.02889	0.00299	0.01727	-0.02744
	(0.02954)	(0.01746)	(0.03087)	(0.02472)	(0.02178)	(0.01573)	(0.02563)	(0.02145)	(0.02243)	(0.02239)
D(Hrs_ii)	-0.00527	0.00179	0.00976	0.00083	0.00407	-0.02957***	0.00087	-0.00019	-0.01267	0.01772
	(0.01432)	(0.00955)	(0.01408)	(0.01032)	(0.01291)	(0.01075)	(0.01389)	(0.01318)	(0.01647)	(0.01899)
D(Hrs_sf)	-0.02382	-0.04841**	-0.06731**	0.03711	-0.00528	0.00749	0.01216	-0.03554	-0.01278	-0.00987
	(0.03327)	(0.02348)	(0.03122)	(0.02760)	(0.02338)	(0.02339)	(0.02350)	(0.02307)	(0.01760)	(0.02473)
D(Hrs_si)	0.00805	-0.00466	0.01697	-0.00791	0.00658	0.00118	-0.00017	0.01407	-0.01472	-0.01075
	(0.01375)	(0.00942)	(0.01262)	(0.00978)	(0.01162)	(0.00997)	(0.01147)	(0.01095)	(0.01267)	(0.01397)
D(Hrs_fam)	0.00265	-0.00219	-0.00401	0.00827	-0.00101	-0.01108	-0.01726	0.01497	-0.00424	0.01523
	(0.01526)	(0.00955)	(0.01559)	(0.01154)	(0.01399)	(0.01187)	(0.01607)	(0.01619)	(0.01689)	(0.01888)
D(Hrs_second)	-0.01077	-0.00237	-0.00988	0.00678	-0.00382	0.01290	-0.01379	0.02500**	-0.01202	0.01173
	(0.01404)	(0.00934)	(0.01411)	(0.01075)	(0.01265)	(0.01051)	(0.01229)	(0.01189)	(0.01121)	(0.01261)
D(Lima)	0.14794***	0.01142	0.11700***	0.01053	0.11371***	-0.00531	0.07311***	0.01557*	0.01773**	0.04519***
	(0.01789)	(0.01384)	(0.01289)	(0.00970)	(0.01056)	(0.00849)	(0.00994)	(0.00926)	(0.00816)	(0.00955)
Constant	-0.04640	0.02941	0.38632	-0.60286**	0.36949	-1.18351***	1.06495***	-0.81589***	1.21488***	-0.71429***
	(0.11042)	(0.10129)	(0.31516)	(0.23997)	(0.34998)	(0.31518)	(0.29382)	(0.27353)	(0.09451)	(0.11340)
Observations	1,602		1,601		1,602		1,601		1,601	
LogL	1423.676		1614.402		1811.801		1772.687		1871.271	

c) SUR–GLS (cities)

	Metropolitan areas		Intermediate cities		Minor cities	
	Informal	Formal	Informal	Formal	Informal	Formal
L(expenditure)	-0.07134***	0.11111***	-0.02711***	0.08614***	-0.00290	0.06479***
	(0.00826)	(0.00677)	(0.00706)	(0.00575)	(0.00625)	(0.00646)
L(fam_size)	0.02615*	0.01457	0.00445	0.01431	0.00075	-0.00333
	(0.01360)	(0.01153)	(0.01260)	(0.00983)	(0.01238)	(0.01123)
D(mem_spo)	0.01697	-0.01200	0.01782	-0.01995*	0.07234***	-0.05896***
	(0.01225)	(0.01105)	(0.01241)	(0.01020)	(0.01247)	(0.01110)
D(mem_son)	-0.01031	0.00383	0.00179	-0.00998	0.03031**	-0.01728
	(0.01331)	(0.01208)	(0.01352)	(0.01095)	(0.01330)	(0.01166)
D(mem_oth)	-0.00858	0.02122***	0.00494	0.00567	0.02436***	0.00138
	(0.00942)	(0.00809)	(0.00880)	(0.00710)	(0.00899)	(0.00693)
P(men)	-0.03090*	0.01215	-0.03092*	0.00768	-0.01129	0.00995
	(0.01735)	(0.01546)	(0.01606)	(0.01352)	(0.01603)	(0.01439)
age_head	0.00394**	0.00213	0.00092	0.00292**	0.00072	0.00612***
	(0.00164)	(0.00150)	(0.00155)	(0.00126)	(0.00139)	(0.00123)
age_head-sq	-0.00004***	0.00000	-0.00001	-0.00001	0.00000	-0.00005***
	(0.00002)	(0.00001)	(0.00002)	(0.00001)	(0.00001)	(0.00001)

D(gender_head)	0.00024	-0.00273	-0.00297	0.00041	-0.03860***	0.03751***
	(0.01284)	(0.01172)	(0.01289)	(0.01035)	(0.01329)	(0.01154)
year_educ_head	-0.00503***	0.00629***	-0.00355***	0.00548***	-0.00289***	0.00510***
	(0.00097)	(0.00084)	(0.00089)	(0.00072)	(0.00087)	(0.00075)
Hrs_ff	-0.00029*	0.00021	-0.00013	0.00010	-0.00038**	0.00034**
	(0.00015)	(0.00015)	(0.00015)	(0.00014)	(0.00015)	(0.00016)
Hrs_if	0.00006	0.00025	-0.00005	0.00013	0.00055	-0.00018
	(0.00027)	(0.00022)	(0.00034)	(0.00030)	(0.00038)	(0.00025)
Hrs_ii	-0.00017	-0.00039***	-0.00015	-0.00022*	-0.00023	-0.00027**
	(0.00013)	(0.00010)	(0.00015)	(0.00012)	(0.00016)	(0.00011)
Hrs_sf	-0.00035	0.00063	-0.00007	0.00032	0.00000	0.00076***
	(0.00041)	(0.00043)	(0.00025)	(0.00028)	(0.00028)	(0.00027)
Hrs_si	0.00035***	-0.00008	0.00024**	0.00012	0.00022*	0.00013
	(0.00013)	(0.00011)	(0.00011)	(0.00010)	(0.00012)	(0.00009)
Hrs_fam	-0.00012	0.00045	0.00047***	-0.00002	-0.00030*	0.00015
	(0.00033)	(0.00034)	(0.00018)	(0.00015)	(0.00018)	(0.00014)
Hrs_second	0.00032	0.00023	0.00051	-0.00014	-0.00011	0.00018
	(0.00040)	(0.00035)	(0.00032)	(0.00022)	(0.00028)	(0.00024)
D(Hrs_ff)	0.00987	-0.00698	0.01425	-0.00583	0.00582	0.00745
	(0.01195)	(0.01148)	(0.01145)	(0.01021)	(0.01157)	(0.01103)
D(Hrs_if)	-0.00710	-0.02788*	-0.00660	-0.00945	-0.02930	0.00023
	(0.01739)	(0.01430)	(0.02124)	(0.01792)	(0.02139)	(0.01564)
D(Hrs_ii)	0.01085	-0.02074**	0.00576	-0.01078	0.00293	-0.00456
	(0.01089)	(0.00934)	(0.01118)	(0.00908)	(0.01149)	(0.00884)
D(Hrs_sf)	0.00857	-0.02315	-0.01049	0.01078	-0.02381	-0.01228
	(0.02512)	(0.02669)	(0.01775)	(0.01720)	(0.01771)	(0.01590)
D(Hrs_si)	-0.00030	-0.00765	0.02293**	-0.01230	0.00908	-0.00710
	(0.00997)	(0.00895)	(0.00928)	(0.00781)	(0.00967)	(0.00778)
D(Hrs_fam)	0.00493	-0.01379	-0.01092	0.01599*	-0.00349	-0.00086
	(0.01703)	(0.01668)	(0.01099)	(0.00866)	(0.01130)	(0.00873)
D(Hrs_second)	-0.02232*	0.01976**	-0.02206**	0.00916	-0.00127	0.00625
	(0.01151)	(0.01007)	(0.01005)	(0.00793)	(0.00946)	(0.00796)
D(Lima)	0.10057***	-0.00171	n.a.	n.a.	0.08524***	0.01887*
	(0.00714)	(0.00626)	n.a.	n.a.	(0.01274)	(0.01135)
Constant	0.89569***	-0.78849***	0.61289***	-0.58450***	0.35247***	-0.45894***
	(0.08194)	(0.06946)	(0.06754)	(0.05224)	(0.05767)	(0.05635)
Observations	2,140		2,718		3,149	
LogL	2376.565		2880.827		2875.030	

d) SUR–GLS (interaction terms)

	Interactions	
	Informal	Formal
L(expenditure)	-0.02302	0.12219***
	(0.02045)	(0.01749)
D(Ahrs_ff)*L(expenditure)	-0.04882**	-0.00773
	(0.02132)	(0.01849)
D(Ahrs_ii)*L(expenditure)	0.01705	-0.07577***
	(0.02306)	(0.02200)
D(Ahrs_sf)*L(expenditure)	-0.01161	-0.04498**
	(0.02311)	(0.01929)
D(Ahrs_si)*L(expenditure)	0.00276	-0.04986***
	(0.02143)	(0.01817)
L(fam_size)	0.01467	-0.00152

	(0.01092)	(0.00955)
D(mem_spo)	0.04176***	-0.03551***
	(0.01074)	(0.00929)
D(mem_son)	-0.00274	-0.00043
	(0.01131)	(0.00935)
D(mem_oth)	0.00504	0.01737***
	(0.00835)	(0.00658)
P(men)	-0.02156	0.00843
	(0.01330)	(0.01156)
age_head	0.00126	0.00370***
	(0.00121)	(0.00106)
age_head-sq	-0.00001	-0.00002*
	(0.00001)	(0.00001)
D(gender_head)	-0.02473**	0.01964**
	(0.01129)	(0.00951)
year_educ_head	-0.00330***	0.00567***
	(0.00078)	(0.00066)
Hrs_ff	-0.00024	0.00017
	(0.00015)	(0.00014)
Hrs_if	0.00002	0.00040
	(0.00041)	(0.00034)
Hrs_ii	-0.00006	-0.00031***
	(0.00016)	(0.00012)
Hrs_sf	0.00032	0.00045
	(0.00027)	(0.00029)
Hrs_si	0.00032***	0.00011
	(0.00011)	(0.00010)
Hrs_fam	0.00007	0.00014
	(0.00016)	(0.00013)
Hrs_second	0.00063**	-0.00004
	(0.00030)	(0.00024)
D(Hrs_ff)	0.44060**	0.09055
	(0.18343)	(0.14900)
D(Hrs_ii)	-0.11011	0.61880***
	(0.19534)	(0.17532)
D(Hrs_sf)	0.07899	0.39775**
	(0.19829)	(0.15453)
D(Hrs_si)	0.01685	0.42429***
	(0.18318)	(0.14524)
D(Hrs_fam)	-0.00999	-0.00116
	(0.01079)	(0.00854)
D(Hrs_second)	-0.01909**	0.00972
	(0.00921)	(0.00788)
D(Lima)	0.08728***	0.01524**
	(0.00737)	(0.00660)
Constant	0.51999***	-0.92893***
	(0.17770)	(0.14075)
Observations	4,216	
LogL	3875.475	

e) SUR–GLS (five-market definition)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expenditure)	-0.01895***	0.06207***	-0.01912***	0.05936***	-0.01550***	0.06221***	-0.01952***	0.06075***
	(0.00308)	(0.00347)	(0.00308)	(0.00344)	(0.00304)	(0.00352)	(0.00306)	(0.00344)
L(fam_size)	-0.03523***	0.02962***	-0.03247***	0.02942***	-0.03494***	0.02927***	-0.03400***	0.03125***
	(0.00543)	(0.00513)	(0.00533)	(0.00507)	(0.00538)	(0.00513)	(0.00535)	(0.00511)
D(mem_spo)	0.00251	-0.00376	0.00281	-0.00181	0.00295	-0.00354	0.00214	-0.00296
	(0.00537)	(0.00477)	(0.00528)	(0.00471)	(0.00535)	(0.00477)	(0.00530)	(0.00474)
D(mem_son)	-0.00410	-0.00336	-0.00592	-0.00294	-0.00225	-0.00348	-0.00316	-0.00305
	(0.00574)	(0.00535)	(0.00560)	(0.00531)	(0.00571)	(0.00535)	(0.00566)	(0.00534)
D(mem_oth)	0.00195	0.00225	0.00401	0.00268	0.00475	0.00227	0.00194	0.00242
	(0.00390)	(0.00351)	(0.00380)	(0.00349)	(0.00386)	(0.00351)	(0.00385)	(0.00349)
P(men)	0.01365*	-0.03935***	0.01579**	-0.03836***	0.01416**	-0.03954***	0.01462**	-0.03992***
	(0.00710)	(0.00686)	(0.00698)	(0.00681)	(0.00704)	(0.00686)	(0.00703)	(0.00685)
age_head	-0.00002	0.00322***	-0.00009	0.00319***	-0.00027	0.00318***	-0.00011	0.00316***
	(0.00064)	(0.00062)	(0.00062)	(0.00061)	(0.00063)	(0.00062)	(0.00063)	(0.00061)
age_head-sq	-0.00000	-0.00001**	-0.00000	-0.00002***	-0.00000	-0.00001**	-0.00000	-0.00001**
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
D(gender_head)	0.00582	-0.00861*	0.00333	-0.00850*	0.00353	-0.00869*	0.00730	-0.00955*
	(0.00558)	(0.00509)	(0.00551)	(0.00503)	(0.00554)	(0.00510)	(0.00551)	(0.00507)
year_educ_head	-0.00127***	0.00570***	-0.00087**	0.00511***	-0.00066*	0.00563***	-0.00158***	0.00568***
	(0.00039)	(0.00036)	(0.00039)	(0.00037)	(0.00039)	(0.00037)	(0.00039)	(0.00036)
Hrs_ff	0.00004	-0.00008	0.00001	-0.00009	0.00004	-0.00008	0.00004	-0.00008
	(0.00006)	(0.00007)	(0.00006)	(0.00007)	(0.00006)	(0.00007)	(0.00006)	(0.00007)
Hrs_if	0.00051***	-0.00024*	0.00049***	-0.00022*	0.00052***	-0.00024*	0.00053***	-0.00023*
	(0.00014)	(0.00013)	(0.00013)	(0.00013)	(0.00014)	(0.00013)	(0.00013)	(0.00013)
Hrs_ii	0.00023***	-0.00041***	0.00019***	-0.00038***	0.00022***	-0.00041***	0.00024***	-0.00041***
	(0.00006)	(0.00005)	(0.00006)	(0.00005)	(0.00006)	(0.00005)	(0.00006)	(0.00005)
Hrs_sf	0.00009	0.00040***	0.00010	0.00043***	0.00007	0.00041***	0.00008	0.00041***
	(0.00013)	(0.00014)	(0.00013)	(0.00014)	(0.00013)	(0.00014)	(0.00013)	(0.00014)
Hrs_si	0.00049***	-0.00015***	0.00044***	-0.00014***	0.00048***	-0.00014***	0.00048***	-0.00015***
	(0.00006)	(0.00005)	(0.00006)	(0.00005)	(0.00006)	(0.00005)	(0.00006)	(0.00005)
Hrs_fam	0.00015	0.00003	0.00012	0.00003	0.00011	0.00004	0.00016*	0.00003
	(0.00010)	(0.00008)	(0.00009)	(0.00008)	(0.00010)	(0.00008)	(0.00010)	(0.00008)
Hrs_second	0.00039***	-0.00004	0.00029**	-0.00006	0.00044***	-0.00004	0.00037***	-0.00006
	(0.00013)	(0.00012)	(0.00013)	(0.00012)	(0.00013)	(0.00012)	(0.00013)	(0.00012)
D(Hrs_ff)	0.03208***	-0.00043	0.03229***	-0.00003	0.02901***	-0.00074	0.03110***	-0.00186
	(0.00528)	(0.00536)	(0.00515)	(0.00530)	(0.00522)	(0.00536)	(0.00519)	(0.00530)
D(Hrs_if)	-0.01241	-0.01045	-0.01195	-0.01070	-0.01199	-0.01070	-0.01374*	-0.01101
	(0.00795)	(0.00803)	(0.00777)	(0.00802)	(0.00803)	(0.00803)	(0.00782)	(0.00802)
D(Hrs_ii)	0.00399	-0.01075**	0.00420	-0.01145***	0.00319	-0.01051**	0.00427	-0.01011**
	(0.00493)	(0.00421)	(0.00482)	(0.00417)	(0.00489)	(0.00421)	(0.00488)	(0.00418)
D(Hrs_sf)	0.01234	-0.01326	0.01487*	-0.01560*	0.01030	-0.01377*	0.01326	-0.01312
	(0.00855)	(0.00834)	(0.00833)	(0.00829)	(0.00844)	(0.00835)	(0.00845)	(0.00832)
D(Hrs_si)	0.00425	-0.01281***	0.00390	-0.01211***	0.00365	-0.01278***	0.00492	-0.01166***
	(0.00420)	(0.00381)	(0.00413)	(0.00378)	(0.00416)	(0.00381)	(0.00415)	(0.00379)
D(Hrs_fam)	0.00677	0.00735	0.00719	0.00695	0.00548	0.00732	0.00597	0.00730
	(0.00560)	(0.00465)	(0.00542)	(0.00465)	(0.00553)	(0.00465)	(0.00557)	(0.00464)
D(Hrs_second)	-0.00652	0.00948**	-0.00605	0.00986**	-0.00756*	0.00907**	-0.00628	0.01020**
	(0.00450)	(0.00417)	(0.00441)	(0.00415)	(0.00445)	(0.00417)	(0.00446)	(0.00416)
D(Lima)	-0.01863***	0.02765***	0.00165	0.03195***	-0.01957***	0.02840***	-0.01424***	0.02812***
	(0.00331)	(0.00401)	(0.00630)	(0.00703)	(0.00338)	(0.00406)	(0.00340)	(0.00407)

L(pop_urb_dist)			-0.01398	-0.00288				
			(0.01811)	(0.01600)				
L(pop_urb_dist)-sq			0.00074	-0.00000				
			(0.00083)	(0.00075)				
pop_den_distx1000			-0.00142***	0.00036				
			(0.00031)	(0.00037)				
D(not_slum)			-0.01554***	0.02957***				
			(0.00310)	(0.00289)				
D(mid_city)			0.00168	0.00164				
			(0.00373)	(0.00304)				
D(border)			0.07303***	0.01892***				
			(0.00557)	(0.00427)				
D(tongue)					0.03993***	-0.00143		
					(0.00389)	(0.00318)		
D(migrant)					0.00170	-0.00149		
					(0.00308)	(0.00287)		
D(social)					0.00276	0.00423		
					(0.00292)	(0.00274)		
Station_distx1000							-0.03382***	-0.00263
							(0.01311)	(0.01163)
Muni_persx1000							-0.00024	0.00242***
							(0.00038)	(0.00045)
L(budget_dist_pc)							0.02465***	0.00295
							(0.00259)	(0.00233)
D(sdp_dist)							0.03876***	-0.02712***
							(0.00748)	(0.00510)
Constant	0.34952***	-0.48084***	0.40570***	-0.43689***	0.31077***	-0.48040***	0.23252***	-0.49226***
	(0.02838)	(0.03059)	(0.09943)	(0.09042)	(0.02822)	(0.03100)	(0.03006)	(0.03226)
Observations	8,007		8,007		8,007		8,007	
LogL	11051.017		11286.781		11123.226		11195.208	

f) OLS under IV sample (grouped hours of work)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expenditure)	-0.02398***	0.08261***	-0.03223***	0.08006***	-0.02213***	0.08166***	-0.02292***	0.08063***
	(0.00548)	(0.00526)	(0.00561)	(0.00521)	(0.00547)	(0.00530)	(0.00547)	(0.00524)
Hrs_for	-0.00022**	0.00017*	-0.00024**	0.00018*	-0.00023**	0.00017*	-0.00024**	0.00018*
	(0.00011)	(0.00010)	(0.00010)	(0.00010)	(0.00011)	(0.00010)	(0.00011)	(0.00010)
Hrs_inf	0.00004	-0.00015***	0.00001	-0.00012**	0.00002	-0.00014***	0.00004	-0.00014***
	(0.00007)	(0.00005)	(0.00007)	(0.00005)	(0.00007)	(0.00005)	(0.00007)	(0.00005)
D(Hrs_for)	0.00694	-0.00180	0.01402*	-0.00260	0.00424	-0.00036	0.00834	-0.00342
	(0.00861)	(0.00764)	(0.00841)	(0.00757)	(0.00863)	(0.00767)	(0.00861)	(0.00760)
D(Hrs_inf)	0.01947***	-0.01997***	0.01898***	-0.01956***	0.01901***	-0.01983***	0.01882***	-0.01848***
	(0.00702)	(0.00615)	(0.00696)	(0.00612)	(0.00699)	(0.00615)	(0.00700)	(0.00612)
L(fam_size)	0.01167	0.00062	0.00838	0.00021	0.01183	0.00078	0.01243	0.00168
	(0.00973)	(0.00814)	(0.00969)	(0.00809)	(0.00964)	(0.00811)	(0.00969)	(0.00813)
D(mem_spo)	0.04193***	-0.02657***	0.04043***	-0.02436***	0.04020***	-0.02644***	0.04181***	-0.02548***
	(0.00906)	(0.00765)	(0.00889)	(0.00761)	(0.00906)	(0.00765)	(0.00904)	(0.00762)
D(mem_son)	0.00763	-0.00628	0.00434	-0.00561	0.00962	-0.00681	0.00669	-0.00591
	(0.00983)	(0.00827)	(0.00968)	(0.00821)	(0.00979)	(0.00827)	(0.00980)	(0.00824)
D(mem_oth)	0.01337*	0.01090*	0.01339*	0.01088*	0.01526**	0.01011*	0.01378*	0.01144**
	(0.00710)	(0.00570)	(0.00704)	(0.00569)	(0.00706)	(0.00568)	(0.00709)	(0.00570)
P(men)	-0.02363*	0.01195	-0.02163*	0.01187	-0.02228*	0.01146	-0.02118*	0.01087

	(0.01227)	(0.01085)	(0.01209)	(0.01080)	(0.01225)	(0.01085)	(0.01225)	(0.01083)
age_head	0.00144	0.00398***	0.00163	0.00395***	0.00128	0.00407***	0.00168	0.00384***
	(0.00107)	(0.00094)	(0.00105)	(0.00094)	(0.00107)	(0.00094)	(0.00107)	(0.00094)
age_head-sq	-0.00001	-0.00002**	-0.00002	-0.00002**	-0.00001	-0.00002**	-0.00002	-0.00002**
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
D(gender_head)	-0.01878*	0.01366*	-0.01707*	0.01418*	-0.02164**	0.01475*	-0.01852*	0.01215
	(0.00961)	(0.00813)	(0.00939)	(0.00811)	(0.00961)	(0.00815)	(0.00958)	(0.00812)
year_educ_head	-0.00471***	0.00660***	-0.00469***	0.00585***	-0.00394***	0.00640***	-0.00475***	0.00659***
	(0.00070)	(0.00058)	(0.00069)	(0.00059)	(0.00072)	(0.00059)	(0.00070)	(0.00058)
D(Lima)	0.07594***	0.01887***	0.06023***	0.01999**	0.07379***	0.01887***	0.06892***	0.02163***
	(0.00634)	(0.00557)	(0.01206)	(0.00999)	(0.00639)	(0.00563)	(0.00647)	(0.00566)
L(pop_urb_dist)			0.14875***	-0.01625				
			(0.03223)	(0.03101)				
L(pop_urb_dist)-sq			-0.00587***	0.00050				
			(0.00149)	(0.00142)				
pop_den_distx1000			-0.00088	0.00093*				
			(0.00062)	(0.00054)				
D(not_slum)			-0.00250	0.03277***				
			(0.00533)	(0.00488)				
D(mid_city)			-0.02046***	0.00378				
			(0.00635)	(0.00544)				
D(border)			0.05884***	0.00464				
			(0.00828)	(0.00707)				
D(tongue)					0.03389***	-0.01242**		
					(0.00647)	(0.00548)		
D(migrant)					0.01424***	-0.00200		
					(0.00543)	(0.00471)		
D(social)					0.00213	-0.00439		
					(0.00524)	(0.00440)		
Station_distx1000							-0.09906***	0.03508**
							(0.01909)	(0.01661)
Muni_persx1000							-0.00118***	0.00237***
							(0.00042)	(0.00055)
L(budget_dist_pc)							0.00652	0.00147
							(0.00423)	(0.00365)
D(sdp_dist)							-0.04401***	-0.01235
							(0.01140)	(0.01077)
Constant	0.55185***	-0.57804***	-0.29124*	-0.44894**	0.51749***	-0.56565***	0.51769***	-0.57772***
	(0.04974)	(0.04633)	(0.17506)	(0.17632)	(0.04971)	(0.04688)	(0.05382)	(0.04952)
Observations	4,874	4,874	4,874	4,874	4,874	4,874	4,874	4,874
R-squared	0.103	0.283	0.131	0.293	0.110	0.284	0.110	0.290
LogL	1612.358	2355.668	1691.885	2387.230	1633.263	2359.278	1631.223	2378.663

g) OLS under IV sample (original hours of work)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expenditure)	-0.02408***	0.08342***	-0.03199***	0.08093***	-0.02231***	0.08250***	-0.02302***	0.08144***
	(0.00549)	(0.00531)	(0.00563)	(0.00526)	(0.00549)	(0.00534)	(0.00548)	(0.00528)
Hrs_ff	-0.00030***	0.00023**	-0.00031***	0.00022**	-0.00030**	0.00022**	-0.00033***	0.00024**
	(0.00012)	(0.00010)	(0.00012)	(0.00010)	(0.00012)	(0.00010)	(0.00012)	(0.00010)
Hrs_if	0.00004	0.00031	0.00002	0.00032	0.00005	0.00030	0.00003	0.00033*
	(0.00024)	(0.00020)	(0.00024)	(0.00020)	(0.00025)	(0.00020)	(0.00024)	(0.00020)
Hrs_ii	-0.00008	-0.00036***	-0.00009	-0.00032***	-0.00010	-0.00036***	-0.00008	-0.00036***

	(0.00010)	(0.00008)	(0.00011)	(0.00008)	(0.00010)	(0.00008)	(0.00010)	(0.00008)
Hrs_sf	0.00007	0.00031	0.00000	0.00035	0.00004	0.00032	0.00009	0.00029
	(0.00030)	(0.00032)	(0.00030)	(0.00033)	(0.00031)	(0.00032)	(0.00030)	(0.00032)
Hrs_si	0.00027**	-0.00003	0.00022**	-0.00001	0.00025**	-0.00003	0.00027**	-0.00003
	(0.00011)	(0.00009)	(0.00011)	(0.00009)	(0.00011)	(0.00009)	(0.00011)	(0.00009)
D(Hrs_ff)	0.01506*	-0.00812	0.01925**	-0.00774	0.01221	-0.00660	0.01723*	-0.01041
	(0.00903)	(0.00807)	(0.00887)	(0.00798)	(0.00905)	(0.00809)	(0.00905)	(0.00799)
D(Hrs_if)	-0.00956	-0.03172**	-0.01060	-0.03013**	-0.00908	-0.03154**	-0.01067	-0.03156**
	(0.01513)	(0.01236)	(0.01499)	(0.01249)	(0.01535)	(0.01246)	(0.01508)	(0.01235)
D(Hrs_ii)	0.00759	-0.01122	0.00654	-0.01240*	0.00666	-0.01119	0.00722	-0.01073
	(0.00861)	(0.00691)	(0.00858)	(0.00691)	(0.00859)	(0.00693)	(0.00858)	(0.00689)
D(Hrs_sf)	-0.02411	0.00016	-0.01099	-0.00329	-0.02501	0.00114	-0.02668	0.00288
	(0.01889)	(0.01917)	(0.01832)	(0.01928)	(0.01901)	(0.01924)	(0.01876)	(0.01917)
D(Hrs_si)	0.01332*	-0.00551	0.01319*	-0.00507	0.01303*	-0.00529	0.01267	-0.00426
	(0.00781)	(0.00655)	(0.00773)	(0.00652)	(0.00775)	(0.00653)	(0.00777)	(0.00654)
L(fam_size)	0.01361	0.00367	0.01044	0.00317	0.01374	0.00388	0.01428	0.00480
	(0.00978)	(0.00817)	(0.00975)	(0.00811)	(0.00969)	(0.00814)	(0.00974)	(0.00816)
D(mem_spo)	0.04134***	-0.02768***	0.03971***	-0.02535***	0.03972***	-0.02757***	0.04126***	-0.02655***
	(0.00907)	(0.00766)	(0.00890)	(0.00763)	(0.00908)	(0.00766)	(0.00905)	(0.00763)
D(mem_son)	0.00648	-0.00784	0.00362	-0.00715	0.00848	-0.00835	0.00540	-0.00730
	(0.00982)	(0.00826)	(0.00967)	(0.00820)	(0.00978)	(0.00826)	(0.00978)	(0.00823)
D(mem_oth)	0.01276*	0.01026*	0.01273*	0.01035*	0.01457**	0.00946*	0.01321*	0.01077*
	(0.00708)	(0.00569)	(0.00702)	(0.00568)	(0.00704)	(0.00567)	(0.00706)	(0.00569)
P(men)	-0.02284*	0.01134	-0.02107*	0.01150	-0.02156*	0.01087	-0.02027*	0.01009
	(0.01225)	(0.01085)	(0.01208)	(0.01080)	(0.01224)	(0.01085)	(0.01223)	(0.01083)
age_head	0.00122	0.00380***	0.00142	0.00377***	0.00107	0.00389***	0.00148	0.00363***
	(0.00107)	(0.00094)	(0.00105)	(0.00094)	(0.00107)	(0.00095)	(0.00107)	(0.00094)
age_head-sq	-0.00001	-0.00002**	-0.00001	-0.00002**	-0.00001	-0.00002**	-0.00001	-0.00002*
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
D(gender_head)	-0.01950**	0.01336*	-0.01766*	0.01373*	-0.02223**	0.01445*	-0.01924**	0.01174
	(0.00960)	(0.00811)	(0.00939)	(0.00809)	(0.00961)	(0.00813)	(0.00957)	(0.00810)
year_educ_head	-0.00481***	0.00673***	-0.00476***	0.00597***	-0.00405***	0.00654***	-0.00484***	0.00673***
	(0.00070)	(0.00058)	(0.00069)	(0.00059)	(0.00072)	(0.00059)	(0.00070)	(0.00058)
D(Lima)	0.07702***	0.01972***	0.06108***	0.02306**	0.07474***	0.01971***	0.06972***	0.02264***
	(0.00632)	(0.00557)	(0.01207)	(0.00995)	(0.00638)	(0.00563)	(0.00646)	(0.00566)
L(pop_urb_dist)			0.14830***	-0.01820				
			(0.03229)	(0.03112)				
L(pop_urb_dist)-sq			-0.00586***	0.00059				
			(0.00149)	(0.00143)				
pop_den_distx1000			-0.00087	0.00074				
			(0.00062)	(0.00054)				
D(not_slum)			-0.00289	0.03270***				
			(0.00532)	(0.00489)				
D(mid_city)			-0.02090***	0.00454				
			(0.00636)	(0.00545)				
D(border)			0.05705***	0.00395				
			(0.00828)	(0.00710)				
D(tongue)					0.03376***	-0.01226**		
					(0.00645)	(0.00548)		
D(migrant)					0.01355**	-0.00195		
					(0.00542)	(0.00471)		
D(social)					0.00094	-0.00460		
					(0.00525)	(0.00441)		
Station_distx1000							-0.10055***	0.03795**

							(0.01893)	(0.01671)
Muni_persx1000							-0.00119***	0.00247***
							(0.00042)	(0.00055)
L(budget_dist_pc)							0.00562	0.00081
							(0.00423)	(0.00363)
D(sdp_dist)							-0.04506***	-0.01311
							(0.01134)	(0.01078)
Constant	0.56224***	-0.58805***	-0.27940	-0.44800**	0.52892***	-0.57588***	0.53214***	-0.58392***
	(0.04940)	(0.04632)	(0.17522)	(0.17721)	(0.04939)	(0.04685)	(0.05346)	(0.04935)
Observations	4,874	4,874	4,874	4,874	4,874	4,874	4,874	4,874
R-squared	0.107	0.287	0.135	0.296	0.115	0.288	0.114	0.294
LogL	1624.738	2367.680	1701.972	2398.588	1644.940	2371.271	1644.318	2392.309

h) First-stage regressions of IV models (expenditure)

	(1)	(2)	(3)	(4)
L(fam_size)	-0.6975723***	-0.6915396***	-0.6902081***	-0.6924406***
	(0.02453)	(0.02402)	(0.02434)	(0.02458)
D(mem_spo)	0.0449464*	0.0431487*	0.0348157	0.0461381***
	(0.02359)	(0.02293)	(0.02338)	(0.02360)
D(mem_son)	0.1183673***	0.1074933***	0.1162343***	0.1191881***
	(0.02726)	(0.02667)	(0.02717)	(0.02727)
D(mem_oth)	0.0933388***	0.0862209***	0.0883539***	0.0940166***
	(0.01941)	(0.01908)	(0.01938)	(0.01937)
P(men)	0.1234433***	0.1177695***	0.1200715***	0.121575***
	(0.03624)	(0.03569)	(0.03605)	(0.03635)
age_head	0.0191748***	0.0194592***	0.0195282***	0.0188211***
	(0.00341)	(0.00335)	(0.00338)	(0.00341)
age_head-sq	-0.0001123***	-0.0001268***	-0.0001165***	-0.0001101***
	(0.00003)	(0.00003)	(0.00003)	(0.00003)
D(gender_head)	-0.0660128***	-0.0472274*	-0.0618247**	-0.0681733***
	(0.02493)	(0.02436)	(0.02481)	(0.02497)
year_educ_head	0.0383932***	0.0339468***	0.0376297***	0.0381612***
	(0.00202)	(0.00201)	(0.00205)	(0.00202)
D(Lima)	0.2211205***	0.1677943***	0.2089685***	0.2240728***
	(0.01756)	(0.03340)	(0.01778)	(0.01769)
Hrs_for	0.0030804***	0.0030164***	0.002984***	0.0030786***
	(0.00032)	(0.00032)	(0.00033)	(0.00032)
Hrs_inf	0.0016085***	0.0015957***	0.0015853***	0.001599***
	(0.00019)	(0.00018)	(0.00019)	(0.00019)
D(Hrs_for)	0.0696895***	0.0879349***	0.083583***	0.0671219***
	(0.02573)	(0.02492)	(0.02576)	(0.02572)
D(Hrs_inf)	-0.0403507*	-0.0407467*	-0.037664*	-0.0371154*
	(0.02110)	(0.02092)	(0.02104)	(0.02108)
L(asset)	0.1968538***	0.1892478***	0.1939483***	0.1958798***
	(0.00601)	(0.00585)	(0.00598)	(0.00602)
D(risk1)	-0.0641129*	-0.0735015**	-0.0662165**	-0.0652723**
	(0.03313)	(0.03264)	(0.03306)	(0.03319)

D(risk2)	-0.0110809	-0.0337895	-0.0161748	-0.0041375
	(0.07150)	(0.07012)	(0.07033)	(0.07138)
D(risk5)	-0.10134*	-0.0912392	-0.0963034	-0.1040545*
	(0.06054)	(0.06168)	(0.06046)	(0.06048)
D(risk7)	0.0502437	0.0460096	0.0512048	0.0516704
	(0.03680)	(0.03679)	(0.03634)	(0.03681)
D(risk8)	-0.2534471***	-0.1579162*	-0.2141763**	-0.2495649***
	(0.09106)	(0.08743)	(0.08972)	(0.09047)
L(pop_urb_dist)		0.579283***		
		(0.10098)		
L(pop_urb_dist)-sq		-0.0247001***		
		(0.00461)		
pop_den_distx1000		0.0045573***		
		(0.00173)		
D(not_slum)		0.0959097***		
		(0.01548)		
D(mid_city)		-0.0092219		
		(0.01820)		
D(border)		0.0716459***		
		(0.02273)		
D(tongue)			-0.0886418***	
			(0.01849)	
D(migrant)			0.0692252***	
			(0.01554)	
D(social)			-0.0352485**	
			(0.01433)	
Station_distx1000				-0.0344683
				(0.05050)
Muni_persx1000				0.0052801***
				(0.00145)
L(budget_dist_pc)				0.0081527
				(0.01095)
D(sdp_dist)				0.0072878
				(0.03344)
Constant	6.169612***	2.877529***	6.17717***	6.124679***
	(0.09429)	(0.55868)	(0.09394)	(0.11144)
Observations	4,874	4,874	4,874	4,874
R-squared	0.5472	0.5635	0.5515	0.5491

i) First-stage regressions of IV models (hours of work)

	Hrs_for				Hrs_inf			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
L(expenditure)	5.15764***	5.088257***	5.018985***	5.126702***	9.628624***	9.688778***	9.581617***	9.585205***
	(0.61681)	(0.62578)	(0.62027)	(0.61582)	(1.14570)	(1.16245)	(1.15085)	(1.14683)
L(fam_size)	9.552103***	9.502377***	9.503541***	9.501758***	29.13651***	28.89112***	29.02519***	29.10518***

	(1.37714)	(1.37445)	(1.38117)	(1.37846)	(2.48376)	(2.47801)	(2.48932)	(2.48617)
D(mem_spo)	1.393285	1.377112	1.180662	1.405535	-2.68102	-2.763261	-3.087059	-2.679788
	(1.14293)	(1.14481)	(1.14411)	(1.14505)	(2.11905)	(2.11578)	(2.11889)	(2.11991)
D(mem_son)	-3.558813***	-3.58792***	-3.537515***	-3.558535***	-7.630036***	-7.713073***	-7.379838***	-7.56945***
	(1.23758)	(1.23693)	(1.23984)	(1.23948)	(2.09092)	(2.08717)	(2.09336)	(2.09027)
D(mem_oth)	1.471237	1.461777	1.448791	1.496924	5.622545***	5.724788***	5.754224***	5.645524***
	(0.91550)	(0.91704)	(0.91980)	(0.91781)	(1.79472)	(1.79133)	(1.79894)	(1.79606)
P(men)	2.768288**	2.857244**	2.765719**	2.792928**	11.22202***	11.49028***	11.32958***	11.2042***
	(1.33954)	(1.33989)	(1.33953)	(1.33960)	(2.41391)	(2.40757)	(2.41005)	(2.41202)
age_head	0.01832	0.0132672	0.0203231	0.0215116	1.376268***	1.354409***	1.369708***	1.371784***
	(0.09332)	(0.09352)	(0.09360)	(0.09339)	(0.19190)	(0.19223)	(0.19217)	(0.19213)
age_head-sq	-0.0007371	-0.0006746	-0.0007592	-0.0007666	-0.0154087***	-0.014886***	-0.0152942***	-0.0153868***
	(0.00090)	(0.00091)	(0.00090)	(0.00090)	(0.00184)	(0.00185)	(0.00184)	(0.00184)
D(gender_head)	-1.52107	-1.557888	-1.54355	-1.545931	3.998572*	3.846052*	3.714862*	3.961498*
	(1.20819)	(1.21044)	(1.20752)	(1.21132)	(2.10051)	(2.09554)	(2.09560)	(2.10154)
year_educ_head	-0.3573299***	-0.3541498***	-0.3533869***	-0.3495436***	-1.141636***	-1.060301***	-1.048528***	-1.137585***
	(0.07949)	(0.08007)	(0.08290)	(0.08015)	(0.14825)	(0.15122)	(0.15172)	(0.14924)
D(Lima)	0.996858	0.655516	0.8816171	0.7608007	3.176063**	4.325263	2.760363*	3.35382**
	(0.76154)	(1.28011)	(0.76641)	(0.78193)	(1.46166)	(2.66824)	(1.45119)	(1.51426)
D(Hrs_for)	57.39319***	57.45467***	57.51025***	57.40119***	-23.29264***	-23.10417***	-23.38378***	-23.32905***
	(0.69395)	(0.69668)	(0.69682)	(0.69675)	(1.39493)	(1.40307)	(1.39613)	(1.39919)
D(Hrs_inf)	-5.507792***	-5.543583***	-5.480241***	-5.525333***	54.63505***	54.33864***	54.49805***	54.64826***
	(0.77043)	(0.76917)	(0.77025)	(0.77216)	(0.95173)	(0.95183)	(0.95047)	(0.95162)
L(asset)	-5.932318***	-6.12688***	-5.968978***	-6.045492***	5.20794**	4.737251**	5.354384**	5.270377**
	(1.25793)	(1.25951)	(1.25558)	(1.25816)	(2.14413)	(2.14362)	(2.13761)	(2.14508)
L(asset)-sq	0.4709992***	0.4848168***	0.4727657***	0.4792468***	-0.4678609***	-0.4297405***	-0.4732164***	-0.4717131***
	(0.09921)	(0.09932)	(0.09904)	(0.09928)	(0.15130)	(0.15135)	(0.15092)	(0.15138)
L(wage_for)	-2.736262***	-2.926127***	-2.80308***	-2.546954***	-2.601397*	-2.773144*	-2.817543**	-2.639351*
	(0.75251)	(0.75632)	(0.75752)	(0.76908)	(1.40251)	(1.43440)	(1.39882)	(1.44543)
L(wage_inf)	-1.116992	-1.369418	-1.309705	-1.248271	-3.333609**	-4.404078***	-3.540152**	-3.936145**
	(0.82136)	(0.84185)	(0.82073)	(0.86007)	(1.48721)	(1.52465)	(1.48405)	(1.59145)
D(house)	-2.565865***	-2.788906***	-2.743967***	-2.453775***	-1.799317	-3.05862	-1.554497	-1.843209
	(0.65753)	(0.68887)	(0.66695)	(0.66353)	(2.35923)	(2.39424)	(2.36537)	(2.37202)
D(risk7)	0.9685094	1.019759	0.9723995	0.9487966	4.062535	4.274697	3.956857	4.131356
	(1.85312)	(1.85304)	(1.85193)	(1.85193)	(2.88844)	(2.90277)	(2.88698)	(2.89331)
non_labour	-0.0001618	-0.000159	-0.000155	-0.0001577	-0.0005024***	-0.00048***	-0.0004984***	--0.0005018***
	(0.00012)	(0.00012)	(0.00012)	(0.00012)	(0.00016)	(0.00017)	(0.00017)	(0.00017)
L(pop_urb_dist)		-0.2038506				-2.935724		
		(3.30402)				(5.83049)		
L(pop_urb_dist)-sq		0.0400299				0.195091		
		(0.15566)				(0.27715)		
pop_den_distx1000		-0.0953555				-0.1452668		
		(0.06967)				(0.13243)		
D(not_slum)		0.0483118				-4.17967***		
		(0.65011)				(1.17931)		
D(mid_city)		-0.5530829				1.983565		

		(0.71896)				(1.32011)		
D(border)		0.8515387				3.945957**		
		(0.94238)				(1.70347)		
D(tongue)			-0.9660547				2.470255*	
			(0.68482)				(1.42868)	
D(migrant)			1.609215**				3.079681***	
			(0.62300)				(1.10092)	
D(social)			-0.2155287				-0.7620457	
			(0.63644)				(1.12408)	
Station_distx1000				-3.893335**				-5.136164
				(1.77584)				(3.65591)
Muni_persx1000				0.0475494				0.1416813**
				(0.05518)				(0.05988)
L(budget_dist_pc)				-0.4275611				-0.1400681
				(0.44043)				(0.85976)
D(sdp_dist)				-0.8743937				2.051915
				(1.50239)				(2.84038)
Constant	-18.70041***	-19.00555	-17.76893***	-16.04093***	-113.8754***	-103.2736***	-116.4059***	-112.5584
	(5.90547)	(17.73516)	(5.91682)	(6.12741)	(11.67244)	(31.81863)	(11.72922)	(12.58776)
Observations	4,874	4,874	4,874	4,874	4,874	4,874	4,874	4,874
R-squared	0.7192	0.7196	0.7196	0.7194	0.5296	0.5329	0.5308	0.5299

j) First-stage regression of IV models (difference in hours of work)

	(1)	(2)	(3)	(4)
L(expenditure)	-4.25354***	-4.39362***	-4.34027***	-4.19901***
	(1.22575)	(1.25233)	(1.23760)	(1.22879)
L(fam_size)	-19.52928***	-19.37375***	-19.46409***	-19.50684***
	(2.94984)	(2.94887)	(2.95912)	(2.95249)
D(mem_spo)	4.09759*	4.15340*	4.27628*	4.13410*
	(2.48627)	(2.48656)	(2.48941)	(2.48919)
D(mem_son)	4.06885	4.09804*	3.84386	4.01706
	(2.48541)	(2.48196)	(2.48987)	(2.48589)
D(mem_oth)	-4.13778**	-4.25365**	-4.29315**	-4.12101**
	(2.05107)	(2.05059)	(2.05782)	(2.05381)
P(men)	-8.43078***	-8.60070***	-8.53929***	-8.40054***
	(2.82872)	(2.82224)	(2.82730)	(2.82698)
age_head	-1.35720***	-1.33668***	-1.34816***	-1.35164***
	(0.21598)	(0.21596)	(0.21639)	(0.21623)
age_head-sq	0.01468***	0.01419***	0.01454***	0.01465***
	(0.00208)	(0.00208)	(0.00208)	(0.00208)
D(gender_head)	-5.58159**	-5.44399**	-5.32574**	-5.61209**
	(2.48829)	(2.48810)	(2.48335)	(2.49055)
year_educ_head	0.79105***	0.71168***	0.70357***	0.79618***
	(0.17106)	(0.17462)	(0.17597)	(0.17217)
D(Lima)	-1.65281	-2.93396	-1.32927	-1.82328
	(1.58804)	(2.95379)	(1.58865)	(1.60798)
D(Hrs_for)	80.56848***	80.47478***	80.77729***	80.55708***
	(1.57068)	(1.58079)	(1.57293)	(1.57312)
D(Hrs_inf)	-60.15839***	-59.91292***	-59.99505***	-60.16554***

	(1.21484)	(1.21353)	(1.21320)	(1.21585)
L(asset)	-11.25296***	-11.06583***	-11.44560***	-11.39197***
	(2.55243)	(2.55794)	(2.54756)	(2.55468)
L(asset)-sq	0.94881***	0.93170***	0.95675***	0.95849***
	(0.18531)	(0.18572)	(0.18496)	(0.18551)
wage_diff	-1.00984	-1.39179	-0.92498	-1.05220
	(1.31664)	(1.33647)	(1.31803)	(1.34378)
D(house)	-0.76078	0.27772	-1.19075	-0.60279
	(2.54671)	(2.59608)	(2.55569)	(2.56238)
D(risk7)	-3.04177	-3.19436	-2.92900	-3.09779
	(3.62868)	(3.63026)	(3.64592)	(3.63025)
non_labour	0.00035*	0.00033*	0.00035*	0.00035*
	(0.00018)	(0.00018)	(0.00019)	(0.00019)
L(pop_urb_dist)		4.83140		
		(6.77642)		
L(pop_urb_dist)-sq		-0.24498		
		(0.32163)		
pop_den_distx1000		0.05073		
		(0.15339)		
D(not_slum)		4.26783***		
		(1.38499)		
D(mid_city)		-2.52719		
		(1.54616)		
D(border)		-2.94033		
		(1.99813)		
D(tongue)			-3.47815**	
			(1.61603)	
D(migrant)			-1.35294	
			(1.29591)	
D(social)			0.49332	
			(1.32991)	
Station_distx1000				1.79953
				(4.29812)
Muni_persx1000				-0.05110
				(0.08639)
L(budget_dist_pc)				-0.22936
				(0.98964)
D(sdp_dist)				-2.94595
				(3.37837)
Constant	96.15398***	74.43575**	99.73086***	97.38025***
	(12.97152)	(37.08107)	(13.03686)	(13.93127)
Observations	4,874	4,874	4,874	4,874
R-squared	0.662	0.663	0.662	0.662

k) Second-stage IV estimation (expenditure)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expenditure)	-0.03660***	0.12819***	-0.04655***	0.12788***	-0.03289***	0.12799***	-0.03406***	0.12567***
	(0.00968)	(0.00860)	(0.01007)	(0.00899)	(0.00985)	(0.00880)	(0.00967)	(0.00859)
L(fam_size)	0.00330	0.03082***	-0.00103	0.03163***	0.00479	0.03108***	0.00512	0.03126***
	(0.01089)	(0.00927)	(0.01099)	(0.00942)	(0.01085)	(0.00930)	(0.01084)	(0.00926)

D(mem_spo)	0.04302***	-0.03050***	0.04161***	-0.02828***	0.04097***	-0.02976***	0.04280***	-0.02946***
	(0.00909)	(0.00783)	(0.00892)	(0.00777)	(0.00908)	(0.00782)	(0.00906)	(0.00778)
D(mem_son)	0.01003	-0.01495*	0.00682	-0.01389	0.01159	-0.01529*	0.00881	-0.01450*
	(0.00993)	(0.00852)	(0.00979)	(0.00847)	(0.00987)	(0.00851)	(0.00989)	(0.00848)
D(mem_oth)	0.01478**	0.00581	0.01484**	0.00603	0.01638**	0.00531	0.01503**	0.00637
	(0.00717)	(0.00588)	(0.00711)	(0.00587)	(0.00712)	(0.00586)	(0.00715)	(0.00587)
P(men)	-0.02220*	0.00677	-0.02008*	0.00668	-0.02111*	0.00644	-0.01997	0.00596
	(0.01233)	(0.01120)	(0.01216)	(0.01115)	(0.01231)	(0.01121)	(0.01230)	(0.01118)
age_head	0.00168	0.00310***	0.00191*	0.00301***	0.00149	0.00316***	0.00189*	0.00299***
	(0.00108)	(0.00097)	(0.00107)	(0.00097)	(0.00109)	(0.00097)	(0.00108)	(0.00097)
age_head-sq	-0.00002	-0.00002*	-0.00002*	-0.00002*	-0.00001	-0.00002*	-0.00002*	-0.00002
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
D(gender_head)	-0.01976**	0.01719**	-0.01782*	0.01667**	-0.02239**	0.01801**	-0.01942**	0.01582*
	(0.00962)	(0.00826)	(0.00939)	(0.00821)	(0.00962)	(0.00829)	(0.00960)	(0.00824)
year_educ_head	-0.00396***	0.00389***	-0.00394***	0.00336***	-0.00333***	0.00376***	-0.00409***	0.00392***
	(0.00084)	(0.00075)	(0.00081)	(0.00074)	(0.00084)	(0.00076)	(0.00084)	(0.00075)
D(Lima)	0.07931***	0.00671	0.06321***	0.01003	0.07649***	0.00724	0.07194***	0.00939
	(0.00665)	(0.00591)	(0.01214)	(0.01018)	(0.00668)	(0.00596)	(0.00679)	(0.00597)
Hrs_for	-0.00017	-0.00003	-0.00018	-0.00002	-0.00018	-0.00002	-0.00019*	-0.00002
	(0.00011)	(0.00010)	(0.00011)	(0.00010)	(0.00011)	(0.00010)	(0.00011)	(0.00010)
Hrs_inf	0.00007	-0.00023***	0.00004	-0.00020***	0.00004	-0.00022***	0.00006	-0.00022***
	(0.00007)	(0.00006)	(0.00007)	(0.00006)	(0.00007)	(0.00006)	(0.00007)	(0.00006)
D(Hrs_for)	0.00823	-0.00644	0.01583*	-0.00865	0.00556	-0.00607	0.00943	-0.00781
	(0.00865)	(0.00774)	(0.00849)	(0.00772)	(0.00870)	(0.00779)	(0.00864)	(0.00770)
D(Hrs_inf)	0.01846***	-0.01631***	0.01787**	-0.01585**	0.01821***	-0.01639***	0.01797**	-0.01507**
	(0.00701)	(0.00624)	(0.00696)	(0.00623)	(0.00698)	(0.00625)	(0.00699)	(0.00622)
L(pop_urb_dist)			0.15939***	-0.05179				
			(0.03319)	(0.03336)				
L(pop_urb_dist)-sq			-0.00632***	0.00201				
			(0.00153)	(0.00152)				
pop_den_distx1000			-0.00083	0.00076				
			(0.00062)	(0.00053)				
D(not_slum)			-0.00043	0.02586***				
			(0.00551)	(0.00504)				
D(mid_city)			-0.02064***	0.00437				
			(0.00637)	(0.00559)				
D(border)			0.05997***	0.00086				
			(0.00830)	(0.00709)				
D(tongue)					0.03219***	-0.00513		
					(0.00668)	(0.00572)		
D(migrant)					0.01525***	-0.00634		
					(0.00549)	(0.00484)		
D(social)					0.00181	-0.00302		
					(0.00524)	(0.00445)		
Station_distx1000							-0.09871***	0.03368**
							(0.01915)	(0.01665)
Muni_persx1000							-0.00108***	0.00199***
							(0.00042)	(0.00055)
L(budget_dist_pc)							0.00642	0.00188
							(0.00423)	(0.00367)
D(sdp_dist)							-0.04394***	-0.01261
							(0.01144)	(0.01092)
Constant	0.64220***	-0.90426***	-0.25013	-0.58620***	0.59457***	-0.89754***	0.59770***	-0.90128***
	(0.07554)	(0.06666)	(0.17628)	(0.18176)	(0.07672)	(0.06819)	(0.07876)	(0.06906)

	(1) Informal	(1) Formal	(2) Informal	(2) Formal	(3) Informal	(3) Formal	(4) Informal	(4) Formal
Observations	4,874	4,874	4,874	4,874	4,874	4,874	4,874	4,874
LogL	1608.341	2285.586	1686.809	2313.037	1630.370	2288.043	1628.093	2310.029

l) Second-stage IV estimation (hours of work under 2SLS)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
Hrs_for	-0.00519***	0.00307*	-0.00269*	0.00278*	-0.00518***	0.00315*	-0.00590***	0.00391**
	(0.00165)	(0.00174)	(0.00158)	(0.00167)	(0.00162)	(0.00172)	(0.00165)	(0.00167)
Hrs_inf	0.00334***	-0.00547***	0.00443***	-0.00540***	0.00330***	-0.00550***	0.00255***	-0.00461***
	(0.00100)	(0.00106)	(0.00095)	(0.00103)	(0.00099)	(0.00106)	(0.00098)	(0.00098)
L(expenditure)	-0.02104	0.10513***	-0.05169***	0.10510***	-0.01990	0.10448***	-0.01068	0.09253***
	(0.01286)	(0.01369)	(0.01236)	(0.01295)	(0.01249)	(0.01342)	(0.01289)	(0.01312)
L(fam_size)	-0.03171	0.11985***	-0.08943***	0.12036***	-0.03139	0.12013***	-0.00292	0.08899**
	(0.03437)	(0.03661)	(0.03172)	(0.03414)	(0.03364)	(0.03610)	(0.03424)	(0.03454)
D(mem_spo)	0.05933***	-0.04715***	0.05781***	-0.04464***	0.05782***	-0.04864***	0.05798***	-0.04485***
	(0.01391)	(0.01464)	(0.01407)	(0.01448)	(0.01394)	(0.01477)	(0.01351)	(0.01352)
D(mem_son)	0.01524	-0.03661**	0.02949*	-0.03684**	0.01634	-0.03571**	0.00577	-0.02659
	(0.01664)	(0.01753)	(0.01629)	(0.01693)	(0.01634)	(0.01738)	(0.01632)	(0.01638)
D(mem_oth)	0.00273	0.03615***	-0.00798	0.03696***	0.00421	0.03612***	0.00869	0.03050**
	(0.01207)	(0.01324)	(0.01237)	(0.01297)	(0.01201)	(0.01331)	(0.01172)	(0.01215)
P(men)	-0.04821**	0.06526***	-0.06643***	0.06634***	-0.04703**	0.06547***	-0.03487*	0.05214**
	(0.02041)	(0.02167)	(0.02022)	(0.02120)	(0.02027)	(0.02172)	(0.02009)	(0.02025)
age_head	-0.00277	0.01107***	-0.00415**	0.01087***	-0.00287	0.01114***	-0.00140	0.00970***
	(0.00192)	(0.00204)	(0.00188)	(0.00196)	(0.00190)	(0.00204)	(0.00188)	(0.00189)
age_head-sq	0.00003	-0.00010***	0.00005**	-0.00010***	0.00003	-0.00010***	0.00002	-0.00008***
	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
D(gender_head)	-0.03911***	0.03977***	-0.03804***	0.03878***	-0.04108***	0.03972***	-0.03664**	0.03593**
	(0.01465)	(0.01516)	(0.01455)	(0.01495)	(0.01451)	(0.01516)	(0.01432)	(0.01407)
year_educ_head	-0.00233	0.00099	-0.00049	0.00072	-0.00188	0.00128	-0.00354**	0.00233
	(0.00164)	(0.00175)	(0.00152)	(0.00161)	(0.00156)	(0.00168)	(0.00161)	(0.00162)
D(Lima)	0.07050***	0.02819***	0.04987***	0.03229*	0.06914***	0.02613***	0.06385***	0.02925***
	(0.00875)	(0.00954)	(0.01711)	(0.01729)	(0.00870)	(0.00955)	(0.00854)	(0.00895)
D(Hrs_for)	0.37055***	-0.29156***	0.25649***	-0.27317***	0.36756***	-0.29650***	0.39432***	-0.32171***
	(0.09800)	(0.10389)	(0.09530)	(0.10125)	(0.09709)	(0.10374)	(0.09740)	(0.09888)
D(Hrs_inf)	-0.19013***	0.29056***	-0.23746***	0.28521***	-0.18890***	0.29202***	-0.15163***	0.24893***
	(0.05568)	(0.05940)	(0.05332)	(0.05798)	(0.05534)	(0.05953)	(0.05482)	(0.05481)
L(pop_urb_dist)			0.17338***	-0.04695				
			(0.04132)	(0.04473)				
L(pop_urb_dist)-sq			-0.00717***	0.00212				
			(0.00195)	(0.00209)				
pop_den_distx1000			-0.00054	0.00049				
			(0.00088)	(0.00089)				
D(not_slum)			0.01709**	0.00940				
			(0.00859)	(0.00907)				
D(mid_city)			-0.02982***	0.01482				
			(0.00911)	(0.00941)				
D(border)			0.04436***	0.02222*				
			(0.01215)	(0.01240)				
D(tongue)					0.02071**	0.00486		
					(0.00946)	(0.01021)		
D(migrant)					0.01204	0.00874		
					(0.00808)	(0.00888)		
D(social)					0.00404	-0.00827		

					(0.00735)	(0.00787)		
Station_distx1000							-0.10534***	0.01811
							(0.02501)	(0.02957)
Muni_persx1000							-0.00118**	0.00246***
							(0.00055)	(0.00068)
L(budget_dist_pc)							0.00443	0.00175
							(0.00561)	(0.00571)
D(sdp_dist)							-0.05331***	-0.00210
							(0.01684)	(0.01863)
Constant	0.61489***	-0.90761***	-0.12201	-0.65989***	0.59148***	-0.91068***	0.49314***	-0.79649***
	(0.12477)	(0.13295)	(0.22485)	(0.24651)	(0.12349)	(0.13306)	(0.12401)	(0.12550)
Observations	4,874	4,874	4,874	4,874	4,874	4,874	4,874	4,874
LogL	107.852	-268.526	-11.446	-193.969	129.851	-290.423	272.774	109.448

m) Second-stage IV estimation (hours of work under LIML)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
Hrs_for	-0.00584***	0.00357	-0.00294*	0.00335	-0.00583***	0.00367*	-0.00677***	0.00446**
	(0.00198)	(0.00225)	(0.00176)	(0.00213)	(0.00193)	(0.00219)	(0.00204)	(0.00210)
Hrs_inf	0.00389***	-0.00654***	0.00483***	-0.00638***	0.00385***	-0.00651***	0.00304**	-0.00541***
	(0.00124)	(0.00142)	(0.00108)	(0.00135)	(0.00123)	(0.00140)	(0.00127)	(0.00127)
L(expenditure)	-0.02150	0.11003***	-0.05330***	0.10932***	-0.02048	0.10899***	-0.00952	0.09532***
	(0.01510)	(0.01709)	(0.01345)	(0.01577)	(0.01455)	(0.01642)	(0.01558)	(0.01601)
L(fam_size)	-0.04063	0.14438***	-0.09799***	0.14199***	-0.04043	0.14306***	-0.00802	0.10586**
	(0.04151)	(0.04725)	(0.03517)	(0.04285)	(0.04044)	(0.04579)	(0.04264)	(0.04339)
D(mem_spo)	0.06197***	-0.05114***	0.05942***	-0.04856***	0.06055***	-0.05280***	0.06076***	-0.04815***
	(0.01521)	(0.01697)	(0.01488)	(0.01669)	(0.01527)	(0.01703)	(0.01489)	(0.01531)
D(mem_son)	0.01713	-0.04293**	0.03166*	-0.04238**	0.01812	-0.04130**	0.00639	-0.03074
	(0.01874)	(0.02096)	(0.01739)	(0.01985)	(0.01832)	(0.02050)	(0.01873)	(0.01915)
D(mem_oth)	0.00069	0.04131***	-0.00987	0.04170***	0.00206	0.04110***	0.00735	0.03412**
	(0.01356)	(0.01584)	(0.01323)	(0.01531)	(0.01349)	(0.01578)	(0.01337)	(0.01421)
P(men)	-0.05277**	0.07610***	-0.07039***	0.07629***	-0.05170**	0.07578***	-0.03811*	0.05990**
	(0.02303)	(0.02595)	(0.02164)	(0.02500)	(0.02282)	(0.02574)	(0.02305)	(0.02367)
age_head	-0.00348	0.01249***	-0.00467**	0.01216***	-0.00358	0.01248***	-0.00201	0.01076***
	(0.00221)	(0.00250)	(0.00203)	(0.00237)	(0.00219)	(0.00247)	(0.00222)	(0.00226)
age_head-sq	0.00004	-0.00012***	0.00005**	-0.00011***	0.00004*	-0.00012***	0.00002	-0.00010***
	(0.00002)	(0.00003)	(0.00002)	(0.00002)	(0.00002)	(0.00003)	(0.00002)	(0.00002)
D(gender_head)	-0.04225***	0.04488**	-0.03997***	0.04350**	-0.04411***	0.04439**	-0.03983**	0.04004**
	(0.01604)	(0.01757)	(0.01538)	(0.01721)	(0.01587)	(0.01744)	(0.01582)	(0.01592)
year_educ_head	-0.00188	-0.00015	-0.00012	-0.00021	-0.00147	0.00030	-0.00323*	0.00153
	(0.00194)	(0.00222)	(0.00166)	(0.00198)	(0.00183)	(0.00209)	(0.00196)	(0.00200)
D(Lima)	0.06957***	0.03006***	0.04893***	0.03461*	0.06837***	0.02750**	0.06293***	0.03060***
	(0.00946)	(0.01090)	(0.01792)	(0.01962)	(0.00938)	(0.01080)	(0.00926)	(0.00998)
D(Hrs_for)	0.42067***	-0.34501**	0.27984***	-0.32879**	0.41807***	-0.35039***	0.45557***	-0.37181***
	(0.11810)	(0.13446)	(0.10678)	(0.13037)	(0.11686)	(0.13255)	(0.12073)	(0.12421)
D(Hrs_inf)	-0.22401***	0.35215***	-0.26070***	0.34267***	-0.22294***	0.35081***	-0.18320***	0.29632***
	(0.06938)	(0.07928)	(0.06060)	(0.07618)	(0.06896)	(0.07845)	(0.07045)	(0.07078)
L(pop_urb_dist)			0.17550***	-0.05237				
			(0.04305)	(0.04967)				
L(pop_urb_dist)-sq			-0.00728***	0.00240				
			(0.00203)	(0.00234)				
pop_den_distx1000			-0.00051	0.00042				
			(0.00092)	(0.00101)				

D(not_slum)			0.01886**	0.00502				
			(0.00917)	(0.01060)				
D(mid_city)			-0.03068***	0.01693				
			(0.00956)	(0.01069)				
D(border)			0.04307***	0.02544*				
			(0.01279)	(0.01418)				
D(tongue)					0.01865*	0.00810		
					(0.01034)	(0.01178)		
D(migrant)					0.01141	0.01082		
					(0.00888)	(0.01025)		
D(social)					0.00439	-0.00901		
					(0.00792)	(0.00892)		
Station_distx1000							-0.10555***	0.01451
							(0.02755)	(0.03320)
Muni_persx1000							-0.00119**	0.00248***
							(0.00059)	(0.00072)
L(budget_dist_pc)							0.00415	0.00173
							(0.00603)	(0.00633)
D(sdp_dist)							-0.05488***	-0.00037
							(0.01844)	(0.02081)
Constant	0.63329***	-0.97678***	-0.10745	-0.69708**	0.61204***	-0.97755***	0.49777***	-0.84093***
	(0.14809)	(0.16802)	(0.23430)	(0.27268)	(0.14582)	(0.16514)	(0.15067)	(0.15373)
Observations	4,874	4,874	4,874	4,874	4,874	4,874	4,874	4,874
LogL	-234.179	-881.185	-234.354	-778.704	-215.719	-877.996	-82.385	-393.124

n) Second-stage IV estimation (hours of work under GMM)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
Hrs_for	-0.00518***	0.00322*	-0.00282*	0.00286*	-0.00517***	0.00326*	-0.00584***	0.00395**
	(0.00164)	(0.00173)	(0.00157)	(0.00166)	(0.00161)	(0.00171)	(0.00165)	(0.00166)
Hrs_inf	0.00342***	-0.00520***	0.00434***	-0.00516***	0.00338***	-0.00525***	0.00270***	-0.00447***
	(0.00099)	(0.00105)	(0.00094)	(0.00102)	(0.00098)	(0.00105)	(0.00098)	(0.00097)
L(expenditure)	-0.02121*	0.10306***	-0.05024***	0.10392***	-0.01998	0.10296***	-0.01163	0.09262***
	(0.01274)	(0.01349)	(0.01224)	(0.01283)	(0.01237)	(0.01323)	(0.01280)	(0.01295)
L(fam_size)	-0.03361	0.11159***	-0.08610***	0.11415***	-0.03331	0.11291***	-0.00704	0.08581**
	(0.03398)	(0.03601)	(0.03125)	(0.03370)	(0.03326)	(0.03561)	(0.03398)	(0.03413)
D(mem_spo)	0.05905***	-0.04619***	0.05800***	-0.04390***	0.05762***	-0.04753***	0.05771***	-0.04467***
	(0.01390)	(0.01462)	(0.01406)	(0.01446)	(0.01393)	(0.01475)	(0.01350)	(0.01350)
D(mem_son)	0.01549	-0.03424**	0.02817*	-0.03513**	0.01647	-0.03373*	0.00682	-0.02600
	(0.01655)	(0.01742)	(0.01620)	(0.01684)	(0.01626)	(0.01729)	(0.01626)	(0.01630)
D(mem_oth)	0.00207	0.03427***	-0.00684	0.03533***	0.00352	0.03438***	0.00738	0.02964**
	(0.01201)	(0.01307)	(0.01229)	(0.01283)	(0.01195)	(0.01317)	(0.01167)	(0.01203)
P(men)	-0.04904**	0.06337***	-0.06491***	0.06488***	-0.04780**	0.06391***	-0.03652*	0.05204***
	(0.02030)	(0.02152)	(0.02007)	(0.02109)	(0.02016)	(0.02159)	(0.02001)	(0.02015)
age_head	-0.00294	0.01044***	-0.00411**	0.01032***	-0.00303	0.01055***	-0.00164	0.00932***
	(0.00190)	(0.00201)	(0.00186)	(0.00194)	(0.00189)	(0.00201)	(0.00187)	(0.00187)
age_head-sq	0.00003	-0.00009***	0.00005**	-0.00009***	0.00003*	-0.00009***	0.00002	-0.00008***
	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
D(gender_head)	-0.03959***	0.03755**	-0.03848***	0.03701**	-0.04158***	0.03764**	-0.03732***	0.03502**
	(0.01464)	(0.01512)	(0.01454)	(0.01491)	(0.01450)	(0.01512)	(0.01431)	(0.01404)
year_educ_head	-0.00223	0.00141	-0.00062	0.00102	-0.00179	0.00163	-0.00333**	0.00254
	(0.00162)	(0.00173)	(0.00150)	(0.00160)	(0.00155)	(0.00167)	(0.00160)	(0.00161)
D(Lima)	0.07095***	0.02854***	0.04971***	0.03393**	0.06957***	0.02675***	0.06411***	0.02962***

	(0.00872)	(0.00950)	(0.01702)	(0.01717)	(0.00867)	(0.00951)	(0.00852)	(0.00893)
D(Hrs_for)	0.37199***	-0.29484***	0.26160***	-0.27336***	0.36833***	-0.29799***	0.39396***	-0.32123***
	(0.09779)	(0.10361)	(0.09491)	(0.10113)	(0.09692)	(0.10351)	(0.09726)	(0.09860)
D(Hrs_inf)	-0.19510***	0.27651***	-0.23332***	0.27289***	-0.19344***	0.27887***	-0.15950***	0.24133***
	(0.05519)	(0.05868)	(0.05279)	(0.05739)	(0.05487)	(0.05892)	(0.05448)	(0.05437)
L(pop_urb_dist)			0.17454***	-0.04500				
			(0.04122)	(0.04467)				
L(pop_urb_dist)-sq			-0.00720***	0.00201				
			(0.00194)	(0.00209)				
pop_den_distx1000			-0.00054	0.00044				
			(0.00087)	(0.00089)				
D(not_slum)			0.01678*	0.01047				
			(0.00857)	(0.00905)				
D(mid_city)			-0.02999***	0.01524				
			(0.00910)	(0.00940)				
D(border)			0.04536***	0.02045*				
			(0.01213)	(0.01235)				
D(tongue)					0.02066**	0.00509		
					(0.00944)	(0.01019)		
D(migrant)					0.01135	0.00696		
					(0.00806)	(0.00884)		
D(social)					0.00452	-0.00766		
					(0.00734)	(0.00786)		
Station_distx1000							-0.10148***	0.02065
							(0.02485)	(0.02941)
Muni_persx1000							-0.00122**	0.00244***
							(0.00055)	(0.00068)
L(budget_dist_pc)							0.00446	0.00212
							(0.00560)	(0.00570)
D(sdp_dist)							-0.05323***	-0.00234
							(0.01684)	(0.01863)
Constant	0.62046***	-0.87748***	-0.14240	-0.64720***	0.59656***	-0.88483***	0.50580***	-0.79153***
	(0.12337)	(0.13057)	(0.22407)	(0.24568)	(0.12212)	(0.13090)	(0.12301)	(0.12362)
Observations	4,874	4,874	4,874	4,874	4,874	4,874	4,874	4,874
LogL	74.277	-137.136	21.735	-72.473	99.654	-166.906	236.391	179.196

o) Second-stage IV estimation (difference of hours of work under C-IV)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
Hrs_formal-Hrs_informal	-0.00447***	0.00521***	-0.00427***	0.00494***	-0.00447***	0.00525***	-0.00415***	0.00503***
	(0.00098)	(0.00102)	(0.00097)	(0.00100)	(0.00099)	(0.00104)	(0.00093)	(0.00101)
L(expenditure)	-0.03291***	0.09189***	-0.04206***	0.09019***	-0.03191***	0.09187***	-0.03133***	0.08959***
	(0.00667)	(0.00687)	(0.00669)	(0.00685)	(0.00678)	(0.00702)	(0.00641)	(0.00671)
L(fam_size)	-0.06976***	0.09211***	-0.06979***	0.08703***	-0.07036***	0.09336***	-0.06340***	0.09003***
	(0.02232)	(0.02319)	(0.02182)	(0.02249)	(0.02248)	(0.02347)	(0.02120)	(0.02269)
D(mem_spo)	0.06173***	-0.04948***	0.05967***	-0.04647***	0.06091***	-0.05050***	0.06035***	-0.04782***
	(0.01398)	(0.01439)	(0.01361)	(0.01393)	(0.01409)	(0.01459)	(0.01341)	(0.01407)
D(mem_son)	0.02636*	-0.02709*	0.02272	-0.02578*	0.02740*	-0.02655*	0.02404*	-0.02586*
	(0.01473)	(0.01517)	(0.01432)	(0.01465)	(0.01469)	(0.01520)	(0.01412)	(0.01479)
D(mem_oth)	-0.00471	0.03132***	-0.00455	0.03092***	-0.00357	0.03145***	-0.00301	0.03110***
	(0.01164)	(0.01199)	(0.01135)	(0.01162)	(0.01171)	(0.01213)	(0.01115)	(0.01170)
P(men)	-0.06305***	0.05658***	-0.06021***	0.05510***	-0.06241***	0.05707***	-0.05789***	0.05402***
	(0.01963)	(0.02024)	(0.01916)	(0.01964)	(0.01971)	(0.02044)	(0.01879)	(0.01978)

age_head	-0.00434**	0.01059***	-0.00386**	0.01016***	-0.00449**	0.01069***	-0.00369**	0.01021***
	(0.00193)	(0.00200)	(0.00188)	(0.00193)	(0.00193)	(0.00201)	(0.00184)	(0.00196)
age_head-sq	0.00005**	-0.00009***	0.00004**	-0.00009***	0.00005***	-0.00009***	0.00004**	-0.00009***
	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
D(gender_head)	-0.04288***	0.04147***	-0.03962***	0.04002***	-0.04472***	0.04155***	-0.04104***	0.03921***
	(0.01494)	(0.01539)	(0.01450)	(0.01485)	(0.01489)	(0.01542)	(0.01433)	(0.01504)
year_educ_head	-0.00068	0.00206	-0.00121	0.00197	-0.00030	0.00228*	-0.00099	0.00219
	(0.00131)	(0.00136)	(0.00124)	(0.00127)	(0.00127)	(0.00132)	(0.00125)	(0.00133)
D(Lima)	0.06839***	0.02750***	0.04984***	0.03181*	0.06765***	0.02594***	0.06180***	0.03014***
	(0.00952)	(0.00979)	(0.01719)	(0.01758)	(0.00956)	(0.00989)	(0.00933)	(0.00975)
D(Hrs_for)	0.35437***	-0.40959***	0.34399***	-0.38790***	0.35307***	-0.41305***	0.32904***	-0.39643***
	(0.07975)	(0.08335)	(0.07865)	(0.08150)	(0.08067)	(0.08467)	(0.07537)	(0.08178)
D(Hrs_inf)	-0.24858***	0.28771***	-0.23706***	0.27188***	-0.24948***	0.28993***	-0.22998***	0.27863***
	(0.06025)	(0.06296)	(0.05926)	(0.06139)	(0.06060)	(0.06359)	(0.05695)	(0.06178)
L(pop_urb_dist)			0.16621***	-0.03533				
			(0.04107)	(0.04198)				
L(pop_urb_dist)-sq			-0.00680***	0.00153				
			(0.00193)	(0.00197)				
pop_den_distx1000			-0.00069	0.00073				
			(0.00092)	(0.00094)				
D(not_slum)			0.01664*	0.01103				
			(0.00899)	(0.00921)				
D(mid_city)			-0.03023***	0.01495				
			(0.00936)	(0.00958)				
D(border)			0.04617***	0.01887				
			(0.01252)	(0.01281)				
D(tongue)					0.01807*	0.00593		
					(0.01000)	(0.01036)		
D(migrant)					0.00780	0.00513		
					(0.00810)	(0.00839)		
D(social)					0.00508	-0.00776		
					(0.00785)	(0.00812)		
Station_distx1000							-0.08556***	0.01990
							(0.03166)	(0.03308)
Muni_persx1000							-0.00126*	0.00246***
							(0.00071)	(0.00074)
L(budget_dist_pc)							0.00600	0.00219
							(0.00582)	(0.00608)
D(sdp_dist)							-0.05392***	-0.00036
							(0.01733)	(0.01812)
Constant	0.74400***	-0.79008***	-0.17140	-0.57990**	0.72626***	-0.79687***	0.69925***	-0.78553***
	(0.07162)	(0.07403)	(0.22081)	(0.22571)	(0.07421)	(0.07716)	(0.07465)	(0.07882)
Observations	4,874	4,874	4,874	4,874	4,874	4,874	4,874	4,874
Wald(overall)	276.63***	643.60***	368.83***	712.37***	295.75***	638.65***	316.49***	690.22***

Appendix 4: Definition of variables and basic tabulations (models in Chapter 6)

a) Model for consumption groups

Variable	Definition	Groups	
		7,253 obs.	
		mean	s.d.
Dependant variables			
Fon	Consumption done in food to be consumed within the household as a share of total expenditure	0.363	0.16157
Foff	Consumption done in food to be consumed outside the household as a share of total expenditure	0.120	0.13079
CC	Consumption done in clothing and personal care as a share of total expenditure	0.088	0.06294
HEA	Consumption done in health goods and services as a share of total expenditure	0.052	0.07438
TC	Consumption done in transportation and communication as a share of total expenditure	0.137	0.09282
ED	Consumption done in entertainment, education and cultural goods and services as a share of total expenditure	0.067	0.07895
OT	Consumption done in other or non-classified goods as a share of total expenditure	0.173	0.09932
D(Fon)	Dummy variable: 1 = household purchase Fon, 0 = otherwise	0.985	0.12112
D(Foff)	Dummy variable: 1 = household purchase Foff, 0 = otherwise	0.865	0.34202
D(CC)	Dummy variable: 1 = household purchase CC, 0 = otherwise	0.990	0.09915
D(HEA)	Dummy variable: 1 = household purchase HEA, 0 = otherwise	0.830	0.37554
D(TC)	Dummy variable: 1 = household purchase TC, 0 = otherwise	0.961	0.19463
D(ED)	Dummy variable: 1 = household purchase ED, 0 = otherwise	0.906	0.29208
D(OT)	Dummy variable: 1 = household purchase OT, 0 = otherwise	0.999	0.03320
Independant variables: SUR model			
L(expenditure)	Natural logarithm of total *per-capita* expenditure (purchased) in annual new *soles*	8.068	0.71693
L(fam_size)	Natural logarithm of the number of family members	1.314	0.55974
D(mem_spo)	Dummy variable: 1 = household has a spouse, 0 = otherwise	0.655	0.47548
D(mem_son)	Dummy variable: 1 = household has a son or daughter, 0 = otherwise	0.811	0.39188
D(mem_oth)	Dummy variable: 1 = household has other member, 0 = otherwise	0.309	0.46199
P(men)	Proportion of men in the household	0.500	0.24141
age_head	Age of the head of the household (in years)	48.660	14.31469
age_head-sq	Age of the head of the household (in years) squared	2572.651	1501.49700
D(gender_head)	Dummy variable: 1 = household head in male, 0 = otherwise	0.742	0.43739
year_educ_head	Number of years of education of the head of the household	10.307	4.82042
Hrs_ff	Sum of the weekly hours of work of all family members, primary activity, sector = formal workers in formal firms (see Chapter 4)	20.797	33.62769
Hrs_if	Sum of the weekly hours of work of all family members, primary activity, sector = informal workers in formal firms (see Chapter 4)	5.403	18.56123
Hrs_ii	Sum of the weekly hours of work of all family members, primary activity, sector = informal workers in informal firms (see Chapter 4)	18.539	35.73597
Hrs_sf	Sum of the weekly hours of work of all family members, primary activity, sector = informal self-employment (see Chapter 4)	7.894	21.96456
Hrs_si	Sum of the weekly hours of work of all family members, primary activity, sector = formal self-employment (see Chapter 4)	29.219	41.04796
Hrs_fam	Sum of the weekly hours of work of all family members, primary activity, sector = family work (see Chapter 4)	8.282	22.38989
Hrs_second	Sum of the weekly hours of work of all family members, secondary activity (see Chapter 4)	6.482	14.50570

D(Hrs_ff)	Dummy variable: 1 = any household member is formal worker in formal firm in his/her primary activity (see Chapter 4), 0 = otherwise	0.344	0.47498
D(Hrs_if)	Dummy variable: 1 = any household member is informal worker in formal firm in his/her primary activity (see Chapter 4), 0 = otherwise	0.102	0.30325
D(Hrs_ii)	Dummy variable: 1 = any household member is informal worker in informal firm in his/her primary activity (see Chapter 4), 0 = otherwise	0.306	0.46101
D(Hrs_sf)	Dummy variable: 1 = any household member is formal self-employer in his/her primary activity (see Chapter 4), 0 = otherwise	0.144	0.35077
D(Hrs_si)	Dummy variable: 1 = any household member is informal self-employer in his/her primary activity (see Chapter 4), 0 = otherwise	0.497	0.50003
D(Hrs_fam)	Dummy variable: 1 = any household member is family worker in his/her primary activity (see Chapter 4), 0 = otherwise	0.205	0.40405
D(Hrs_second)	Dummy variable: 1 = any household member has a second job (see Chapter 4), 0 = otherwise	0.282	0.44998
D(Lima)	Dummy variable: 1 = household lives in Lima, 0 = otherwise	0.150	0.35710
Independant variables: PROBIT model			
D(ca_food)_1	Dummy variable: 1 = the household considers it has an adequate consumption of food, 0 = otherwise	0.740	0.43876
D(ca_food)_2	Dummy variable: 1 = the household considers it has an adequate consumption of tubers and cereals, 0 = otherwise	0.629	0.48318
D(ca_food)_3	Dummy variable: 1 = the household considers it has an adequate consumption of fruits and vegetables, 0 = otherwise	0.551	0.49742
D(ca_food)_4	Dummy variable: 1 = the household considers it has an adequate consumption of meat, fish and chicken, 0 = otherwise	0.547	0.49780
Agropec_firmPC	Number of agropecuarian firms per 1,000 habs (province level)	0.489	1.31330
D(agro_dist)	Dummy variable: 1 = agrarian activity is important in the district, 0 = otherwise (declared by the mayor)	0.590	0.49185
D(pec_dist)	Dummy variable: 1 = pecuarian activity is important in the district, 0 = otherwise (declared by the mayor)	0.298	0.45720
D(fish_dist)	Dummy variable: 1 = fishing activity is important in the district, 0 = otherwise (declared by the mayor)	0.110	0.31260
Prop(workagro_dist)	Proportion of urban workers in agropecuarian sector in the district	0.097	0.13243
Prop(workfish_dist)	Proportion of urban workers in fishing sector in the district	0.010	0.02806
Hotel_firmPC	Number of tourism (hotel and restaurants) firms per 1,000 habs (province level)	3.547	1.34299
D(rest_dist)	Dummy variable: 1 = restaurant activity is important in the district, 0 = otherwise (declared by the mayor)	0.765	0.42373
D(tour_dist)	Dummy variable: 1 = tourism activity is important in the district, 0 = otherwise (declared by the mayor)	0.223	0.41652
Prop(workrest_dist)	Proportion of urban workers in restaurant and hotel sector in the district	0.058	0.01435
Com_firmPC	Number of commercial firms per 1,000 habs (province level)	23.024	6.84234
Prop(workcom_dist)	Proportion of urban workers in commercial sector in the district	0.228	0.05515
D(ca_cloth)_1	Dummy variable: 1 = the household considers it has an adequate consumption of clothes, 0 = otherwise	0.393	0.48835
D(ca_cloth)_2	Dummy variable: 1 = the household considers it has an adequate consumption of shoes, 0 = otherwise	0.377	0.48476
D(ca_health)_1	Dummy variable: 1 = the household considers it has an adequate consumption of medicines, 0 = otherwise	0.422	0.49398
D(ca_health)_2	Dummy variable: 1 = the household considers it has an adequate consumption of soaps, 0 = otherwise	0.506	0.50000
n_chronic	Number family members with a chronic illness	1.063	1.11382
Health1_distx1000	Number of health centres per 1,000 habitants (district level)	0.697	0.69784
Health2_distx1000	Number of hospitals per 1,000 habitants (district level)	0.026	0.04061
Health3_distx1000	Number of health micro-centres per 1,000 habitants (district level)	0.098	0.23481

Health4_distx1000	Number of other health providers per 1,000 habitants (district level)	0.526	0.59335
Health_firmPC	Number of health firms per 1,000 habitants (province level)	0.612	0.35445
Prop(workhealth_dist)	Proportion of urban workers in health sector in the district	0.028	0.01200
D(ca_transport)	Dummy variable: 1 = the household considers it has an adequate consumption of transport, 0 = otherwise	0.522	0.49957
comm1_distx1000	Number of communal phones per 1,000 habitants (district level)	0.170	0.30509
comm2_distx1000	Number of telephone centrals per 1,000 habitants (district level)	0.024	0.09046
comm3_distx1000	Number of post offices per 1,000 habitants (district level)	0.020	0.04793
comm4_distx1000	Number of satellite dishes per 1,000 habitants (district level)	0.068	0.20115
comm5_distx1000	Number of internet cabins per 1,000 habitants (district level)	0.818	1.26340
comm6_distx1000	Number of other communication infrastructures per 1,000 habitants (district level)	0.014	0.06966
D(comm_dist)	Dummy variable: 1 = if the district has other transportation and communication infrastructure , 0 = otherwise (declared by the mayor)	0.665	0.47218
D(transp_dist)	Dummy variable: 1 = transportation activity is important in the district, 0 = otherwise (declared by the mayor)	0.912	0.28386
transp_distx1000	Number of transport lines per 1,000 habitants (district level)	0.304	0.36089
D(good_roads_dist)	Dummy variable: 1 =the roads in the district are good, 0 = otherwise (declared by the mayor)	0.927	0.25937
Transport_firmPC	Number of transport firms per 1,000 habitants (province level)	0.799	0.41888
D(TC_dist)	Dummy variable: 1 = there are workers on transport and communication in the district, 0 = otherwise (declared by the mayor)	0.143	0.34993
Prop(worktransp_dist)	Proportion of urban workers in transport and communication sector in the district	0.102	0.02749
Commun_firmPC	Number of communication firms per 1,000 habitants (province level)	1.754	0.71898
D(ca_educ)	Dummy variable: 1 = the household considers it has an adequate consumption of education, 0 = otherwise	0.455	0.49797
D(ca_entert)	Dummy variable: 1 = the household considers it has an adequate consumption of entertainment, 0 = otherwise	0.132	0.33860
school_distx1000	Number of school rooms per 1,000 habitants (district level)	4.736	1.84052
Educa_firmPC	Number of educational firms per 1,000 habitants (province level)	0.523	0.31554
Entert_firmPC	Number of entertainment firms per 1,000 habitants (province level)	0.244	0.09157
Prop(workeduca_dist)	Proportion of urban workers in educational sector in the district	0.082	0.03701

b) Models for markets within consumption groups (1)

Variable	**Definition**	**Fon**		**Foff**		**CC**	
		7,145 obs.		**6,272 obs.**		**7,181 obs.**	
		Mean	**s.d.**	**Mean**	**s.d.**	**Mean**	**s.d.**
Dependent variables							
Fon_inf	Consumption of Fon done at informal markets as a share of total Fon expenditure	0.603	0.28492				
Fon_semi	Consumption of Fon done at semi-formal markets as a share of total Fon expenditure	0.332	0.27992				
Fon_for	Consumption of Fon done at formal markets as a share of total Fon expenditure	0.065	0.14366				
D(Fon_inf)	Dummy variable: 1 = household purchase from Fon_inf, 0 = otherwise	0.970	0.17162				

D(Fon_semi)	Dummy variable: 1 = household purchase from Fon_semi, 0 = otherwise	0.934	0.24914				
D(Fon_for)	Dummy variable: 1 = household purchase from Fon_for, 0 = otherwise	0.467	0.49893				
Foff_inf	Consumption of Foff done at informal markets as a share of total Foff expenditure			0.279	0.37052		
Foff_semi	Consumption of Foff done at semi-formal markets as a share of total Foff expenditure			0.196	0.33898		
Foff_for	Consumption of Foff done at formal markets as a share of total Foff expenditure			0.525	0.41753		
D(Foff_inf)	Dummy variable: 1 = household purchase from Foff_inf, 0 = otherwise			0.588	0.49232		
D(Foff_semi)	Dummy variable: 1 = household purchase from Foff_semi, 0 = otherwise			0.299	0.45783		
D(Foff_for)	Dummy variable: 1 = household purchase from Foff_for, 0 = otherwise			0.766	0.42354		
CC_inf	Consumption of CC done at informal markets as a share of total CC expenditure					0.417	0.32832
CC_semi	Consumption of CC done at semi-formal markets as a share of total CC expenditure					0.515	0.32139
CC_for	Consumption of CC done at formal markets as a share of total CC expenditure					0.068	0.15302
D(CC_inf)	Dummy variable: 1 = household purchase from CC_inf, 0 = otherwise					0.823	0.38169
D(CC_semi)	Dummy variable: 1 = household purchase from CC_semi, 0 = otherwise					0.954	0.20849
D(CC_for)	Dummy variable: 1 = household purchase from CC_for, 0 = otherwise					0.300	0.45845
Independent variables: SUR model							
L(expend_Fon)	Natural logarithm of total *per-capita* expenditure (purchased) in annual new *soles* (Fon)	6.935	0.74619				
L(expenditure_Foff)	Natural logarithm of total *per-capita* expenditure (purchased) in annual new *soles* (Foff)			5.649	1.33060		
L(expenditure_CC)	Natural logarithm of total *per-capita* expenditure (purchased) in annual new *soles* (CC)					5.422	1.03924
L(fam_size)	Natural logarithm of the number of family members	1.326	0.54964	1.349	0.54910	1.321	0.55516
D(mem_spo)	Dummy variable: 1 = household has a spouse, 0 = otherwise	0.660	0.47359	0.667	0.47135	0.658	0.47428
D(mem_son)	Dummy variable: 1 = household has a son or daughter, 0 = otherwise	0.818	0.38560	0.830	0.37535	0.817	0.38679
D(mem_oth)	Dummy variable: 1 = household has other member, 0 = otherwise	0.311	0.46310	0.319	0.46608	0.309	0.46194
P(men)	Proportion of men in the household	0.498	0.23776	0.509	0.23685	0.501	0.23962
age_head	Age of the head of the household (in years)	48.704	14.24073	48.036	13.72804	48.518	14.19388
age_head-sq	Age of the head of the household (in years) squared	2574.889	1493.76600	2495.844	1417.64800	2555.411	1484.79200

D(gender_head)	Dummy variable: 1 = household head in male, 0 = otherwise	0.742	0.43730	0.759	0.42796	0.745	0.43580
year_educ_head	Number of years of education of the head of the household	10.297	4.81120	10.511	4.76424	10.357	4.79832
Hrs_ff	Sum of the weekly hours of work of all family members, primary activity, sector = formal workers in formal firms (see Chapter 4)	20.955	33.73765	22.757	34.76091	20.987	33.72882
Hrs_if	Sum of the weekly hours of work of all family members, primary activity, sector = informal workers in formal firms (see Chapter 4)	5.420	18.55995	5.905	19.50229	5.457	18.64613
Hrs_ii	Sum of the weekly hours of work of all family members, primary activity, sector = informal workers in informal firms (see Chapter 4)	18.698	35.89226	20.198	37.19842	18.644	35.83561
Hrs_sf	Sum of the weekly hours of work of all family members, primary activity, sector = informal self-employment (see Chapter 4)	7.925	22.00129	7.903	22.04935	7.901	21.98871
Hrs_si	Sum of the weekly hours of work of all family members, primary activity, sector = formal self-employment (see Chapter 4)	29.268	41.03178	30.672	42.14880	29.323	41.11847
Hrs_fam	Sum of the weekly hours of work of all family members, primary activity, sector = family work (see Chapter 4)	8.306	22.38653	8.177	21.77709	8.331	22.47213
Hrs_second	Sum of the weekly hours of work of all family members, secondary activity (see Chapter 4)	6.507	14.50371	6.888	15.02152	6.519	14.54816
D(Hrs_ff)	Dummy variable: 1 = any household member is formal worker in formal firm in his/her primary activity (see Chapter 4), 0 = otherwise	0.346	0.47572	0.373	0.48358	0.347	0.47597
D(Hrs_if)	Dummy variable: 1 = any household member is informal worker in formal firm in his/her primary activity (see Chapter 4), 0 = otherwise	0.103	0.30417	0.110	0.31313	0.103	0.30459
D(Hrs_ii)	Dummy variable: 1 = any household member is informal worker in informal firm in his/her primary activity (see Chapter 4), 0 = otherwise	0.309	0.46201	0.328	0.46945	0.307	0.46137
D(Hrs_sf)	Dummy variable: 1 = any household member is formal self-employer in his/her primary activity (see Chapter 4), 0 = otherwise	0.144	0.35141	0.142	0.34947	0.144	0.35082
D(Hrs_si)	Dummy variable: 1 = any household member is informal self-employer in his/her primary activity (see Chapter 4), 0 = otherwise	0.499	0.50003	0.509	0.49995	0.498	0.50003
D(Hrs_fam)	Dummy variable: 1 = any household member is family worker in his/her primary activity (see Chapter 4), 0 = otherwise	0.206	0.40478	0.207	0.40492	0.206	0.40473
D(Hrs_second)	Dummy variable: 1 = any household member has a second job (see Chapter 4), 0 = otherwise	0.284	0.45089	0.295	0.45592	0.283	0.45067
D(Lima)	Dummy variable: 1 = household lives in Lima, 0 = otherwise	0.151	0.35836	0.149	0.35635	0.150	0.35721
Independent variables: PROBIT model							
D(ca_food)_1	Dummy variable: 1 = the household considers it has an adequate consumption of food, 0 = otherwise	0.740	0.43861				
D(ca_food)_2	Dummy variable: 1 = the household considers it has an adequate consumption of tubers and cereals, 0 = otherwise	0.629	0.48304				
D(ca_food)_3	Dummy variable: 1 = the household considers it has an adequate consumption of fruits and vegetables, 0 = otherwise	0.552	0.49729				
D(ca_food)_4	Dummy variable: 1 = the household considers it has an adequate consumption of meat, fish and chicken, 0 = otherwise	0.548	0.49776				
Agropec_firmPC	Number of agropecuarian firms per 1,000 habitants (province level)	0.488	1.31486				
D(agro_dist)	Dummy variable: 1 = agrarian activity is important in the district, 0 = otherwise (declared by the mayor)	0.588	0.49231				
D(pec_dist)	Dummy variable: 1 = pecuarian activity is important in the district, 0 = otherwise (declared by the mayor)	0.295	0.45628				

D(fish_dist)	Dummy variable: 1 = fishing activity is important in the district, 0 = otherwise (declared by the mayor)	0.110	0.31344				
Prop(workagro_dist)	Proportion of urban workers in agropecuarian sector in the district	0.096	0.13200				
Prop(workfish_dist)	Proportion of urban workers in fishing sector in the district	0.010	0.02794				
Hotel_firmPC	Number of tourism (hotel and restaurants) firms per 1,000 habitants (province level)			3.605	1.34239		
D(rest_dist)	Dummy variable: 1 = restaurant activity is important in the district, 0 = otherwise (declared by the mayor)			0.776	0.41675		
D(tour_dist)	Dummy variable: 1 = tourism activity is important in the district, 0 = otherwise (declared by the mayor)			0.223	0.41612		
Prop(workrest_dist)	Proportion of urban workers in restaurant and hotel sector in the district			0.059	0.01391		
Com_firmPC	Number of commercial firms per 1,000 habitants (province level)					23.062	6.82462
Prop(workcom_dist)	Proportion of urban workers in commercial sector in the district					0.229	0.05489
D(ca_cloth)_1	Dummy variable: 1 = the household considers it has an adequate consumption of clothes, 0 = otherwise					0.395	0.48884
D(ca_cloth)_2	Dummy variable: 1 = the household considers it has an adequate consumption of shoes, 0 = otherwise					0.380	0.48532
P(pop_slums_dist)	Proportion of people in the district living in a slum	0.328	0.33782	0.337	0.33937	0.330	0.33815
P(migrant_dist)	Proportion of migrants living in the district (urban places)	0.312	0.10376	0.315	0.10145	0.312	0.10365
P(race_dist)	Proportion of indigenous people (by declared race) living in the district (urban places)	0.089	0.11848	0.089	0.11391	0.088	0.11710
informal_index	Index of tax informality at the province level, varies from 1 (high informality) to 0 (no informality) Index = (Tax of the IVA - (IVA payments/Value Added))/Tax of the IVA	0.283	0.30138	0.277	0.30031	0.282	0.30125
D(conf_gov_1)	Dummy variable: 1 = the head of the household considers that the public administration performance is good or very good , 0 = otherwise	0.310	0.46236	0.304	0.46018	0.309	0.46218
D(conf_gov_2)	Dummy variable: 1 = the head of the household considers that the performance of the government is not an obstacle for development , 0 = otherwise	0.192	0.39403	0.184	0.38777	0.191	0.39294
D(conf_gov_3)	Dummy variable: 1 = if the head of the household has a positive opinion in all responses related to the government , 0 = otherwise	0.083	0.27653	0.080	0.27113	0.083	0.27569
D(overall_satis)	Dummy variable: 1 = the head of the household considers his/her family as non-poor, 0 = otherwise	0.615	0.48657	0.603	0.48938	0.613	0.48700
D(income_satis)	Dummy variable: 1 = the head of the household is satisfied with his/her income, 0 = otherwise	0.355	0.47853	0.343	0.47463	0.355	0.47866
D(income_stab)	Dummy variable: 1 = the head of the household considers that his/her income is stable, 0 = otherwise	0.782	0.41324	0.779	0.41462	0.781	0.41382
D(sub_pov)	Dummy variable: 1 = if income is below the minimum income question, 0 = otherwise	0.320	0.46665	0.309	0.46218	0.323	0.46779
D(road_1)	Dummy variable: 1 = main road of the district is a non-asphalted highway, 0 = otherwise	0.178	0.38280	0.173	0.37799	0.179	0.38310
D(road_2)	Dummy variable: 1 = main road of the district is an asphalted highway, 0 = otherwise	0.670	0.47025	0.674	0.46863	0.669	0.47076
D(road_3)	Dummy variable: 1 = main road of the district is any other type, 0 = otherwise	0.029	0.16734	0.025	0.15672	0.029	0.16772

distance	Distance from the district to the capital of the province in minutes	75.658	125.83730	73.845	125.09920	75.689	126.41520
P(nowater_dist)	Proportion of households without water service in the district (urban places)	0.208	0.19994	0.204	0.19364	0.208	0.19871
P(nodrain_dist)	Proportion of houses without drain service in the district (urban places)	0.280	0.21226	0.273	0.20610	0.280	0.21203
P(noelec_dist)	Proportion of houses without electricity in the district (urban places)	0.113	0.08178	0.110	0.07881	0.113	0.08189
P(nofloor_dist)	Proportion of houses with floor of inadequate materials in the district (urban places)	0.300	0.16361	0.295	0.15776	0.300	0.16256
P(nowall_dist)	Proportion of houses with walls of inadequate materials in the district (urban places)	0.422	0.23411	0.416	0.23027	0.422	0.23367
P(noinsu_dist)	Proportion of people without insurance in the district (urban places)	0.578	0.09535	0.579	0.09476	0.578	0.09507
P(illiterate_dist)	Proportion of illiterate people in the district (urban places)	0.038	0.02871	0.037	0.02661	0.038	0.02830
P(selfwork_dist)	Proportion of self-employed in the district (urban places)	0.383	0.07291	0.382	0.07164	0.383	0.07271
P(unreg_dist)	Proportion of unregistered people in the district (urban places)	0.974	0.01438	0.974	0.01387	0.974	0.01439
P(undernut_dist)	Proportion of under-nourished children below 5 years old in the district	0.179	0.12251	0.175	0.11854	0.179	0.12237
Mort_distx1000	Child mortality, number per 1,000 born alive	17.031	6.37148	16.975	6.27833	17.033	6.34784
P(poor_dist)	Proportion of poor people in the district	0.294	0.16316	0.288	0.15676	0.293	0.16227
icf_index	Targeting index (between 0 and 1) of the government for social programs (district level)	0.130	0.16463	0.123	0.15474	0.129	0.16373
lifeexp_dist	Average life expectancy of the district	73.733	2.06017	73.747	2.03231	73.733	2.05428

c) Models for markets within consumption groups (2)

Variable	Definition	HEA		TC		ED	
		6,021 obs.		6,967 obs.		6,570 obs.	
		Mean	**s.d.**	**Mean**	**s.d.**	**Mean**	**s.d.**
Dependent variables							
HEA_inf	Consumption of HEA done at informal markets as a share of total HEA expenditure	0.071	0.20316				
HEA_semi	Consumption of HEA done at semi-formal markets as a share of total HEA expenditure	0.154	0.30664				
HEA_for	Consumption of HEA done at formal markets as a share of total HEA expenditure	0.776	0.34955				
D(HEA_inf)	Dummy variable: 1 = household purchases from HEA_inf, 0 = otherwise	0.199	0.39951				
D(HEA_semi)	Dummy variable: 1 = household purchases from HEA_semi, 0 = otherwise	0.384	0.48632				
D(HEA_for)	Dummy variable: 1 = household purchases from HEA_for, 0 = otherwise	0.892	0.31098				
TC_inf	Consumption of TC done at informal markets as a share of total TC expenditure			0.455	0.31597		
TC_semi	Consumption of TC done at semi-formal markets as a share of total TC expenditure			0.099	0.16694		
TC_for	Consumption of TC done at formal markets as a share of total TC expenditure			0.445	0.32183		
D(TC_inf)	Dummy variable: 1 = household purchases from TC_inf, 0 = otherwise			0.890	0.31320		
D(TC_semi)	Dummy variable: 1 = household purchases from TC_semi, 0 = otherwise			0.647	0.47801		

D(TC_for)	Dummy variable: 1 = household purchases from TC_for, 0 = otherwise			0.837	0.36944		
ED_inf	Consumption of ED done at informal markets as a share of total ED expenditure					0.279	0.33212
ED_semi	Consumption of ED done at semi-formal markets as a share of total ED expenditure					0.247	0.27189
ED_for	Consumption of ED done at formal markets as a share of total ED expenditure					0.474	0.32686
D(ED_inf)	Dummy variable: 1 = household purchases from ED_inf, 0 = otherwise					0.751	0.43274
D(ED_semi)	Dummy variable: 1 = household purchases from ED_semi, 0 = otherwise					0.711	0.45335
D(ED_for)	Dummy variable: 1 = household purchases from ED_for, 0 = otherwise					0.877	0.32844
Independent variables: SUR model							
L(expenditure_HEA)	Natural logarithm of total *per-capita* expenditure (purchased) in annual new *soles* (HEA)	4.612	1.57623				
L(expenditure_TC)	Natural logarithm of total *per-capita* expenditure (purchased) in annual new *soles* (TC)			5.912	1.20291		
L(expenditure_ED)	Natural logarithm of total *per-capita* expenditure (purchased) in annual new *soles* (ED)					4.941	1.36778
L(fam_size)	Natural logarithm of the number of family members	1.380	0.52056	1.329	0.54909	1.385	0.50933
D(mem_spo)	Dummy variable: 1 = household has a spouse, 0 = otherwise	0.681	0.46602	0.661	0.47351	0.684	0.46494
D(mem_son)	Dummy variable: 1 = household has a son or daughter, 0 = otherwise	0.846	0.36110	0.821	0.38337	0.861	0.34562
D(mem_oth)	Dummy variable: 1 = household has other member, 0 = otherwise	0.333	0.47144	0.313	0.46371	0.319	0.46631
P(men)	Proportion of men in the household	0.494	0.22703	0.499	0.23799	0.505	0.22667
age_head	Age of the head of the household (in years)	48.750	13.89760	48.564	14.13041	47.770	13.72667
age_head-sq	Age of the head of the household (in years) squared	2569.678	1458.76500	2558.138	1478.30200	2470.396	1421.74600
D(gender_head)	Dummy variable: 1 = household head is male, 0 = otherwise	0.747	0.43464	0.744	0.43673	0.760	0.42732
year_educ_head	Number of years of education of the head of the household	10.454	4.77763	10.457	4.77208	10.587	4.73053
Hrs_ff	Sum of the weekly hours of work of all family members, primary activity, sector = formal workers in formal firms (see Chapter 4)	21.966	34.61431	21.553	34.01283	22.274	34.43539
Hrs_if	Sum of the weekly hours of work of all family members, primary activity, sector = informal workers in formal firms (see Chapter 4)	5.849	19.34720	5.507	18.73742	5.617	18.80602
Hrs_ii	Sum of the weekly hours of work of all family members, primary activity, sector = informal workers in informal firms (see Chapter 4)	19.460	36.46896	18.911	36.16941	18.973	36.33052
Hrs_sf	Sum of the weekly hours of work of all family members, primary activity, sector = informal self-employment (see Chapter 4)	8.324	22.49613	7.962	22.13226	8.129	22.39012
Hrs_si	Sum of the weekly hours of work of all family members, primary activity, sector = formal self-employment (see Chapter 4)	30.394	42.17762	29.414	41.43377	29.985	41.62005
Hrs_fam	Sum of the weekly hours of work of all family members, primary activity, sector = family work (see Chapter 4)	8.727	23.13826	8.330	22.53500	8.737	22.88572
Hrs_second	Sum of the weekly hours of work of all family members, secondary activity (see Chapter 4)	6.872	14.94628	6.585	14.65703	6.812	14.87236
D(Hrs_ff)	Dummy variable: 1 = any household member is formal worker in formal firm in his/her primary activity (see Chapter 4), 0 = otherwise	0.359	0.47972	0.356	0.47884	0.367	0.48202
D(Hrs_if)	Dummy variable: 1 = any household member is informal worker in formal firm in his/her primary activity (see Chapter 4), 0 = otherwise	0.111	0.31409	0.104	0.30592	0.107	0.30953
D(Hrs_ii)	Dummy variable: 1 = any household member is informal worker in informal firm in his/her primary activity (see Chapter 4), 0 = otherwise	0.319	0.46602	0.310	0.46248	0.311	0.46298
D(Hrs_sf)	Dummy variable: 1 = any household member is formal self-employer in his/her primary activity (see Chapter 4), 0 = otherwise	0.151	0.35773	0.144	0.35093	0.146	0.35325
D(Hrs_si)	Dummy variable: 1 = any household member is informal self-employer in his/her primary activity (see Chapter 4), 0 = otherwise	0.508	0.49997	0.495	0.50002	0.506	0.50000
D(Hrs_fam)	Dummy variable: 1 = any household member is family worker in his/her primary activity (see Chapter 4), 0 = otherwise	0.214	0.41010	0.206	0.40454	0.216	0.41185
D(Hrs_second)	Dummy variable: 1 = any household member has a second job (see Chapter 4), 0 = otherwise	0.296	0.45644	0.285	0.45141	0.294	0.45566
D(Lima)	Dummy variable: 1 = household lives in Lima, 0 = otherwise	0.157	0.36378	0.153	0.36016	0.150	0.35717
Independent variables: PROBIT model							

D(ca_health)_1	Dummy variable: 1 = the household considers it has an adequate consumption of medicines, 0 = otherwise	0.430	0.49509				
D(ca_health)_2	Dummy variable: 1 = the household considers it has an adequate consumption of soaps, 0 = otherwise	0.519	0.49969				
n_chronic	Number family members with a chronic illness	1.152	1.14623				
Health1_distx1000	Number of health centres per 1,000 habitants (district level)	0.698	0.70087				
Health2_distx1000	Number of hospitals per 1,000 habitants (district level)	0.026	0.03945				
Health3_distx1000	Number of health micro-centres per 1,000 habitants (district level)	0.092	0.22241				
Health4_distx1000	Number of other health providers per 1,000 habitants (district level)	0.534	0.60490				
Health_firmPC	Number of health firms per 1,000 habitants (province level)	0.619	0.35202				
Prop(workhealth_dist)	Proportion of urban workers in health sector in the district	0.028	0.01191				
D(ca_transport)	Dummy variable: 1 = the household considers it has an adequate consumption of transport, 0 = otherwise			0.535	0.49879		
comm1_distx1000	Number of communal phones per 1,000 habitants (district level)			0.163	0.29462		
comm2_distx1000	Number of telephone centrals per 1,000 habitants (district level)			0.023	0.08987		
comm3_distx1000	Number of post offices per 1,000 habitants (district level)			0.019	0.04563		
comm4_distx1000	Number of satellite dishes per 1,000 habitants (district level)			0.066	0.20047		
comm5_distx1000	Number of internet cabins per 1,000 habitants (district level)			0.827	1.27892		
comm6_distx1000	Number of other communication infrastructures per 1,000 habitants (district level)			0.013	0.06567		
D(comm_dist)	Dummy variable: 1 = if the district has other transportation and communication infrastructure , 0 = otherwise (declared by the mayor)			0.669	0.47055		
D(transp_dist)	Dummy variable: 1 = transportation activity is important in the district, 0 = otherwise (declared by the mayor)			0.915	0.27886		
transp_distx1000	Number of transport lines per 1,000 habitants (district level)			0.305	0.36144		
D(good_roads_dist)	Dummy variable: 1 =the roads in the district are good, 0 = otherwise (declared by the mayor)			0.930	0.25548		
Transport_firmPC	Number of transport firms per 1,000 habitants (province level)			0.809	0.41754		
D(TC_dist)	Dummy variable: 1 = there are workers on transport and communication in the district, 0 = otherwise (declared by the mayor)			0.143	0.35050		
Prop(worktransp_dist)	Proportion of urban workers in transport and communication sector in the district			0.103	0.02633		
Commun_firmPC	Number of communication firms per 1,000 habitants (province level)			1.771	0.71247		
D(ca_educ)	Dummy variable: 1 = the household considers it has an adequate consumption of education, 0 = otherwise					0.491	0.49996
D(ca_entert)	Dummy variable: 1 = the household considers it has an adequate consumption of entertainment, 0 = otherwise					0.137	0.34386
school_distx1000	Number of school rooms per 1,000 habitants (district level)					4.712	1.80304
Educa_firmPC	Number of educational firms per 1,000 habitants (province level)					0.526	0.31555
Entert_firmPC	Number of entertainment firms per 1,000 habitants (province level)					0.245	0.09117
Prop(workeduca_dist)	Proportion of urban workers in educational sector in the district					0.082	0.03669
P(pop_slums_dist)	Proportion of people in the district living in a slum	0.335	0.33681	0.332	0.33827	0.333	0.33820
P(migrant_dist)	Proportion of migrants living in the district (urban places)	0.313	0.10123	0.314	0.10180	0.312	0.10298
P(race_dist)	Proportion of indigenous people (by declared race) living in the district (urban places)	0.083	0.11215	0.087	0.11327	0.088	0.11696
informal_index	Index of tax informality at the province level, varies from 1 (high informality) to 0 (no informality) Index = (Tax of the IVA - (IVA payments/Value Added))/Tax of the IVA	0.269	0.29506	0.278	0.29922	0.280	0.30005
D(conf_gov_1)	Dummy variable: 1 = the head of the household considers that the public administration performance is good or very good , 0 = otherwise	0.306	0.46091	0.307	0.46135	0.309	0.46230
D(conf_gov_2)	Dummy variable: 1 = the head of the household considers that the performance of the government is not an obstacle for development , 0 = otherwise	0.183	0.38699	0.185	0.38810	0.185	0.38827

D(conf_gov_3)	Dummy variable: 1 = if the head of the household has a positive opinion in all responses related to the government , 0 = otherwise	0.079	0.26985	0.081	0.27234	0.080	0.27141
D(overall_satis)	Dummy variable: 1 = the head of the household considers his/her family as non-poor, 0 = otherwise	0.605	0.48895	0.605	0.48886	0.603	0.48937
D(income_satis)	Dummy variable: 1 = the head of the household is satisfied with his//her income, 0 = otherwise	0.354	0.47828	0.349	0.47680	0.353	0.47779
D(income_stab)	Dummy variable: 1 = the head of the household considers that his/her income is stable, 0 = otherwise	0.782	0.41320	0.777	0.41613	0.780	0.41403
D(sub_pov)	Dummy variable: 1 = if income is below the minimum income question, 0 = otherwise	0.299	0.45769	0.316	0.46514	0.317	0.46542
D(road_1)	Dummy variable: 1 = main road of the district is a non-asphalted highway, 0 = otherwise	0.168	0.37397	0.173	0.37787	0.179	0.38312
D(road_2)	Dummy variable: 1 = main road of the district is an asphalted highway, 0 = otherwise	0.681	0.46608	0.675	0.46834	0.670	0.47008
D(road_3)	Dummy variable: 1 = main road of the district is any other type, 0 = otherwise	0.027	0.16182	0.028	0.16618	0.030	0.17098
distance	Distance from the district to the capital of the province in minutes	73.609	124.44530	74.018	123.01220	75.455	125.66960
P(nowater_dist)	Proportion of households without water service in the district (urban places)	0.206	0.19478	0.204	0.19478	0.208	0.19716
P(nodrain_dist)	Proportion of houses without drain service in the district (urban places)	0.276	0.20856	0.275	0.20715	0.279	0.21082
P(noelec_dist)	Proportion of houses without electricity in the district (urban places)	0.111	0.07961	0.111	0.07844	0.113	0.08202
P(nofloor_dist)	Proportion of houses with floor of inadequate materials in the district (urban places)	0.295	0.15916	0.296	0.15855	0.299	0.16205
P(nowall_dist)	Proportion of houses with walls of inadequate materials in the district (urban places)	0.417	0.23084	0.417	0.23148	0.420	0.23271
P(noinsu_dist)	Proportion of people without insurance in the district (urban places)	0.579	0.09344	0.578	0.09441	0.578	0.09438
P(illiterate_dist)	Proportion of illiterate people in the district (urban places)	0.037	0.02667	0.037	0.02653	0.038	0.02811
P(selfwork_dist)	Proportion of self-employers in the district (urban places)	0.382	0.07047	0.381	0.07199	0.382	0.07225
P(unreg_dist)	Proportion of unregistered people in the district (urban places)	0.974	0.01411	0.974	0.01404	0.974	0.01410
P(undernut_dist)	Proportion of under-nourished children below 5 years old in the district	0.176	0.11902	0.176	0.11999	0.179	0.12207
Mort_distx1000	Child mortality, number per 1,000 born alive	16.863	6.24006	16.923	6.30119	17.067	6.40444
P(poor_dist)	Proportion of poor people in the district	0.289	0.15769	0.289	0.15889	0.292	0.16167
icf_index	Targeting index (between 0 and 1) of the government for social programs (district level)	0.124	0.15791	0.124	0.15705	0.128	0.16156
lifeexp_dist	Average life expectancy of the district	73.785	2.02323	73.768	2.04196	73.727	2.07170

Appendix 5: Full regression results (models in Chapter 6)

a) PROBIT for consumption groups

	D(Fon)	D(Foff)	D(CC)	D(HEA)	D(TC)	D(ED)
L(expenditure)	0.52061***	0.69847***	0.90564***	0.52624***	0.89699***	0.44086***
	(0.08090)	(0.04344)	(0.11204)	(0.03819)	(0.06636)	(0.04926)
L(fam_size)	1.08460***	0.58821***	0.81506***	0.79104***	0.81035***	1.41548***
	(0.23774)	(0.08424)	(0.17941)	(0.07833)	(0.13562)	(0.11366)
D(mem_spo)	0.05625	-0.49868***	-0.09311	-0.03524	0.03287	-0.37282***
	(0.16574)	(0.08591)	(0.22523)	(0.07626)	(0.15228)	(0.10702)
D(mem_son)	0.17887	-0.16428*	0.59261***	-0.03831	-0.03046	0.38365***
	(0.21576)	(0.08411)	(0.21761)	(0.07684)	(0.13142)	(0.10046)
D(mem_oth)	0.01597	-0.02618	-0.24251	0.07916	0.10801	-0.20389***
	(0.19840)	(0.06366)	(0.16495)	(0.05995)	(0.11261)	(0.07885)
P(men)	-0.23213	0.23459**	0.48787	-0.17470*	-0.02004	0.36110***
	(0.18920)	(0.10071)	(0.31122)	(0.09181)	(0.17940)	(0.13245)
age_head	0.02891*	0.00720	0.01920	0.00962	0.02564*	-0.01808*
	(0.01596)	(0.00873)	(0.01828)	(0.00778)	(0.01340)	(0.01042)
age_head-sq	-0.00028*	-0.00019**	-0.00033**	-0.00011	-0.00022*	0.00002
	(0.00016)	(0.00008)	(0.00016)	(0.00008)	(0.00013)	(0.00010)
D(gender_head)	-0.03338	0.44099***	0.07756	-0.06591	-0.35036**	0.25598**
	(0.13980)	(0.08563)	(0.25712)	(0.07786)	(0.16166)	(0.11417)
year_educ_head	-0.05087***	-0.02301***	0.00777	0.01153**	0.03700***	0.02244***
	(0.01330)	(0.00567)	(0.01466)	(0.00524)	(0.00999)	(0.00697)
Hrs_ff	-0.00468	0.00394**	0.00623	-0.00089	-0.00215	-0.00451**
	(0.00295)	(0.00173)	(0.00774)	(0.00122)	(0.00489)	(0.00201)
Hrs_if	-0.01237***	0.01248***		-0.00174	-0.00368	-0.00834***
	(0.00314)	(0.00351)		(0.00203)	(0.00354)	(0.00244)
Hrs_ii	-0.00456***	0.00513***	0.00604	-0.00140	0.00393*	-0.00177
	(0.00160)	(0.00146)	(0.00478)	(0.00102)	(0.00218)	(0.00124)
Hrs_sf	-0.00667	-0.00223	-0.01281***	-0.00195	-0.00538	0.00157
	(0.00497)	(0.00221)	(0.00420)	(0.00201)	(0.00335)	(0.00262)
Hrs_si	-0.00422**	0.00549***	-0.00174	0.00121	0.00456***	-0.00191*
	(0.00198)	(0.00102)	(0.00301)	(0.00076)	(0.00137)	(0.00102)
Hrs_fam	-0.00339	-0.00332**	0.00566	0.00092	0.00134	-0.00243
	(0.00292)	(0.00147)	(0.00375)	(0.00139)	(0.00237)	(0.00224)
Hrs_second	-0.00797*	0.00730***	0.00601	0.00055	0.00786*	0.00625*
	(0.00443)	(0.00258)	(0.00921)	(0.00212)	(0.00443)	(0.00325)
D(Hrs_ff)	0.12858	0.34688***	-0.50985	-0.06680	0.60333**	0.23636*
	(0.20111)	(0.10813)	(0.45326)	(0.08365)	(0.29482)	(0.13771)
D(Hrs_if)	0.55166**	-0.14763		0.20298	0.23909	0.27328
	(0.27598)	(0.17993)		(0.13016)	(0.24099)	(0.17462)
D(Hrs_ii)	0.37509**	0.40973***	-0.52180**	0.19520***	0.07408	-0.15264
	(0.16138)	(0.09053)	(0.23377)	(0.07370)	(0.13463)	(0.09443)
D(Hrs_sf)	0.34694	0.33319**	0.40825	0.26839**	0.39355*	-0.16874
	(0.32441)	(0.13309)	(0.36243)	(0.12728)	(0.20794)	(0.16132)
D(Hrs_si)	0.09239	0.14633**	-0.12990	0.01546	-0.28798***	-0.00086
	(0.16413)	(0.06816)	(0.21813)	(0.05843)	(0.10282)	(0.08338)
D(Hrs_fam)	-0.22097	0.17410**	0.03211	-0.05383	0.09755	0.16364
	(0.17785)	(0.08551)	(0.21628)	(0.07469)	(0.13097)	(0.12261)
D(Hrs_second)	0.25881	-0.08248	0.10468	0.04733	-0.05680	-0.00248
	(0.17366)	(0.07449)	(0.26407)	(0.06470)	(0.11922)	(0.09587)
D(Lima)	0.11073	-0.27119***	-0.41491**	0.08609	0.07224	-0.14835

	(0.18467)	(0.06088)	(0.19207)	(0.05793)	(0.16180)	(0.09212)
D(ca_food)_1	-0.04541					
	(0.14421)					
D(ca_food)_2	-0.01748					
	(0.12457)					
D(ca_food)_3	0.21109*					
	(0.12728)					
D(ca_food)_4	-0.17106					
	(0.13443)					
Agropec_firmPC	0.00032					
	(0.02582)					
D(agro_dist)	-0.15155					
	(0.12371)					
D(pec_dist)	-0.00735					
	(0.10517)					
D(fish_dist)	0.49692**					
	(0.21894)					
Prop(workagro_dist)	-0.24235					
	(0.31854)					
Prop(workfish_dist)	-4.53789***					
	(1.35840)					
Hotel_firmPC		0.05773***				
		(0.01982)				
D(rest_dist)		0.07081				
		(0.05229)				
D(tour_dist)		-0.18134***				
		(0.05208)				
Prop(workrest_dist)		6.19586***				
		(1.87182)				
Com_firmPC			0.00287			
			(0.01039)			
Prop(workcom_dist)			2.03269**			
			(0.94270)			
D(ca_cloth)_1			-0.21381			
			(0.24818)			
D(ca_cloth)_2			0.16059			
			(0.26181)			
D(ca_health)_1				-0.16271***		
				(0.05105)		
D(ca_health)_2				0.03938		
				(0.05059)		
n_chronic				0.22926***		
				(0.02345)		
Health1_distx1000				-0.49065		
				(0.31091)		
Health2_distx1000				0.15624		
				(0.58177)		
Health3_distx1000				0.37626		
				(0.34413)		
Health4_distx1000				0.60433*		
				(0.32122)		
Health_firmPC				0.00121		
				(0.05888)		
Prop(workhealth_dist)				-7.40630***		

				(1.73438)		
D(ca_transport)					0.24479***	
					(0.08106)	
comm1_distx1000					-0.13124	
					(0.11137)	
comm2_distx1000					0.04017	
					(0.34286)	
comm3_distx1000					-2.62139***	
					(0.55440)	
comm4_distx1000					0.14933	
					(0.21787)	
comm5_distx1000					0.05075	
					(0.03540)	
comm6_distx1000					-0.63236	
					(0.40835)	
D(comm_dist)					0.08191	
					(0.07870)	
D(transp_dist)					0.14888	
					(0.12774)	
transp_distx1000					0.19422	
					(0.13118)	
D(good_roads_dist)					-0.08045	
					(0.12682)	
Transport_firmPC					-0.20621	
					(0.16900)	
D(TC_dist)					-0.14870	
					(0.10174)	
Prop(worktransp_dist)					0.58087	
					(1.44796)	
Commun_firmPC					0.20324***	
					(0.07888)	
D(ca_educ)						0.56772***
						(0.06795)
D(ca_entert)						0.07634
						(0.09375)
school_distx1000						0.00158
						(0.01708)
Educa_firmPC						0.14586
						(0.14177)
Entert_firmPC						-0.31799
						(0.33271)
Prop(workeduca_dist)						1.55043**
						(0.77032)
Constant	-2.83153***	-5.93074***	-6.15729***	-4.41544***	-7.44405***	-3.55692***
	(0.68246)	(0.38857)	(0.90524)	(0.33608)	(0.59006)	(0.48626)
Observations	7,253	7,253	7,253	7,253	7,253	7,253
LogL	-404.675	-2220.828	-211.496	-2807.751	-764.385	-1400.052
Pseudo-R2	0.279	0.227	0.476	0.150	0.366	0.381
Wald (overall)	369.818***	861.934***	236.714***	820.379***	606.535***	1304.011***
Wald (indent. vars.)	19.86**	53.39***	5.79	131.78***	67.28***	78.83***

b) SUR for consumption groups

	Fon	Foff	CC	HEA	TC	ED
L(expenditure)	-0.08205***	-0.02567***	0.00796***	0.00515*	0.03033***	0.02868***
	(0.00434)	(0.00416)	(0.00170)	(0.00303)	(0.00277)	(0.00235)
L(fam_size)	0.03725***	-0.09061***	0.00220	-0.01909***	-0.00086	0.04649***
	(0.00702)	(0.00658)	(0.00283)	(0.00473)	(0.00396)	(0.00471)
D(mem_spo)	0.06054***	-0.04215***	-0.01467***	0.01834***	-0.00517	-0.01844***
	(0.00637)	(0.00653)	(0.00276)	(0.00388)	(0.00372)	(0.00350)
D(mem_son)	0.02327***	-0.03037***	0.00539*	-0.00063	-0.01642***	0.01648***
	(0.00731)	(0.00689)	(0.00308)	(0.00447)	(0.00439)	(0.00446)
D(mem_oth)	-0.00500	0.00435	-0.00352*	0.01126***	0.00078	-0.00633**
	(0.00448)	(0.00402)	(0.00196)	(0.00276)	(0.00282)	(0.00263)
P(men)	-0.01837**	0.05703***	-0.00042	-0.00931*	0.00512	-0.02244***
	(0.00892)	(0.00847)	(0.00368)	(0.00528)	(0.00532)	(0.00517)
age_head	-0.00015	-0.00039	-0.00162***	0.00037	0.00181***	0.00114**
	(0.00085)	(0.00081)	(0.00035)	(0.00050)	(0.00050)	(0.00047)
age_head-sq	0.00001	-0.00000	0.00001***	0.00000	-0.00001**	-0.00001**
	(0.00001)	(0.00001)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
D(gender_head)	-0.03023***	0.02463***	0.01103***	-0.01046**	0.00244	0.00240
	(0.00660)	(0.00683)	(0.00280)	(0.00410)	(0.00387)	(0.00372)
year_educ_head	-0.00305***	-0.00182***	-0.00005	-0.00034	0.00141***	0.00296***
	(0.00046)	(0.00042)	(0.00021)	(0.00028)	(0.00030)	(0.00028)
Hrs_ff	-0.00052***	0.00069***	0.00001	-0.00001	0.00013**	-0.00017***
	(0.00007)	(0.00007)	(0.00004)	(0.00005)	(0.00005)	(0.00006)
Hrs_if	-0.00077***	0.00093***	0.00009	-0.00005	0.00017*	-0.00019*
	(0.00014)	(0.00013)	(0.00009)	(0.00008)	(0.00010)	(0.00010)
Hrs_ii	-0.00064***	0.00100***	0.00005	0.00002	0.00010**	-0.00035***
	(0.00007)	(0.00007)	(0.00003)	(0.00004)	(0.00004)	(0.00003)
Hrs_sf	-0.00068***	0.00011	-0.00005	0.00015	0.00013	-0.00001
	(0.00016)	(0.00014)	(0.00008)	(0.00010)	(0.00011)	(0.00010)
Hrs_si	-0.00061***	0.00057***	0.00014***	-0.00004	0.00001	-0.00009***
	(0.00006)	(0.00006)	(0.00003)	(0.00003)	(0.00004)	(0.00003)
Hrs_fam	-0.00049***	0.00027***	0.00010**	0.00002	0.00019***	-0.00004
	(0.00012)	(0.00011)	(0.00005)	(0.00006)	(0.00007)	(0.00006)
Hrs_second	-0.00060***	-0.00000	0.00029***	-0.00010	0.00031***	-0.00004
	(0.00017)	(0.00014)	(0.00008)	(0.00009)	(0.00010)	(0.00008)
D(Hrs_ff)	-0.02244***	0.00129	0.00817***	-0.00894**	0.01647***	0.00901**
	(0.00558)	(0.00540)	(0.00287)	(0.00362)	(0.00413)	(0.00420)
D(Hrs_if)	0.01017	-0.00659	-0.00613	-0.00906*	0.00756	-0.00027
	(0.00860)	(0.00786)	(0.00493)	(0.00526)	(0.00615)	(0.00627)
D(Hrs_ii)	-0.00661	0.00829	-0.00298	-0.01060***	0.00119	0.00901***
	(0.00559)	(0.00533)	(0.00260)	(0.00335)	(0.00334)	(0.00301)
D(Hrs_sf)	-0.00454	0.01187	0.00846*	-0.00572	0.00575	-0.00069
	(0.00996)	(0.00890)	(0.00480)	(0.00611)	(0.00676)	(0.00601)
D(Hrs_si)	0.00490	-0.00229	-0.00483**	0.00190	0.00103	-0.00409
	(0.00487)	(0.00469)	(0.00214)	(0.00290)	(0.00305)	(0.00272)
D(Hrs_fam)	-0.00805	-0.00026	-0.00136	-0.00072	0.00150	0.01098***
	(0.00639)	(0.00543)	(0.00274)	(0.00348)	(0.00384)	(0.00345)
D(Hrs_second)	-0.00194	-0.00210	-0.00212	0.00355	0.00284	0.00195
	(0.00511)	(0.00451)	(0.00244)	(0.00303)	(0.00326)	(0.00292)
D(Lima)	0.03407***	-0.00239	-0.02898***	-0.00062	0.00749***	-0.00571**
	(0.00491)	(0.00404)	(0.00162)	(0.00280)	(0.00288)	(0.00287)
PHI	0.20134***	-0.03038*	0.10260***	-0.03036***	0.03529**	0.01777*

	(0.05094)	(0.01556)	(0.02660)	(0.01102)	(0.01550)	(0.01014)
Constant	1.03831***	0.46284***	0.06957***	0.03642	-0.18485***	-0.26387***
	(0.04245)	(0.04317)	(0.01680)	(0.03200)	(0.02641)	(0.02444)
Observations	7,253					
LogL	48858.641					

c) PROBIT for markets within groups

	Fon		Foff		CC		HEA		TC		ED	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expend_Fon)	0.57678***	0.54711***										
	(0.04131)	(0.03205)										
L(expend_Foff)			0.03803**	0.19543***								
			(0.01529)	(0.01682)								
L(expend_CC)					0.40686***	0.36147***						
					(0.02279)	(0.02160)						
L(expend_HEA)							0.21079***	0.50681***				
							(0.01452)	(0.02024)				
L(expend_TC)									0.37532***	0.77407***		
									(0.02412)	(0.02813)		
L(expend_ED)											0.04890***	0.73460***
											(0.01629)	(0.03922)
L(fam_size)	0.53882***	0.14501**	0.37251***	0.37216***	0.72708***	0.43976***	0.38808***	0.78177***	0.44288***	0.87911***	0.25441***	2.10373***
	(0.14028)	(0.06002)	(0.06348)	(0.06932)	(0.07260)	(0.06277)	(0.07278)	(0.09525)	(0.08360)	(0.08511)	(0.06424)	(0.15101)
D(mem_spo)	0.08585	-0.09365	-0.09390	-0.06989	-0.01843	0.03910	-0.12981*	-0.08395	0.20817**	-0.27679***	0.04160	-0.07820
	(0.13592)	(0.06211)	(0.06375)	(0.07092)	(0.07390)	(0.06547)	(0.07604)	(0.09522)	(0.08451)	(0.08640)	(0.06637)	(0.12716)
D(mem_son)	0.16947	0.05835	-0.20955***	0.01068	0.03610	0.12542*	0.08682	-0.00542	0.01540	-0.02174	-0.39602***	0.63278***
	(0.13954)	(0.06599)	(0.06834)	(0.07613)	(0.07324)	(0.07008)	(0.08244)	(0.10565)	(0.08936)	(0.09103)	(0.07718)	(0.12236)
D(mem_oth)	-0.00502	0.02852	-0.01179	0.03539	-0.17553***	0.02309	0.02622	-0.02173	0.06496	-0.07219	-0.11302**	0.08304
	(0.12155)	(0.04588)	(0.04803)	(0.05379)	(0.05626)	(0.04803)	(0.05318)	(0.07378)	(0.06946)	(0.06564)	(0.04934)	(0.10694)
P(men)	-0.36097**	0.06347	0.06230	0.16023*	-0.22688**	-0.27269***	-0.15582	-0.28325**	-0.06132	-0.24893**	0.27975***	-0.18032
	(0.17417)	(0.08110)	(0.08393)	(0.09257)	(0.09596)	(0.08465)	(0.09959)	(0.12576)	(0.11113)	(0.10828)	(0.08914)	(0.16649)
age_head	0.01583	-0.00391	0.00014	0.02784***	0.00169	-0.01058	0.00776	-0.02284**	0.00793	0.00970	0.02014**	-0.03615***
	(0.01496)	(0.00756)	(0.00752)	(0.00843)	(0.00796)	(0.00762)	(0.00945)	(0.01145)	(0.01007)	(0.00977)	(0.00820)	(0.01334)
age_head-sq	-0.00007	0.00007	-0.00005	-0.00021**	-0.00008	0.00014*	-0.00008	0.00017	-0.00007	-0.00006	-0.00011	0.00001
	(0.00015)	(0.00007)	(0.00007)	(0.00008)	(0.00008)	(0.00007)	(0.00009)	(0.00011)	(0.00010)	(0.00009)	(0.00008)	(0.00013)
D(gender_head)	-0.00770	0.12598*	0.07983	0.01243	0.06183	-0.16910**	0.06178	0.09614	-0.09514	0.17973**	0.16842**	-0.18812
	(0.14252)	(0.06538)	(0.06738)	(0.07492)	(0.07849)	(0.06874)	(0.07923)	(0.09786)	(0.09054)	(0.08649)	(0.06832)	(0.13736)
year_educ_head	-0.01820*	0.03389***	-0.02272***	0.03940***	-0.02722***	0.04452***	0.00378	-0.00176	-0.01390**	0.04713***	0.01005**	-0.03100***
	(0.01044)	(0.00442)	(0.00453)	(0.00500)	(0.00518)	(0.00470)	(0.00530)	(0.00740)	(0.00634)	(0.00633)	(0.00497)	(0.00879)
Hrs_ff	0.00065	0.00221**	0.00217**	0.00363***	-0.00157	0.00147	-0.00003	-0.00114	0.00154	0.00017	-0.00020	-0.00491**

	(0.00249)	(0.00094)	(0.00096)	(0.00132)	(0.00111)	(0.00090)	(0.00102)	(0.00155)	(0.00156)	(0.00176)	(0.00101)	(0.00237)
Hrs_if	-0.00296	-0.00152	0.00129	0.00065	-0.00294	-0.00130	-0.00075	0.00138	-0.00075	-0.00083	0.00329*	-0.00673**
	(0.00322)	(0.00167)	(0.00177)	(0.00184)	(0.00199)	(0.00186)	(0.00191)	(0.00235)	(0.00250)	(0.00226)	(0.00174)	(0.00295)
Hrs_ii	-0.00076	-0.00140*	0.00132*	-0.00176**	-0.00015	-0.00069	0.00015	0.00083	0.00263**	-0.00211**	-0.00098	-0.00348**
	(0.00164)	(0.00073)	(0.00077)	(0.00077)	(0.00097)	(0.00077)	(0.00084)	(0.00122)	(0.00102)	(0.00090)	(0.00079)	(0.00147)
Hrs_sf	0.00472	0.00164	-0.00041	0.00176	-0.00303*	-0.00103	-0.00147	0.00505*	-0.00536***	0.00102	0.00371**	0.00253
	(0.00487)	(0.00159)	(0.00166)	(0.00201)	(0.00173)	(0.00157)	(0.00167)	(0.00270)	(0.00194)	(0.00255)	(0.00180)	(0.00316)
Hrs_si	0.00174	0.00104*	0.00338***	-0.00116*	0.00043	-0.00044	0.00054	0.00047	-0.00170**	-0.00012	0.00284***	-0.00318***
	(0.00162)	(0.00060)	(0.00067)	(0.00063)	(0.00078)	(0.00062)	(0.00067)	(0.00094)	(0.00082)	(0.00078)	(0.00067)	(0.00115)
Hrs_fam	-0.00577***	-0.00096	0.00184	0.00041	-0.00057	-0.00189	-0.00031	-0.00353**	0.00028	0.00083	0.00032	-0.00804***
	(0.00210)	(0.00112)	(0.00127)	(0.00137)	(0.00137)	(0.00115)	(0.00114)	(0.00145)	(0.00146)	(0.00153)	(0.00122)	(0.00193)
Hrs_second	0.00723	0.00021	0.00246	0.00106	-0.00423**	-0.00265	-0.00068	0.00410	-0.00217	0.00406*	0.00200	-0.00390
	(0.00521)	(0.00162)	(0.00172)	(0.00178)	(0.00197)	(0.00165)	(0.00176)	(0.00284)	(0.00220)	(0.00232)	(0.00170)	(0.00321)
D(Hrs_ff)	0.11001	-0.04926	-0.15256**	0.01590	-0.10675	-0.03251	-0.01712	-0.00310	0.07470	-0.02326	0.08240	0.00113
	(0.17754)	(0.06828)	(0.06865)	(0.08650)	(0.08057)	(0.06776)	(0.07688)	(0.11485)	(0.10520)	(0.11388)	(0.07360)	(0.15196)
D(Hrs_if)	0.09048	0.07559	0.02329	0.02496	0.07451	-0.04425	0.11290	-0.07638	0.16700	-0.15899	-0.26787**	-0.00218
	(0.20631)	(0.10372)	(0.10798)	(0.11547)	(0.12622)	(0.11115)	(0.11653)	(0.14817)	(0.15027)	(0.13986)	(0.10527)	(0.19501)
D(Hrs_ii)	0.05747	0.04829	-0.03984	-0.07226	-0.04014	-0.09824	0.02745	0.01985	0.00977	-0.03002	0.08047	-0.10023
	(0.12616)	(0.05680)	(0.05861)	(0.06145)	(0.06949)	(0.06038)	(0.06635)	(0.09293)	(0.07765)	(0.07432)	(0.06213)	(0.11877)
D(Hrs_sf)	-0.16443	0.07941	0.08352	0.11900	-0.03259	0.02755	0.27747**	-0.24879	0.21863*	0.06686	-0.11166	-0.34425*
	(0.27550)	(0.10079)	(0.10669)	(0.12689)	(0.11342)	(0.10250)	(0.10831)	(0.16263)	(0.13176)	(0.16081)	(0.11198)	(0.20333)
D(Hrs_si)	0.07959	-0.07493	0.24549***	0.00292	0.04337	-0.09421*	0.01411	0.13380*	0.10989	-0.03568	-0.01538	-0.05618
	(0.12190)	(0.04861)	(0.05142)	(0.05509)	(0.05820)	(0.05077)	(0.05665)	(0.07969)	(0.06996)	(0.06936)	(0.05352)	(0.09986)
D(Hrs_fam)	0.09443	0.00240	0.06051	-0.01080	0.04513	-0.02728	0.00767	0.09133	0.13941*	-0.19127**	-0.00558	0.20574
	(0.15335)	(0.06178)	(0.06736)	(0.07299)	(0.07827)	(0.06486)	(0.06820)	(0.09443)	(0.08286)	(0.08373)	(0.06652)	(0.13154)
D(Hrs_second)	0.10066	0.06716	-0.00222	0.04968	0.16609***	0.10486*	0.06855	-0.08839	0.12504*	-0.03416	-0.05483	0.07476
	(0.14208)	(0.05203)	(0.05480)	(0.05936)	(0.06395)	(0.05354)	(0.05943)	(0.08520)	(0.07411)	(0.07351)	(0.05601)	(0.10964)
D(Lima)	-0.64649***	0.63497***	0.01997	0.11248	-0.04760	0.59147***	-0.13410*	-0.11653	-0.26503**	0.06870	-0.17259**	0.12539
	(0.14692)	(0.06814)	(0.07004)	(0.07759)	(0.07864)	(0.07117)	(0.07937)	(0.10681)	(0.12312)	(0.12739)	(0.07564)	(0.14085)
P(pop_slums_dist)	-0.01926	0.07539	-0.03661	0.06154	0.09562	-0.01855	-0.04743	-0.10909	0.01244	-0.07266	-0.13504**	0.10547
	(0.13323)	(0.05923)	(0.06216)	(0.06851)	(0.06669)	(0.06077)	(0.07047)	(0.09081)	(0.09128)	(0.08689)	(0.06180)	(0.10950)
P(migrant_dist)	0.10588	-0.26504	-0.94421***	0.78673***	-0.57126**	1.06470***	0.23424	0.58936	-0.53901*	0.96047***	-1.06281***	0.13568
	(0.52576)	(0.25319)	(0.27183)	(0.30318)	(0.28072)	(0.26796)	(0.30636)	(0.38424)	(0.32574)	(0.36143)	(0.26495)	(0.51268)
P(race_dist)	0.50669	-0.20853	1.52876***	-0.45823	1.37996***	-0.62942**	1.14813***	-1.34521***	0.82596***	0.19191	1.38603***	0.38369
	(0.44698)	(0.25382)	(0.27993)	(0.29103)	(0.27799)	(0.27167)	(0.29352)	(0.34915)	(0.29017)	(0.35226)	(0.26061)	(0.49720)

informal_index	0.00021	-0.11798*	0.12792*	-0.37229***	-0.16659**	0.02162	-0.07861	-0.00790	0.31836***	-0.05488	-0.18205***	0.09012
	(0.15280)	(0.07144)	(0.07659)	(0.08536)	(0.07728)	(0.07470)	(0.08492)	(0.10827)	(0.09996)	(0.10455)	(0.07031)	(0.13225)
D(conf_gov_1)	0.03412	-0.05348	0.03709	0.05773	-0.07499	-0.01469	-0.06441	0.03419	0.04229	0.04139	-0.11937***	-0.06975
	(0.09743)	(0.04031)	(0.04216)	(0.04772)	(0.04746)	(0.04222)	(0.04919)	(0.06433)	(0.05739)	(0.05748)	(0.04413)	(0.08024)
D(conf_gov_2)	-0.06951	-0.13375**	0.06961	0.00640	0.00124	-0.03481	-0.01071	-0.05370	-0.01046	-0.08812	-0.11307*	0.24409**
	(0.11314)	(0.05564)	(0.05755)	(0.06308)	(0.06323)	(0.05991)	(0.06750)	(0.08768)	(0.07398)	(0.07403)	(0.05895)	(0.11159)
D(conf_gov_3)	0.11272	0.13577	-0.02969	-0.14741	0.08081	0.07604	0.19493*	-0.00852	-0.00620	-0.04082	0.05425	-0.09865
	(0.18190)	(0.08653)	(0.09032)	(0.09733)	(0.09737)	(0.09120)	(0.10276)	(0.13439)	(0.11463)	(0.11151)	(0.09018)	(0.17455)
D(overall_satis)	0.09992	-0.11632***	0.04705	-0.08701**	0.11857***	-0.10480***	0.03209	0.08730	0.06625	-0.11808**	-0.05468	0.16516**
	(0.08548)	(0.03748)	(0.03875)	(0.04400)	(0.04452)	(0.03891)	(0.04538)	(0.06177)	(0.05290)	(0.05691)	(0.04197)	(0.07560)
D(income_satis)	0.09328	-0.07579**	0.06190	-0.07853*	0.04364	-0.08282**	0.01582	0.04761	0.09412*	-0.04810	-0.18022***	0.26645***
	(0.08803)	(0.03806)	(0.03876)	(0.04170)	(0.04444)	(0.03996)	(0.04503)	(0.05781)	(0.05083)	(0.05001)	(0.04021)	(0.07800)
D(income_stab)	0.01965	-0.13720***	0.13253***	-0.24995***	0.10684**	-0.00120	-0.06983	0.08452	-0.08189	0.04764	-0.05367	0.00965
	(0.10099)	(0.04276)	(0.04415)	(0.05420)	(0.04882)	(0.04458)	(0.05096)	(0.07027)	(0.06361)	(0.06821)	(0.04861)	(0.08892)
D(sub_pov)	0.00613	-0.07608**	-0.00464	-0.02307	0.04168	-0.07747**	0.07936*	0.03795	0.00544	-0.11088**	-0.07229*	0.10938
	(0.07982)	(0.03683)	(0.03849)	(0.04258)	(0.04236)	(0.03920)	(0.04521)	(0.05755)	(0.05096)	(0.04972)	(0.04020)	(0.07613)
D(road_1)	-0.19164	-0.04943	-0.12215*	0.00459	0.10957	0.08891	0.06768	0.02433	-0.10511	0.08708	0.33161***	0.20270
	(0.13152)	(0.06474)	(0.06601)	(0.07467)	(0.06949)	(0.06868)	(0.07809)	(0.10266)	(0.08420)	(0.09066)	(0.06662)	(0.14220)
D(road_2)	-0.20660	0.12750**	0.00389	-0.06444	-0.02301	0.16877***	0.04273	-0.02161	0.10073	0.06257	0.14541**	0.09166
	(0.13310)	(0.05861)	(0.05999)	(0.06779)	(0.06184)	(0.06242)	(0.07190)	(0.09029)	(0.07936)	(0.08336)	(0.06357)	(0.13198)
D(road_3)	0.07782	0.09523	0.16257	-0.06925	0.08137	-0.03103	-0.19424	0.11426	0.38046**	-0.18299	0.16255	-0.11611
	(0.29151)	(0.13260)	(0.13957)	(0.14586)	(0.14231)	(0.15499)	(0.17266)	(0.21272)	(0.17768)	(0.17515)	(0.13259)	(0.23723)
distance	-0.00072**	0.00000	-0.00017	-0.00039**	-0.00054***	-0.00069***	-0.00048**	0.00047*	-0.00054**	0.00014	-0.00021	-0.00046
	(0.00029)	(0.00019)	(0.00018)	(0.00019)	(0.00020)	(0.00024)	(0.00024)	(0.00029)	(0.00025)	(0.00030)	(0.00019)	(0.00028)
P(nowater_dist)	0.46438*	-0.09334	0.21216	-0.04669	0.14186	-0.25805	-0.08030	-0.29116	-0.22437	0.49711**	-0.13360	0.49558*
	(0.25546)	(0.14382)	(0.15139)	(0.16119)	(0.15526)	(0.15917)	(0.17549)	(0.22249)	(0.17425)	(0.19488)	(0.14778)	(0.29575)
P(nodrain_dist)	-0.94802***	-0.36022*	-0.28560	0.20152	0.04284	-0.29329	0.09349	0.36729	0.21719	-0.85834***	0.30060	-0.39150
	(0.35902)	(0.20333)	(0.19772)	(0.21669)	(0.21024)	(0.21179)	(0.22507)	(0.28722)	(0.23475)	(0.24205)	(0.19868)	(0.37192)
P(noelec_dist)	2.08140***	-1.53627***	0.02980	-0.58840	1.12733**	0.21179	-0.71216	0.44110	-0.80537	-0.22851	-0.03031	-0.30679
	(0.76864)	(0.43369)	(0.43453)	(0.45664)	(0.47439)	(0.47011)	(0.49490)	(0.58899)	(0.50002)	(0.52709)	(0.41449)	(0.73652)
P(nofloor_dist)	0.14305	0.27361	0.68361***	-0.87088***	1.21486***	-0.31759	0.32299	-0.05629	0.51782*	0.06323	0.31469	-0.21796
	(0.53080)	(0.24829)	(0.25885)	(0.28304)	(0.25265)	(0.26640)	(0.29710)	(0.36878)	(0.30934)	(0.34171)	(0.25077)	(0.44539)
P(nowall_dist)	0.26490	0.53784***	-0.23721	0.44298***	-0.56647***	0.16864	-0.20931	-0.24305	-0.36592*	-0.24442	0.08709	0.51421*
	(0.35490)	(0.14495)	(0.15341)	(0.16997)	(0.15822)	(0.15230)	(0.17768)	(0.21914)	(0.19027)	(0.19314)	(0.15253)	(0.29533)
P(noinsu_dist)	0.85548	-0.31202	0.08562	-0.29431	0.64543**	-1.37420***	0.20498	1.33418***	0.33692	-0.99919***	0.65117**	-0.05661

	(0.54973)	(0.26911)	(0.26455)	(0.28864)	(0.31888)	(0.30652)	(0.33288)	(0.39755)	(0.35022)	(0.37356)	(0.27480)	(0.47962)
P(illiterate_dist)	-4.03362	1.57066	-2.75280**	-0.33339	-6.49851***	10.86513***	-3.53556**	5.54417***	0.17177	4.72671***	-3.66117***	-5.13211**
	(2.53951)	(1.43001)	(1.37049)	(1.42593)	(1.35178)	(1.33270)	(1.39665)	(1.82768)	(1.50538)	(1.79706)	(1.29435)	(2.25370)
P(selfwork_dist)	0.88699	0.09053	0.33797	0.75589	1.04196**	-0.20968	0.46764	-0.24091	-0.92390*	0.71481	0.18041	0.84232
	(0.86826)	(0.45291)	(0.44021)	(0.46790)	(0.50624)	(0.50486)	(0.51266)	(0.60005)	(0.53083)	(0.55683)	(0.44792)	(0.77864)
P(unreg_dist)	8.02098**	-6.16399***	-0.29771	0.95537	10.27742***	-1.52350	0.77133	7.47605***	-2.76201	3.23073	7.43306***	0.80149
	(3.82354)	(2.08350)	(2.27549)	(2.50941)	(2.28664)	(2.26369)	(2.48487)	(2.87945)	(2.84131)	(3.13127)	(2.27070)	(3.87793)
icf_index	-0.93053*	0.32601	-0.93361***	0.61160**	-0.12060	-1.19358***	0.41600	0.11260	0.66672*	0.15037	-0.17009	-0.07731
	(0.48940)	(0.26730)	(0.26144)	(0.28689)	(0.27338)	(0.30002)	(0.29543)	(0.36690)	(0.34825)	(0.38212)	(0.26868)	(0.51826)
P(undernut_dist)	0.07292	0.71765**	-0.23224	0.71353**	-0.38856	-0.02867	0.92721***	0.04331	0.67840*	-0.84485*	0.94210***	-0.48312
	(0.64388)	(0.29917)	(0.31275)	(0.33921)	(0.32516)	(0.31330)	(0.35198)	(0.42805)	(0.39045)	(0.44140)	(0.31818)	(0.57506)
Mort_distx1000	-0.00626	0.01500	0.00983	-0.01299	0.04554***	-0.06313***	0.03245*	-0.02967	-0.04620**	0.03020*	0.02969*	-0.08433
	(0.01853)	(0.01630)	(0.01560)	(0.02223)	(0.01671)	(0.01790)	(0.01946)	(0.02312)	(0.02127)	(0.01741)	(0.01790)	(0.08357)
P(poor_dist)	2.36792***	-1.09205***	0.70435**	-0.43255	-0.20692	0.56751*	-0.00942	-0.54729	-0.68895*	0.16339	-1.28428***	1.88036***
	(0.62910)	(0.28503)	(0.30465)	(0.33296)	(0.32321)	(0.30227)	(0.32872)	(0.41500)	(0.36767)	(0.38862)	(0.29715)	(0.52288)
lifeexp_dist	0.09604	0.04307	0.07759	-0.11506	0.16002***	-0.25127***	0.08446	-0.05942	-0.07698	0.05780	0.11905**	-0.27203
	(0.07000)	(0.05250)	(0.05083)	(0.07090)	(0.05463)	(0.05756)	(0.06344)	(0.07694)	(0.07097)	(0.05832)	(0.05788)	(0.25715)
D(ca_food)_1	0.06122	0.00702										
	(0.10380)	(0.05118)										
D(ca_food)_2	0.02809	0.06851										
	(0.11089)	(0.05253)										
D(ca_food)_3	0.18152	0.00570										
	(0.11546)	(0.05066)										
D(ca_food)_4	-0.12192	-0.02398										
	(0.10832)	(0.04908)										
Agropec_firmPC	-0.06398	-0.02960										
	(0.04187)	(0.01930)										
D(agro_dist)	0.11864	0.03421										
	(0.10758)	(0.04500)										
D(pec_dist)	-0.02205	-0.06270										
	(0.11955)	(0.05163)										
D(fish_dist)	0.46597**	0.00076										
	(0.19783)	(0.06300)										
Prop(workagro_dist)	-1.44438***	-0.80262***										
	(0.51303)	(0.25520)										

Prop(workfish_dist)	0.83220	2.07161**								
	(1.72810)	(0.93450)								
Hotel_firmPC			-0.03313	0.08169***						
			(0.02222)	(0.02507)						
D(rest_dist)			-0.07115	0.08811*						
			(0.04730)	(0.05130)						
D(tour_dist)			-0.03315	-0.06682						
			(0.04672)	(0.05234)						
Prop(workrest_dist)			8.35082***	-7.23300***						
			(2.17364)	(2.40466)						
Com_firmPC					-0.00376	-0.00103				
					(0.00458)	(0.00435)				
Prop(workcom_dist)					0.09871	2.66049***				
					(0.68942)	(0.68028)				
D(ca_cloth)_1					-0.20071**	-0.01062				
					(0.09175)	(0.08762)				
D(ca_cloth)_2					-0.00642	0.15777*				
					(0.09292)	(0.08833)				
D(ca_health)_1							-0.01459	0.10911		
							(0.05119)	(0.06842)		
D(ca_health)_2							0.01236	-0.04536		
							(0.04999)	(0.06619)		
n_chronic							0.04517**	0.04863*		
							(0.01872)	(0.02845)		
Health1_distx1000							-0.58038	0.72646		
							(0.38441)	(0.58983)		
Health2_distx1000							0.35439	-0.76783		
							(0.72508)	(0.96679)		
Health3_distx1000							0.33831	-0.63814		
							(0.43025)	(0.64520)		
Health4_distx1000							0.56963	-0.78309		
							(0.39229)	(0.59615)		
Health_firmPC							0.19715**	-0.24585**		
							(0.09120)	(0.11795)		
Prop(workhealth_dist)							2.65729	1.25137		

							(2.63676)	(3.43334)				
D(ca_transport)									0.07290	0.15265***		
									(0.05128)	(0.05018)		
comm1_distx1000									-0.08174	0.18855**		
									(0.08332)	(0.09431)		
comm2_distx1000									0.47212*	-0.04100		
									(0.27648)	(0.26839)		
comm3_distx1000									-1.21592**	-0.71579		
									(0.54528)	(0.63151)		
comm4_distx1000									-0.30926**	-0.19020		
									(0.15444)	(0.15083)		
comm5_distx1000									0.00091	0.00124		
									(0.02128)	(0.02236)		
comm6_distx1000									0.92129**	0.31068		
									(0.42816)	(0.38633)		
D(comm_dist)									0.14591**	-0.02443		
									(0.05902)	(0.05893)		
D(transp_dist)									0.16061*	0.10601		
									(0.09061)	(0.10327)		
transp_distx1000									-0.09879	0.05864		
									(0.07088)	(0.07739)		
D(good_roads_dist)									0.01783	-0.18528*		
									(0.10273)	(0.10845)		
Transport_firmPC									0.07488	-0.31498**		
									(0.13883)	(0.14454)		
D(TC_dist)									0.05232	0.00070		
									(0.07866)	(0.07354)		
Prop(worktransp_dist)									6.25037***	-0.51688		
									(1.27880)	(1.16777)		
Commun_firmPC									0.30634***	0.03042		
									(0.06595)	(0.06656)		
D(ca_educ)											0.02843	0.46454***
											(0.03994)	(0.07769)
D(ca_entert)											-0.00227	-0.24531**
											(0.05667)	(0.10033)

school_distx1000											-0.03432**	0.03713
											(0.01378)	(0.02473)
Educa_firmPC											0.38821***	-0.13270
											(0.11388)	(0.21306)
Entert_firmPC											-1.38455***	-0.07381
											(0.28426)	(0.51712)
Prop(workeduca_dist)											-0.46820	-0.05660
											(0.72711)	(1.41979)
Constant	-19.05277***	-1.57650	-6.37904	5.85095	-24.65814***	17.46112***	-10.82364*	-4.30091	6.28307	-12.27537**	-16.74992***	17.20068
	(7.06877)	(4.79282)	(4.83008)	(6.30671)	(5.02155)	(5.21635)	(5.67418)	(6.70530)	(6.43912)	(5.82459)	(5.24351)	(21.23984)
Observations	7,145	7,145	6,272	6,272	7,181	7,181	6,021	6,021	6,967	6,967	6,570	6,570
LogL	-626.346	-4082.925	-3843.844	-2999.566	-2800.179	-3659.533	-2753.117	-1480.547	-1895.040	-1964.155	-3378.221	-897.204
Pseudo-R2	0.356	0.173	0.096	0.121	0.165	0.166	0.084	0.284	0.216	0.366	0.085	0.634
Wald (overall)	571.647***	1326.961***	728.585***	723.333***	918.361***	1187.793***	458.021***	880.028***	938.327***	1453.861***	580.604***	1165.773***
Wald (indent. vars.)	164.02***	235.10***	209.63***	174.79***	368.38***	271.75***	175.36***	89.01***	330.99***	129.96***	309.39***	149.63***

d) SUR for markets within groups

	Fon		Foff		CC		HEA		TC		ED	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expend_Fon)	-0.01111	0.01161										
	(0.00720)	(0.01386)										
L(expend_Foff)			-0.13601***	-0.07062***								
			(0.00659)	(0.00701)								
L(expend_CC)					-0.08257***	-0.11007***						
					(0.00656)	(0.01116)						
L(expend_HEA)							-0.07617***	0.00938				
							(0.01278)	(0.00610)				
L(expend_TC)									-0.07099***	0.13331***		
									(0.00679)	(0.00652)		
L(expend_ED)											-0.14352***	0.14999***
											(0.00426)	(0.00352)
L(fam_size)	-0.01432	-0.05074***	-0.20732***	-0.14122***	-0.14284***	-0.15584***	-0.17865***	-0.02675	-0.09010***	0.06830***	-0.24900***	0.13550***
	(0.01286)	(0.01369)	(0.02630)	(0.02372)	(0.01775)	(0.02363)	(0.04322)	(0.01810)	(0.01605)	(0.01534)	(0.01558)	(0.01395)
D(mem_spo)	0.02389*	-0.01506	0.01731	-0.02077	0.00158	-0.00479	-0.03255	0.00961	0.06974***	-0.01976	0.02626	-0.02316*

	(0.01305)	(0.01519)	(0.02589)	(0.02284)	(0.01702)	(0.02383)	(0.04502)	(0.01681)	(0.01612)	(0.01520)	(0.01651)	(0.01367)
D(mem_son)	0.01098	-0.02507	-0.01108	0.03219	-0.04986***	-0.04571*	0.05018	0.00130	-0.00378	0.00198	-0.13960***	0.07354***
	(0.01421)	(0.01639)	(0.03014)	(0.02520)	(0.01935)	(0.02729)	(0.05048)	(0.01807)	(0.01646)	(0.01598)	(0.02036)	(0.02011)
D(mem_oth)	0.01146	-0.00166	0.01382	0.00020	0.00459	0.03434**	0.02238	0.00274	0.02163*	0.00246	-0.02137*	0.01978**
	(0.00927)	(0.00930)	(0.01875)	(0.01675)	(0.01235)	(0.01422)	(0.03074)	(0.01207)	(0.01153)	(0.01126)	(0.01138)	(0.00920)
P(men)	-0.05660***	0.01308	0.01213	0.03282	0.00528	0.01539	0.05865	0.00752	-0.03130	-0.02513	0.06078***	-0.06723***
	(0.01772)	(0.02128)	(0.03638)	(0.03144)	(0.02245)	(0.02856)	(0.06493)	(0.02322)	(0.02094)	(0.02004)	(0.02291)	(0.01838)
age_head	0.00547***	-0.00356*	-0.00136	-0.00048	-0.00465**	0.00407	-0.00710	-0.00149	-0.00348*	0.00410**	0.00103	0.00110
	(0.00162)	(0.00194)	(0.00344)	(0.00306)	(0.00214)	(0.00257)	(0.00605)	(0.00208)	(0.00191)	(0.00185)	(0.00225)	(0.00174)
age_head-sq	-0.00003**	0.00004**	0.00001	0.00003	0.00004*	-0.00002	0.00007	0.00001	0.00002	-0.00002	0.00003	-0.00002
	(0.00002)	(0.00002)	(0.00003)	(0.00003)	(0.00002)	(0.00002)	(0.00006)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
D(gender_head)	-0.01802	-0.00084	0.00004	0.01502	0.00912	-0.00073	0.04244	-0.02140	-0.00354	0.01247	-0.01403	-0.01246
	(0.01367)	(0.01655)	(0.02773)	(0.02415)	(0.01775)	(0.02324)	(0.04499)	(0.01697)	(0.01668)	(0.01569)	(0.01798)	(0.01406)
year_educ_head	0.00041	0.00632***	-0.00933***	0.01134***	-0.00237**	0.00454***	0.00246	-0.00387***	-0.01007***	0.00841***	0.00211*	-0.00383***
	(0.00087)	(0.00116)	(0.00190)	(0.00181)	(0.00120)	(0.00176)	(0.00309)	(0.00117)	(0.00110)	(0.00109)	(0.00116)	(0.00095)
Hrs_ff	0.00025	0.00002	0.00016	-0.00009	-0.00018	0.00016	-0.00033	-0.00017	0.00045**	-0.00033*	0.00054***	-0.00022
	(0.00017)	(0.00018)	(0.00032)	(0.00027)	(0.00023)	(0.00023)	(0.00049)	(0.00023)	(0.00018)	(0.00017)	(0.00017)	(0.00017)
Hrs_if	0.00047	0.00056*	0.00088	-0.00091*	0.00012	-0.00031	0.00051	0.00026	0.00079**	-0.00073**	0.00039	-0.00039
	(0.00033)	(0.00029)	(0.00055)	(0.00051)	(0.00047)	(0.00054)	(0.00081)	(0.00041)	(0.00033)	(0.00031)	(0.00037)	(0.00034)
Hrs_ii	-0.00001	-0.00001	0.00036	-0.00151***	-0.00021	-0.00006	-0.00054	0.00026	0.00089***	-0.00080***	0.00050**	-0.00048***
	(0.00015)	(0.00014)	(0.00024)	(0.00026)	(0.00018)	(0.00023)	(0.00037)	(0.00017)	(0.00016)	(0.00016)	(0.00020)	(0.00015)
Hrs_sf	0.00165***	0.00004	0.00022	0.00095*	0.00001	0.00005	0.00020	0.00034	-0.00095***	0.00058*	0.00040	0.00014
	(0.00031)	(0.00032)	(0.00063)	(0.00049)	(0.00043)	(0.00043)	(0.00090)	(0.00035)	(0.00033)	(0.00031)	(0.00032)	(0.00029)
Hrs_si	0.00044***	-0.00011	0.00113***	-0.00009	0.00004	-0.00009	0.00004	0.00001	0.00014	-0.00012	0.00051***	-0.00060***
	(0.00011)	(0.00011)	(0.00021)	(0.00021)	(0.00015)	(0.00019)	(0.00033)	(0.00014)	(0.00015)	(0.00015)	(0.00014)	(0.00012)
Hrs_fam	0.00042*	0.00057**	0.00029	-0.00003	0.00047*	0.00065	0.00031	0.00008	0.00010	-0.00020	0.00022	-0.00038*
	(0.00023)	(0.00028)	(0.00043)	(0.00038)	(0.00028)	(0.00055)	(0.00059)	(0.00027)	(0.00029)	(0.00027)	(0.00024)	(0.00021)
Hrs_second	0.00005	-0.00023	-0.00041	-0.00013	0.00117***	-0.00014	0.00013	-0.00007	0.00033	-0.00016	0.00007	-0.00024
	(0.00037)	(0.00031)	(0.00057)	(0.00055)	(0.00043)	(0.00052)	(0.00119)	(0.00042)	(0.00036)	(0.00036)	(0.00034)	(0.00032)
D(Hrs_ff)	0.01573	0.00192	-0.05621**	0.00859	0.02708	0.02263	0.02202	-0.01109	0.02258	-0.02130	-0.01867	-0.00167
	(0.01282)	(0.01367)	(0.02589)	(0.02193)	(0.01752)	(0.01969)	(0.04152)	(0.01797)	(0.01434)	(0.01373)	(0.01428)	(0.01314)
D(Hrs_if)	-0.02548	-0.03638**	-0.07784**	0.00935	0.00114	0.01951	-0.02893	-0.00581	0.01315	-0.03330	-0.00469	0.02155
	(0.02014)	(0.01756)	(0.03548)	(0.03269)	(0.02793)	(0.03530)	(0.05311)	(0.02524)	(0.02186)	(0.02089)	(0.02347)	(0.02045)
D(Hrs_ii)	0.01580	-0.01113	-0.08654***	-0.03064	0.02628*	0.00204	0.01135	-0.00005	0.02617*	-0.04603***	-0.02115	0.02078*
	(0.01154)	(0.01199)	(0.02141)	(0.02129)	(0.01482)	(0.01898)	(0.03579)	(0.01487)	(0.01363)	(0.01370)	(0.01487)	(0.01174)

D(Hrs_sf)	-0.07534***	-0.01234	-0.04346	0.00201	0.00882	0.00808	-0.01851	-0.03674	0.03083	0.00472	-0.01186	-0.01540
	(0.02102)	(0.02106)	(0.04005)	(0.03370)	(0.02705)	(0.02966)	(0.05965)	(0.02488)	(0.02357)	(0.02249)	(0.02381)	(0.02006)
D(Hrs_si)	0.00344	-0.01360	-0.00050	-0.02048	-0.00487	-0.00721	-0.02609	0.00354	0.00356	-0.00612	-0.02237*	0.02511**
	(0.00966)	(0.01043)	(0.02134)	(0.01776)	(0.01298)	(0.01640)	(0.03068)	(0.01284)	(0.01220)	(0.01173)	(0.01204)	(0.00981)
D(Hrs_fam)	0.00209	0.00054	-0.01466	-0.05535**	0.00612	-0.00988	0.02754	0.00418	0.01531	-0.02681*	-0.03437**	0.03274***
	(0.01241)	(0.01264)	(0.02424)	(0.02275)	(0.01568)	(0.02447)	(0.03918)	(0.01586)	(0.01576)	(0.01508)	(0.01398)	(0.01148)
D(Hrs_second)	-0.00863	0.01529	-0.03671*	0.00610	-0.04036***	-0.00740	0.04617	-0.01994	-0.01246	0.01006	-0.00150	0.00235
	(0.01122)	(0.01141)	(0.02032)	(0.01887)	(0.01358)	(0.01730)	(0.03663)	(0.01423)	(0.01284)	(0.01248)	(0.01213)	(0.01019)
D(Lima)	0.10451***	0.07177***	0.06927***	0.00340	-0.01502	0.03858**	-0.01113	-0.00647	0.09336***	-0.07546***	0.09591***	-0.02489***
	(0.00977)	(0.01185)	(0.01944)	(0.01647)	(0.01242)	(0.01610)	(0.02987)	(0.01202)	(0.00987)	(0.00961)	(0.01223)	(0.00958)
PHI	-0.22651***	-0.04543*	-0.04389	-0.05292	-0.41433***	-0.07401**	0.04710	0.16085***	0.44836***	0.56496***	-0.00381	0.39844***
	(0.06863)	(0.02421)	(0.04509)	(0.05273)	(0.03763)	(0.03021)	(0.05461)	(0.06201)	(0.02923)	(0.02370)	(0.03288)	(0.03602)
Constant	0.51113***	0.15951	1.69780***	1.15139***	1.42909***	0.95526***	0.99426***	0.94808***	1.10856***	-0.64058***	1.36824***	-0.43323***
	(0.06816)	(0.13148)	(0.10002)	(0.11041)	(0.07018)	(0.12183)	(0.22011)	(0.06540)	(0.06152)	(0.06063)	(0.06881)	(0.04682)
Observations	7,145		6,272		7,181		6,021		6,967		6,570	
LogL	4043.816		-3196.704		2225.751		309.848		2056.629		29.746	

Appendix 6: Definition of variables and basic tabulations (models in Chapter 7)

Variable	Definition	Price 7,513 obs. Mean	s.d.
Dependant variables			
Fon_inf	Consumption of Fon done at informal markets as a share of total Fon expenditure	0.600	0.28714
Fon_semi	Consumption of Fon done at semi-formal markets as a share of total Fon expenditure	0.332	0.28153
Fon_for	Consumption of Fon done at formal markets as a share of total Fon expenditure	0.068	0.14901
D(Fon_inf)	Dummy variable: 1 = household purchase from Fon_inf, 0 = otherwise	0.967	0.17832
D(Fon_semi)	Dummy variable: 1 = household purchase from Fon_semi, 0 = otherwise	0.932	0.25225
D(Fon_for)	Dummy variable: 1 = household purchase from Fon_for, 0 = otherwise	0.471	0.49916
Independant variables: SUR model			
L(expend_Fon)	Natural logarithm of total *per-capita* expenditure (purchased) in annual new *soles* (Fon)	6.943	0.75845
O1L(expend_Fon)	Natural logarithm of total *per-capita* expenditure (purchased) in annual new *soles* (Fon) adjusted by Stone price index using UV-1	5.648	0.74036
O2L(expend_Fon)	Natural logarithm of total *per-capita* expenditure (purchased) in annual new *soles* (Fon) adjusted by Stone price index using UV-2	5.650	0.74486
O3L(expend_Fon)	Natural logarithm of total *per-capita* expenditure (purchased) in annual new *soles* (Fon) adjusted by Stone price index using UV-3	5.643	0.74680
L(O1_uv_for)	Natural logarithm of UV-1 (formal)	1.970	0.36204
L(O1_uv_semi)	Natural logarithm of UV-1 (semi-formal)	1.402	0.11151
L(O1_uv_inf)	Natural logarithm of UV-1 (informal)	1.160	0.16511
L(O2_uv_for)	Natural logarithm of UV-2 (formal)	1.969	0.33268
L(O2_uv_semi)	Natural logarithm of UV-2 (semi-formal)	1.399	0.08443
L(O2_uv_inf)	Natural logarithm of UV-2 (informal)	1.158	0.12423
L(O3_uv_for)	Natural logarithm of UV-3 (formal)	1.963	0.26767
L(O3_uv_semi)	Natural logarithm of UV-3 (semi-formal)	1.404	0.07567
L(O3_uv_inf)	Natural logarithm of UV-3 (informal)	1.168	0.09634

L(fam_size)	Natural logarithm of the number of family members	1.299	0.56504
D(mem_spo)	Dummy variable: 1 = household has a spouse, 0 = otherwise	0.651	0.47673
D(mem_son)	Dummy variable: 1 = household has a son or daughter, 0 = otherwise	0.799	0.40076
D(mem_oth)	Dummy variable: 1 = household has other member, 0 = otherwise	0.307	0.46136
P(men)	Proportion of men in the household	0.502	0.24325
age_head	Age of the head of the household (in years)	49.048	14.45457
age_head-sq	Age of the head of the household (in years) squared	2614.644	1523.76500
D(gender_head)	Dummy variable: 1 = household head in male, 0 = otherwise	0.743	0.43709
year_educ_head	Number of years of education of the head of the household	10.303	4.82066
Hrs_ff	Sum of the weekly hours of work of all family members, primary activity, sector = formal workers in formal firms (see Chapter 4)	20.907	33.79519
Hrs_if	Sum of the weekly hours of work of all family members, primary activity, sector = informal workers in formal firms (see Chapter 4)	5.399	18.48301
Hrs_ii	Sum of the weekly hours of work of all family members, primary activity, sector = informal workers in informal firms (see Chapter 4)	18.420	35.53746
Hrs_sf	Sum of the weekly hours of work of all family members, primary activity, sector = informal self-employment (see Chapter 4)	7.952	22.00029
Hrs_si	Sum of the weekly hours of work of all family members, primary activity, sector = formal self-employment (see Chapter 4)	28.571	40.57227
Hrs_fam	Sum of the weekly hours of work of all family members, primary activity, sector = family work (see Chapter 4)	8.034	22.01353
Hrs_second	Sum of the weekly hours of work of all family members, secondary activity (see Chapter 4)	6.389	14.38882
D(Hrs_ff)	Dummy variable: 1 = any household member is formal worker in formal firm in his/her primary activity (see Chapter 4), 0 = otherwise	0.345	0.47540
D(Hrs_if)	Dummy variable: 1 = any household member is informal worker in formal firm in his/her primary activity (see Chapter 4), 0 = otherwise	0.103	0.30401
D(Hrs_ii)	Dummy variable: 1 = any household member is informal worker in informal firm in his/her primary activity (see Chapter 4), 0 = otherwise	0.306	0.46092
D(Hrs_sf)	Dummy variable: 1 = any household member is formal self-employer in his/her primary activity (see Chapter 4), 0 = otherwise	0.145	0.35194
D(Hrs_si)	Dummy variable: 1 = any household member is informal self-employer in his/her primary activity (see Chapter 4), 0 = otherwise	0.491	0.49995
D(Hrs_fam)	Dummy variable: 1 = any household member is family worker in his/her primary activity (see Chapter 4), 0 = otherwise	0.200	0.40007
D(Hrs_second)	Dummy variable: 1 = any household member has a second job (see Chapter 4), 0 = otherwise	0.279	0.44840
D(Lima)	Dummy variable: 1 = household lives in Lima, 0 = otherwise	0.159	0.36575
L(pop_urb_dist)	Natural logarithm of the population size of the district where the household lives	11.058	1.27918

L(pop_urb_dist)-sq	Natural logarithm of the population size of the district where the household lives squared	123.907	27.47688
pop_den_distx1000	Population density: inhabitants per 1,000 km^2	2.880	5.29031
D(not_slum)	Dummy variable: 1 = household does not live in a slum, 0 = otherwise	0.471	0.49919
D(mid_city)	Dummy variable: 1 = household lives in a departmental capital (except Lima), 0 = otherwise	0.514	0.49983
D(border)	Dummy variable: 1 = household lives in a border city (Tumbes, Tacna or Puno), 0 = otherwise	0.114	0.31759
D(tongue)	Dummy variable: 1 = any of the two parents of the household has indigenous mother tongue , 0 = otherwise	0.242	0.42823
D(migrant)	Dummy variable: 1 = any of the two parents of the household is a migrant, 0 = otherwise	0.646	0.47822
D(social)	Dummy variable: 1 = any of the household members pertain to social organizations (unions, social clubs, social programs, etc.) , 0 = otherwise	0.406	0.49108
Station_distx1000	Number of police stations in the district per 1,000 inhabitants	0.059	0.10864
Muni_persx1000	Number of workers in the local government of the district per 1,000 inhabitants	3.605	4.85469
L(budget_dist_pc)	Natural logarithm of the *per-capita* budget of the local government, in thousands of new *soles*	5.116	0.65747
D(sdp_dist)	Dummy variable: 1 = the district has a slum development plan, 0 = otherwise	0.049	0.21612
Independant variables: PROBIT model			
D(ca_food)_1	Dummy variable: 1 = the household considers it has an adequate consumption of food, 0 = otherwise	0.739	0.43914
D(ca_food)_2	Dummy variable: 1 = the household considers it has an adequate consumption of tubers and cereals, 0 = otherwise	0.630	0.48281
D(ca_food)_3	Dummy variable: 1 = the household considers it has an adequate consumption of fruits and vegetables, 0 = otherwise	0.555	0.49699
D(ca_food)_4	Dummy variable: 1 = the household considers it has an adequate consumption of meat, fish and chicken, 0 = otherwise	0.551	0.49741
Agropec_firmPC	Number of agropecuarian firms per 1,000 habitants (province level)	0.489	1.31546
D(agro_dist)	Dummy variable: 1 = agrarian activity is important in the district, 0 = otherwise (declared by the mayor)	0.582	0.49325
D(pec_dist)	Dummy variable: 1 = pecuarian activity is important in the district, 0 = otherwise (declared by the mayor)	0.290	0.45393
D(fish_dist)	Dummy variable: 1 = fishing activity is important in the district, 0 = otherwise (declared by the mayor)	0.111	0.31416
Prop(workagro_dist)	Proportion of urban workers in agropecuarian sector in the district	0.095	0.13118
Prop(workfish_dist)	Proportion of urban workers in fishing sector in the district	0.010	0.02792
P(pop_slums_dist)	Proportion of people in the district living in a slum	0.334	0.34174
P(migrant_dist)	Proportion of migrants living in the district (urban places)	0.313	0.10293

P(race_dist)	Proportion of indigenous people (by declared race) living in the district (urban places)	0.089	0.11666
informal_index	Index of tax informality at the province level, varies from 1 (high informality) to 0 (no informality) Index = (Tax of the IVA - (IVA payments/Value Added))/Tax of the IVA	0.279	0.30053
D(conf_gov_1)	Dummy variable: 1 = the head of the household considers that the public administration performance is good or very good , 0 = otherwise	0.308	0.46175
D(conf_gov_2)	Dummy variable: 1 = the head of the household considers that the performance of the government is not an obstacle for development , 0 = otherwise	0.190	0.39217
D(conf_gov_3)	Dummy variable: 1 = if the head of the household has a positive opinion in all responses related to the government , 0 = otherwise	0.082	0.27437
D(overall_satis)	Dummy variable: 1 = the head of the household considers his/her family as non-poor, 0 = otherwise	0.611	0.48751
D(income_satis)	Dummy variable: 1 = the head of the household is satisfied with his/her income, 0 = otherwise	0.356	0.47874
D(income_stab)	Dummy variable: 1 = the head of the household considers that his/her income is stable, 0 = otherwise	0.775	0.41731
D(sub_pov)	Dummy variable: 1 = if income is below the minimum income question, 0 = otherwise	0.325	0.46852
D(road_1)	Dummy variable: 1 = main road of the district is a non-asphalted highway, 0 = otherwise	0.175	0.38002
D(road_2)	Dummy variable: 1 = main road of the district is an asphalted highway, 0 = otherwise	0.678	0.46726
D(road_3)	Dummy variable: 1 = main road of the district is any other type, 0 = otherwise	0.029	0.16711
distance	Distance from the district to the capital of the province in minutes	75.054	126.70420
P(nowater_dist)	Proportion of households without water service in the district (urban places)	0.206	0.19863
P(nodrain_dist)	Proportion of houses without drain service in the district (urban places)	0.276	0.21222
P(noelec_dist)	Proportion of houses without electricity in the district (urban places)	0.111	0.08137
P(nofloor_dist)	Proportion of houses with floor of inadequate materials in the district (urban places)	0.295	0.16419
P(nowall_dist)	Proportion of houses with walls of inadequate materials in the district (urban places)	0.413	0.23527
P(noinsu_dist)	Proportion of people without insurance in the district (urban places)	0.577	0.09569
P(illiterate_dist)	Proportion of illiterate people in the district (urban places)	0.037	0.02837
P(selfwork_dist)	Proportion of self-employers in the district (urban places)	0.381	0.07310
P(unreg_dist)	Proportion of unregistered people in the district (urban places)	0.974	0.01436
P(undernut_dist)	Proportion of under-nourished children below 5 years old in the district	0.176	0.12226
Mort_distx1000	Child mortality, number per 1,000 born alive	16.943	6.31018
P(poor_dist)	Proportion of poor people in the district	0.291	0.16235

icf_index	Targeting index (between 0 and 1) of the government for social programs (district level)	0.127	0.16325
lifeexp_dist	Average life expectancy of the district	73.762	2.04419

Appendix 7: Full regression results (models in Chapter 7)

a) Results for models with UV-1

	Restriction 1				Restriction 2				Restriction 3			
	PROBIT		SUR		PROBIT		SUR		PROBIT		SUR	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expend_Fon)					0.56693***	0.55184***			0.56693***	0.55184***		
					(0.03851)	(0.03062)			(0.03851)	(0.03062)		
O1L(expend_Fon)	0.57087***	0.56310***	-0.00596	-0.00366			-0.00675	-0.00268			-0.00704	0.00207
	(0.03863)	(0.03096)	(0.00720)	(0.01331)			(0.00718)	(0.01300)			(0.00716)	(0.01300)
L(O1_uv_inf)	-0.35757	0.38637**	-0.14533***	0.07775***			-0.17200***	0.07993***			-0.19317***	0.16278***
	(0.26768)	(0.17089)	(0.02502)	(0.00910)			(0.02398)	(0.00909)			(0.02442)	(0.01843)
L(O1_uv_semi)	0.65455	0.43651**	0.06758***	-0.05840***			0.09207***	-0.01953*			0.11373***	-0.09647***
	(0.42218)	(0.19326)	(0.02530)	(0.01294)			(0.02447)	(0.01184)			(0.02393)	(0.01887)
L(O1_uv_for)	0.10512	-0.46372***	0.07775***	-0.01935*			0.07993***	-0.06040***			0.07944***	-0.06631***
	(0.11170)	(0.06111)	(0.00910)	(0.01113)			(0.00909)	(0.00951)			(0.00900)	(0.00933)
L(fam_size)	0.49040***	0.14956**	-0.00535	-0.06104***	0.49213***	0.14577**	-0.00684	-0.06049***	0.49213***	0.14577**	-0.00729	-0.05847***
	(0.13673)	(0.05905)	(0.01246)	(0.01336)	(0.13455)	(0.05881)	(0.01245)	(0.01336)	(0.13455)	(0.05881)	(0.01245)	(0.01334)
D(mem_spo)	0.09863	-0.09592	0.03533***	-0.01888	0.08425	-0.10800*	0.03581***	-0.01836	0.08425	-0.10800*	0.03621***	-0.02013
	(0.13132)	(0.06068)	(0.01256)	(0.01440)	(0.13235)	(0.06034)	(0.01257)	(0.01441)	(0.13235)	(0.06034)	(0.01255)	(0.01426)
D(mem_son)	0.22910*	0.03304	0.02381*	-0.03098**	0.22643*	0.03798	0.02397*	-0.03318**	0.22643*	0.03798	0.02406*	-0.03357**
	(0.13466)	(0.06376)	(0.01374)	(0.01561)	(0.13365)	(0.06352)	(0.01375)	(0.01565)	(0.13365)	(0.06352)	(0.01374)	(0.01555)
D(mem_oth)	0.00603	0.02390	0.01425	-0.00065	-0.00176	0.02727	0.01472	-0.00082	-0.00176	0.02727	0.01513*	-0.00156
	(0.11902)	(0.04540)	(0.00908)	(0.00925)	(0.11806)	(0.04526)	(0.00908)	(0.00927)	(0.11806)	(0.04526)	(0.00907)	(0.00919)
P(men)	-0.31675*	0.08064	-0.07255***	0.01582	-0.30304*	0.06990	-0.07144***	0.01577	-0.30304*	0.06990	-0.07112***	0.01559
	(0.16823)	(0.07929)	(0.01695)	(0.02062)	(0.16909)	(0.07881)	(0.01695)	(0.02073)	(0.16909)	(0.07881)	(0.01694)	(0.02059)
age_head	0.01903	-0.00611	0.00588***	-0.00351*	0.01998	-0.00455	0.00587***	-0.00377**	0.01998	-0.00455	0.00584***	-0.00367*
	(0.01379)	(0.00731)	(0.00157)	(0.00191)	(0.01376)	(0.00722)	(0.00157)	(0.00190)	(0.01376)	(0.00722)	(0.00157)	(0.00189)
age_head-sq	-0.00009	0.00008	-0.00004**	0.00004**	-0.00010	0.00007	-0.00004**	0.00004**	-0.00010	0.00007	-0.00004**	0.00004**

	(0.00014)	(0.00007)	(0.00002)	(0.00002)	(0.00014)	(0.00007)	(0.00002)	(0.00002)	(0.00014)	(0.00007)	(0.00002)	(0.00002)
D(gender_head)	-0.02166	0.12332*	-0.02318*	0.00283	-0.02722	0.14705**	-0.02296*	0.00038	-0.02722	0.14705**	-0.02320*	0.00199
	(0.13788)	(0.06430)	(0.01313)	(0.01573)	(0.13934)	(0.06383)	(0.01314)	(0.01577)	(0.13934)	(0.06383)	(0.01311)	(0.01560)
year_educ_head	-0.01185	0.03595***	-0.00047	0.00626***	-0.01170	0.03421***	-0.00050	0.00646***	-0.01170	0.03421***	-0.00048	0.00677***
	(0.00991)	(0.00434)	(0.00085)	(0.00117)	(0.00988)	(0.00431)	(0.00085)	(0.00114)	(0.00988)	(0.00431)	(0.00085)	(0.00114)
D(Lima)	-0.52076***	0.55424***	0.11759***	0.07080***	-0.57041***	0.66697***	0.11864***	0.06124***	-0.57041***	0.66697***	0.11915***	0.06295***
	(0.14952)	(0.07015)	(0.00924)	(0.01146)	(0.13895)	(0.06615)	(0.00924)	(0.01153)	(0.13895)	(0.06615)	(0.00914)	(0.01147)
Hrs_ff	-0.00108	0.00219**	0.00026	0.00010	-0.00105	0.00205**	0.00026	0.00012	-0.00105	0.00205**	0.00026	0.00013
	(0.00253)	(0.00092)	(0.00017)	(0.00019)	(0.00251)	(0.00091)	(0.00017)	(0.00019)	(0.00251)	(0.00091)	(0.00017)	(0.00019)
Hrs_if	-0.00254	-0.00165	0.00042	0.00045	-0.00237	-0.00163	0.00042	0.00047*	-0.00237	-0.00163	0.00042	0.00046
	(0.00325)	(0.00164)	(0.00032)	(0.00028)	(0.00337)	(0.00164)	(0.00032)	(0.00028)	(0.00337)	(0.00164)	(0.00032)	(0.00028)
Hrs_ii	-0.00129	-0.00128*	-0.00005	0.00003	-0.00106	-0.00137*	-0.00005	0.00002	-0.00106	-0.00137*	-0.00005	0.00002
	(0.00166)	(0.00072)	(0.00015)	(0.00014)	(0.00166)	(0.00072)	(0.00015)	(0.00014)	(0.00166)	(0.00072)	(0.00015)	(0.00014)
Hrs_sf	0.00627	0.00145	0.00169***	-0.00003	0.00618	0.00123	0.00167***	0.00001	0.00618	0.00123	0.00167***	0.00000
	(0.00508)	(0.00156)	(0.00031)	(0.00031)	(0.00491)	(0.00155)	(0.00031)	(0.00032)	(0.00491)	(0.00155)	(0.00031)	(0.00032)
Hrs_si	0.00216	0.00127**	0.00040***	-0.00008	0.00226	0.00113*	0.00039***	-0.00006	0.00226	0.00113*	0.00039***	-0.00005
	(0.00161)	(0.00060)	(0.00011)	(0.00011)	(0.00162)	(0.00060)	(0.00011)	(0.00011)	(0.00162)	(0.00060)	(0.00011)	(0.00011)
Hrs_fam	-0.00509**	-0.00074	0.00030	0.00058**	-0.00517**	-0.00090	0.00030	0.00061**	-0.00517**	-0.00090	0.00030	0.00060**
	(0.00224)	(0.00111)	(0.00023)	(0.00029)	(0.00223)	(0.00112)	(0.00023)	(0.00028)	(0.00223)	(0.00112)	(0.00023)	(0.00028)
Hrs_second	0.00659	0.00015	0.00008	-0.00027	0.00656	-0.00010	0.00009	-0.00027	0.00656	-0.00010	0.00011	-0.00031
	(0.00507)	(0.00160)	(0.00037)	(0.00031)	(0.00495)	(0.00160)	(0.00036)	(0.00031)	(0.00495)	(0.00160)	(0.00037)	(0.00031)
D(Hrs_ff)	0.16868	-0.04612	0.00803	0.00890	0.17409	-0.04611	0.00796	0.00831	0.17409	-0.04611	0.00751	0.00989
	(0.17759)	(0.06684)	(0.01271)	(0.01437)	(0.17599)	(0.06647)	(0.01271)	(0.01434)	(0.17599)	(0.06647)	(0.01271)	(0.01426)
D(Hrs_if)	0.12381	0.08103	-0.01943	-0.03329*	0.09041	0.07711	-0.01903	-0.03271*	0.09041	0.07711	-0.01865	-0.03346**
	(0.20804)	(0.10187)	(0.01947)	(0.01705)	(0.21040)	(0.10160)	(0.01950)	(0.01705)	(0.21040)	(0.10160)	(0.01948)	(0.01695)
D(Hrs_ii)	0.16169	0.03777	0.01216	-0.01080	0.15223	0.03578	0.01177	-0.00927	0.15223	0.03578	0.01168	-0.00847
	(0.12719)	(0.05586)	(0.01132)	(0.01174)	(0.12696)	(0.05583)	(0.01132)	(0.01175)	(0.12696)	(0.05583)	(0.01132)	(0.01170)
D(Hrs_sf)	-0.27003	0.08473	-0.07334***	-0.00590	-0.26919	0.09041	-0.07218***	-0.00655	-0.26919	0.09041	-0.07207***	-0.00564
	(0.27554)	(0.09917)	(0.02044)	(0.02065)	(0.26821)	(0.09839)	(0.02039)	(0.02087)	(0.26821)	(0.09839)	(0.02038)	(0.02084)
D(Hrs_si)	0.03962	-0.08039*	0.00542	-0.01606	0.03676	-0.08468*	0.00524	-0.01522	0.03676	-0.08468*	0.00503	-0.01556

	(0.11668)	(0.04788)	(0.00942)	(0.01010)	(0.11666)	(0.04771)	(0.00941)	(0.01015)	(0.11666)	(0.04771)	(0.00941)	(0.01013)
D(Hrs_fam)	0.08108	-0.00969	0.00075	0.00224	0.07843	0.00380	0.00091	0.00212	0.07843	0.00380	0.00078	0.00239
	(0.15763)	(0.06120)	(0.01230)	(0.01274)	(0.15723)	(0.06134)	(0.01226)	(0.01260)	(0.15723)	(0.06134)	(0.01227)	(0.01260)
D(Hrs_second)	0.09510	0.06436	-0.00759	0.01271	0.10310	0.07213	-0.00766	0.01232	0.10310	0.07213	-0.00766	0.01234
	(0.13677)	(0.05157)	(0.01097)	(0.01131)	(0.13484)	(0.05136)	(0.01095)	(0.01127)	(0.13484)	(0.05136)	(0.01095)	(0.01122)
P(pop_slums_dist)	0.00776	0.00256			-0.04430	0.07969			-0.04430	0.07969		
	(0.12847)	(0.05903)			(0.12734)	(0.05740)			(0.12734)	(0.05740)		
P(migrant_dist)	0.13901	-0.25379			0.18024	-0.24649			0.18024	-0.24649		
	(0.52690)	(0.25199)			(0.51392)	(0.25037)			(0.51392)	(0.25037)		
P(race_dist)	0.23823	0.52517*			0.42377	-0.28657			0.42377	-0.28657		
	(0.46498)	(0.27302)			(0.42861)	(0.25057)			(0.42861)	(0.25057)		
informal_index	0.01237	-0.14893**			-0.05393	-0.12182*			-0.05393	-0.12182*		
	(0.14979)	(0.07183)			(0.14623)	(0.07013)			(0.14623)	(0.07013)		
D(conf_gov_1)	0.02827	-0.04830			0.01997	-0.05265			0.01997	-0.05265		
	(0.09302)	(0.03945)			(0.09246)	(0.03936)			(0.09246)	(0.03936)		
D(conf_gov_2)	-0.10453	-0.12698**			-0.10316	-0.13439**			-0.10316	-0.13439**		
	(0.10912)	(0.05519)			(0.10850)	(0.05469)			(0.10850)	(0.05469)		
D(conf_gov_3)	-0.01382	0.12617			0.00501	0.12391			0.00501	0.12391		
	(0.16558)	(0.08628)			(0.16659)	(0.08521)			(0.16659)	(0.08521)		
D(overall_satis)	0.08578	-0.10330***			0.08940	-0.10733***			0.08940	-0.10733***		
	(0.08157)	(0.03671)			(0.08074)	(0.03652)			(0.08074)	(0.03652)		
D(income_satis)	0.11146	-0.08317**			0.10792	-0.08179**			0.10792	-0.08179**		
	(0.08498)	(0.03724)			(0.08444)	(0.03710)			(0.08444)	(0.03710)		
D(income_stab)	0.03804	-0.13723***			0.02924	-0.13535***			0.02924	-0.13535***		
	(0.09325)	(0.04177)			(0.09282)	(0.04148)			(0.09282)	(0.04148)		
D(sub_pov)	0.00152	-0.08601**			0.00182	-0.08120**			0.00182	-0.08120**		
	(0.07634)	(0.03590)			(0.07602)	(0.03584)			(0.07602)	(0.03584)		
D(road_1)	-0.18999	-0.01815			-0.19268	-0.04821			-0.19268	-0.04821		
	(0.13034)	(0.06547)			(0.12846)	(0.06417)			(0.12846)	(0.06417)		
D(road_2)	-0.23358*	0.10567*			-0.20665	0.11462**			-0.20665	0.11462**		

	(0.12911)	(0.05842)		(0.12780)	(0.05798)		(0.12780)	(0.05798)	
D(road_3)	0.15471	0.19519		0.17902	0.12951		0.17902	0.12951	
	(0.28498)	(0.12900)		(0.29140)	(0.12928)		(0.29140)	(0.12928)	
distance	-0.00034	0.00009		-0.00058**	-0.00003		-0.00058**	-0.00003	
	(0.00030)	(0.00019)		(0.00029)	(0.00018)		(0.00029)	(0.00018)	
P(nowater_dist)	0.47109*	-0.14208		0.53204**	-0.11970		0.53204**	-0.11970	
	(0.25231)	(0.14792)		(0.24867)	(0.14239)		(0.24867)	(0.14239)	
P(nodrain_dist)	-0.85538**	-0.26899		-0.95589***	-0.36117*		-0.95589***	-0.36117*	
	(0.35820)	(0.20669)		(0.34665)	(0.19988)		(0.34665)	(0.19988)	
P(noelec_dist)	1.92485**	-1.51051***		2.07547***	-1.58577***		2.07547***	-1.58577***	
	(0.78637)	(0.43312)		(0.74530)	(0.43348)		(0.74530)	(0.43348)	
P(nofloor_dist)	-0.02600	0.44540*		-0.12444	0.26527		-0.12444	0.26527	
	(0.53125)	(0.24887)		(0.52026)	(0.24438)		(0.52026)	(0.24438)	
P(nowall_dist)	0.30119	0.49940***		0.31744	0.56225***		0.31744	0.56225***	
	(0.34643)	(0.14265)		(0.34226)	(0.14214)		(0.34226)	(0.14214)	
P(noinsu_dist)	1.19515**	-0.08951		1.28200**	-0.31394		1.28200**	-0.31394	
	(0.53968)	(0.27504)		(0.52609)	(0.26523)		(0.52609)	(0.26523)	
P(illiterate_dist)	-3.51012	-0.92582		-3.75728	1.58955		-3.75728	1.58955	
	(2.50419)	(1.46958)		(2.49769)	(1.41485)		(2.49769)	(1.41485)	
P(selfwork_dist)	0.76447	0.26934		0.95424	0.20626		0.95424	0.20626	
	(0.85148)	(0.45522)		(0.84814)	(0.44523)		(0.84814)	(0.44523)	
P(unreg_dist)	8.84917**	-4.61979**		9.62524**	-5.29650***		9.62524**	-5.29650***	
	(3.81623)	(2.13932)		(3.76598)	(2.04666)		(3.76598)	(2.04666)	
icf_index	-0.80592*	0.03468		-1.11717**	0.35269		-1.11717**	0.35269	
	(0.47259)	(0.27775)		(0.46755)	(0.26136)		(0.46755)	(0.26136)	
P(undernut_dist)	-0.30501	0.99478***		0.05551	0.72415**		0.05551	0.72415**	
	(0.63911)	(0.30142)		(0.62044)	(0.29309)		(0.62044)	(0.29309)	
Mort_distx1000	-0.00368	0.00823		-0.01036	0.01884		-0.01036	0.01884	
	(0.01909)	(0.01891)		(0.01948)	(0.01663)		(0.01948)	(0.01663)	
P(poor_dist)	2.29411***	-1.10006***		2.57002***	-1.07810***		2.57002***	-1.07810***	

	(0.66781)	(0.29692)			(0.62227)	(0.28032)			(0.62227)	(0.28032)		
lifeexp_dist	0.08563	0.01885			0.06327	0.05670			0.06327	0.05670		
	(0.07127)	(0.06012)			(0.07323)	(0.05339)			(0.07323)	(0.05339)		
D(ca_food)_1	0.07921	0.00308			0.07998	0.01147			0.07998	0.01147		
	(0.09969)	(0.05066)			(0.09924)	(0.05050)			(0.09924)	(0.05050)		
D(ca_food)_2	0.01063	0.07968			0.02962	0.07169			0.02962	0.07169		
	(0.10885)	(0.05188)			(0.10721)	(0.05172)			(0.10721)	(0.05172)		
D(ca_food)_3	0.19425*	0.01044			0.19070*	0.00564			0.19070*	0.00564		
	(0.11260)	(0.04950)			(0.11247)	(0.04961)			(0.11247)	(0.04961)		
D(ca_food)_4	-0.08436	-0.02041			-0.10697	-0.02248			-0.10697	-0.02248		
	(0.10532)	(0.04807)			(0.10445)	(0.04806)			(0.10445)	(0.04806)		
Agropec_firmPC	-0.04049	-0.02426			-0.05484	-0.03096			-0.05484	-0.03096		
	(0.04396)	(0.01895)			(0.04315)	(0.01899)			(0.04315)	(0.01899)		
D(agro_dist)	0.12789	0.03851			0.13768	0.02670			0.13768	0.02670		
	(0.10229)	(0.04392)			(0.10168)	(0.04389)			(0.10168)	(0.04389)		
D(pec_dist)	-0.06012	-0.07484			-0.06448	-0.05682			-0.06448	-0.05682		
	(0.11222)	(0.05142)			(0.11543)	(0.05052)			(0.11543)	(0.05052)		
D(fish_dist)	0.34882*	0.01670			0.43587**	-0.00207			0.43587**	-0.00207		
	(0.17953)	(0.06253)			(0.18180)	(0.06171)			(0.18180)	(0.06171)		
Prop(workagro_dist)	-1.11172**	-0.93570***			-1.14757**	-0.80565***			-1.14757**	-0.80565***		
	(0.52946)	(0.25786)			(0.50026)	(0.24798)			(0.50026)	(0.24798)		
Prop(workfish_dist)	1.01626	1.72546*			1.22674	2.36604***			1.22674	2.36604***		
	(1.73818)	(0.91032)			(1.77318)	(0.91223)			(1.77318)	(0.91223)		
PHI			-0.13311**	-0.05838**			-0.14399**	-0.05859***			-0.14704**	-0.04373*
			(0.06728)	(0.02300)			(0.06633)	(0.02224)			(0.06627)	(0.02237)
Constant	-19.40841***	-0.84400	0.35993***	0.31995***	-18.60654***	-3.53289	0.35873***	0.34547***	-18.60654***	-3.53289	0.35633***	0.32099***
	(7.24520)	(5.35204)	(0.06404)	(0.10813)	(7.09483)	(4.81140)	(0.06363)	(0.10771)	(7.09483)	(4.81140)	(0.06350)	(0.10720)
Observations	7,513	7,513	7,513		7,513	7,513	7,513		7,513	7,513	7,513	
LogL	-692.833	-4228.773	4124.244		-697.076	-4266.134	4134.204		-697.076	-4266.134	4152.785	
Wald (ident. vars.)	100.09***	227.72***			176.93***	252.03***			176.93***	252.03***		

b) Results for models with UV-2

	Restriction 1				Restriction 2				Restriction 3			
	PROBIT		SUR		PROBIT		SUR		PROBIT		SUR	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expend_Fon)					0.56693***	0.55184***			0.56693***	0.55184***		
					(0.03851)	(0.03062)			(0.03851)	(0.03062)		
O2L(expend_Fon)	0.57560***	0.55904***	-0.01073	-0.00006			-0.01149	0.00342			-0.01124	0.00801
	(0.03834)	(0.03082)	(0.00717)	(0.01376)			(0.00718)	(0.01351)			(0.00716)	(0.01357)
L(O2_uv_inf)	-1.17828***	0.07250	-0.05553*	0.06436***			-0.10769***	0.06903***			-0.12805***	0.14365***
	(0.44486)	(0.24273)	(0.03120)	(0.00929)			(0.03058)	(0.00918)			(0.03119)	(0.01806)
L(O2_uv_semi)	1.35390**	0.85780***	-0.00883	-0.08200***			0.03866	-0.05243***			0.05794*	-0.12075***
	(0.54006)	(0.25295)	(0.03095)	(0.01406)			(0.03052)	(0.01345)			(0.03010)	(0.02029)
L(O2_uv_for)	0.28939**	-0.34219***	0.06436***	0.01764			0.06903***	-0.01661			0.07010***	-0.02289**
	(0.12750)	(0.06844)	(0.00929)	(0.01222)			(0.00918)	(0.01127)			(0.00881)	(0.01060)
L(fam_size)	0.50081***	0.14925**	-0.00841	-0.05955***	0.49213***	0.14577**	-0.01004	-0.05826***	0.49213***	0.14577**	-0.01003	-0.05669***
	(0.13663)	(0.05900)	(0.01244)	(0.01335)	(0.13455)	(0.05881)	(0.01246)	(0.01342)	(0.13455)	(0.05881)	(0.01246)	(0.01344)
D(mem_spo)	0.08304	-0.10861*	0.03311***	-0.01887	0.08425	-0.10800*	0.03338***	-0.01924	0.08425	-0.10800*	0.03348***	-0.02017
	(0.12941)	(0.06049)	(0.01261)	(0.01436)	(0.13235)	(0.06034)	(0.01265)	(0.01447)	(0.13235)	(0.06034)	(0.01264)	(0.01436)
D(mem_son)	0.22699*	0.03328	0.02153	-0.03163**	0.22643*	0.03798	0.02157	-0.03317**	0.22643*	0.03798	0.02173	-0.03339**
	(0.13473)	(0.06362)	(0.01376)	(0.01562)	(0.13365)	(0.06352)	(0.01378)	(0.01572)	(0.13365)	(0.06352)	(0.01378)	(0.01566)
D(mem_oth)	0.00014	0.02680	0.01209	-0.00047	-0.00176	0.02727	0.01294	-0.00004	-0.00176	0.02727	0.01321	-0.00050
	(0.11829)	(0.04531)	(0.00912)	(0.00923)	(0.11806)	(0.04526)	(0.00912)	(0.00929)	(0.11806)	(0.04526)	(0.00912)	(0.00923)
P(men)	-0.30690*	0.07491	-0.07067***	0.01636	-0.30304*	0.06990	-0.06903***	0.01598	-0.30304*	0.06990	-0.06891***	0.01589
	(0.16687)	(0.07901)	(0.01702)	(0.02063)	(0.16909)	(0.07881)	(0.01703)	(0.02080)	(0.16909)	(0.07881)	(0.01702)	(0.02071)
age_head	0.02179	-0.00559	0.00565***	-0.00363*	0.01998	-0.00455	0.00568***	-0.00370*	0.01998	-0.00455	0.00568***	-0.00374**
	(0.01371)	(0.00727)	(0.00157)	(0.00190)	(0.01376)	(0.00722)	(0.00157)	(0.00190)	(0.01376)	(0.00722)	(0.00157)	(0.00190)
age_head-sq	-0.00012	0.00008	-0.00004**	0.00004**	-0.00010	0.00007	-0.00004**	0.00004**	-0.00010	0.00007	-0.00004**	0.00004**
	(0.00014)	(0.00007)	(0.00002)	(0.00002)	(0.00014)	(0.00007)	(0.00002)	(0.00002)	(0.00014)	(0.00007)	(0.00002)	(0.00002)

D(gender_head)	-0.01839	0.13933**	-0.02312*	0.00366	-0.02722	0.14705**	-0.02305*	0.00260	-0.02722	0.14705**	-0.02308*	0.00354
	(0.13547)	(0.06404)	(0.01321)	(0.01569)	(0.13934)	(0.06383)	(0.01324)	(0.01584)	(0.13934)	(0.06383)	(0.01323)	(0.01570)
year_educ_head	-0.01412	0.03512***	-0.00029	0.00590***	-0.01170	0.03421***	-0.00029	0.00629***	-0.01170	0.03421***	-0.00028	0.00661***
	(0.01007)	(0.00434)	(0.00085)	(0.00117)	(0.00988)	(0.00431)	(0.00085)	(0.00114)	(0.00988)	(0.00431)	(0.00085)	(0.00114)
D(Lima)	-0.41779**	0.55336***	0.11768***	0.07656***	-0.57041***	0.66697***	0.12048***	0.07152***	-0.57041***	0.66697***	0.12131***	0.07290***
	(0.17091)	(0.07575)	(0.00956)	(0.01183)	(0.13895)	(0.06615)	(0.00956)	(0.01197)	(0.13895)	(0.06615)	(0.00941)	(0.01188)
Hrs_ff	-0.00120	0.00210**	0.00028*	0.00008	-0.00105	0.00205**	0.00029*	0.00010	-0.00105	0.00205**	0.00029*	0.00011
	(0.00257)	(0.00092)	(0.00017)	(0.00019)	(0.00251)	(0.00091)	(0.00017)	(0.00019)	(0.00251)	(0.00091)	(0.00017)	(0.00019)
Hrs_if	-0.00177	-0.00159	0.00043	0.00047*	-0.00237	-0.00163	0.00044	0.00046	-0.00237	-0.00163	0.00045	0.00044
	(0.00352)	(0.00163)	(0.00032)	(0.00028)	(0.00337)	(0.00164)	(0.00032)	(0.00028)	(0.00337)	(0.00164)	(0.00032)	(0.00028)
Hrs_ii	-0.00127	-0.00129*	-0.00002	0.00002	-0.00106	-0.00137*	-0.00002	0.00001	-0.00106	-0.00137*	-0.00002	0.00000
	(0.00169)	(0.00072)	(0.00015)	(0.00014)	(0.00166)	(0.00072)	(0.00015)	(0.00014)	(0.00166)	(0.00072)	(0.00015)	(0.00014)
Hrs_sf	0.00619	0.00141	0.00170***	-0.00007	0.00618	0.00123	0.00168***	-0.00002	0.00618	0.00123	0.00168***	-0.00002
	(0.00499)	(0.00155)	(0.00031)	(0.00031)	(0.00491)	(0.00155)	(0.00031)	(0.00032)	(0.00491)	(0.00155)	(0.00031)	(0.00032)
Hrs_si	0.00230	0.00119**	0.00041***	-0.00009	0.00226	0.00113*	0.00041***	-0.00007	0.00226	0.00113*	0.00041***	-0.00007
	(0.00161)	(0.00060)	(0.00011)	(0.00011)	(0.00162)	(0.00060)	(0.00011)	(0.00011)	(0.00162)	(0.00060)	(0.00011)	(0.00011)
Hrs_fam	-0.00513**	-0.00074	0.00032	0.00058**	-0.00517**	-0.00090	0.00033	0.00059**	-0.00517**	-0.00090	0.00033	0.00058**
	(0.00223)	(0.00111)	(0.00023)	(0.00029)	(0.00223)	(0.00112)	(0.00023)	(0.00028)	(0.00223)	(0.00112)	(0.00023)	(0.00028)
Hrs_second	0.00708	0.00028	0.00004	-0.00030	0.00656	-0.00010	0.00007	-0.00028	0.00656	-0.00010	0.00008	-0.00032
	(0.00486)	(0.00161)	(0.00036)	(0.00030)	(0.00495)	(0.00160)	(0.00036)	(0.00031)	(0.00495)	(0.00160)	(0.00036)	(0.00031)
D(Hrs_ff)	0.17230	-0.04041	0.00868	0.00850	0.17409	-0.04611	0.00806	0.00850	0.17409	-0.04611	0.00778	0.00954
	(0.17970)	(0.06689)	(0.01277)	(0.01437)	(0.17599)	(0.06647)	(0.01276)	(0.01443)	(0.17599)	(0.06647)	(0.01276)	(0.01440)
D(Hrs_if)	0.07768	0.08389	-0.02152	-0.03411**	0.09041	0.07711	-0.02155	-0.03245*	0.09041	0.07711	-0.02145	-0.03233*
	(0.21909)	(0.10147)	(0.01957)	(0.01692)	(0.21040)	(0.10160)	(0.01960)	(0.01714)	(0.21040)	(0.10160)	(0.01958)	(0.01709)
D(Hrs_ii)	0.15892	0.03514	0.01172	-0.01060	0.15223	0.03578	0.01169	-0.01070	0.15223	0.03578	0.01159	-0.01005
	(0.12920)	(0.05590)	(0.01133)	(0.01182)	(0.12696)	(0.05583)	(0.01131)	(0.01181)	(0.12696)	(0.05583)	(0.01131)	(0.01180)
D(Hrs_sf)	-0.26828	0.08763	-0.07518***	-0.00404	-0.26919	0.09041	-0.07386***	-0.00536	-0.26919	0.09041	-0.07419***	-0.00421
	(0.27073)	(0.09831)	(0.02047)	(0.02079)	(0.26821)	(0.09839)	(0.02046)	(0.02094)	(0.26821)	(0.09839)	(0.02045)	(0.02098)
D(Hrs_si)	0.03469	-0.07886*	0.00603	-0.01554	0.03676	-0.08468*	0.00569	-0.01579	0.03676	-0.08468*	0.00560	-0.01647
	(0.11608)	(0.04777)	(0.00944)	(0.01009)	(0.11666)	(0.04771)	(0.00943)	(0.01018)	(0.11666)	(0.04771)	(0.00943)	(0.01016)

D(Hrs_fam)	0.07099	-0.00780	0.00042	0.00295	0.07843	0.00380	-0.00072	0.00298	0.07843	0.00380	-0.00102	0.00360
	(0.15608)	(0.06130)	(0.01227)	(0.01271)	(0.15723)	(0.06134)	(0.01224)	(0.01258)	(0.15723)	(0.06134)	(0.01224)	(0.01257)
D(Hrs_second)	0.09093	0.05925	-0.00779	0.01418	0.10310	0.07213	-0.00825	0.01305	0.10310	0.07213	-0.00836	0.01356
	(0.13300)	(0.05145)	(0.01093)	(0.01126)	(0.13484)	(0.05136)	(0.01092)	(0.01129)	(0.13484)	(0.05136)	(0.01092)	(0.01126)
P(pop_slums_dist)	0.02679	0.01785			-0.04430	0.07969			-0.04430	0.07969		
	(0.13031)	(0.05896)			(0.12734)	(0.05740)			(0.12734)	(0.05740)		
P(migrant_dist)	0.01127	-0.24910			0.18024	-0.24649			0.18024	-0.24649		
	(0.52854)	(0.25391)			(0.51392)	(0.25037)			(0.51392)	(0.25037)		
P(race_dist)	0.18627	0.30824			0.42377	-0.28657			0.42377	-0.28657		
	(0.47120)	(0.27789)			(0.42861)	(0.25057)			(0.42861)	(0.25057)		
informal_index	-0.03032	-0.14629**			-0.05393	-0.12182*			-0.05393	-0.12182*		
	(0.15688)	(0.07183)			(0.14623)	(0.07013)			(0.14623)	(0.07013)		
D(conf_gov_1)	0.02939	-0.04536			0.01997	-0.05265			0.01997	-0.05265		
	(0.09387)	(0.03943)			(0.09246)	(0.03936)			(0.09246)	(0.03936)		
D(conf_gov_2)	-0.10150	-0.12632**			-0.10316	-0.13439**			-0.10316	-0.13439**		
	(0.10952)	(0.05495)			(0.10850)	(0.05469)			(0.10850)	(0.05469)		
D(conf_gov_3)	-0.03549	0.12589			0.00501	0.12391			0.00501	0.12391		
	(0.16730)	(0.08611)			(0.16659)	(0.08521)			(0.16659)	(0.08521)		
D(overall_satis)	0.09118	-0.10563***			0.08940	-0.10733***			0.08940	-0.10733***		
	(0.08172)	(0.03662)			(0.08074)	(0.03652)			(0.08074)	(0.03652)		
D(income_satis)	0.10250	-0.08192**			0.10792	-0.08179**			0.10792	-0.08179**		
	(0.08486)	(0.03719)			(0.08444)	(0.03710)			(0.08444)	(0.03710)		
D(income_stab)	0.03139	-0.13639***			0.02924	-0.13535***			0.02924	-0.13535***		
	(0.09349)	(0.04162)			(0.09282)	(0.04148)			(0.09282)	(0.04148)		
D(sub_pov)	-0.00655	-0.08900**			0.00182	-0.08120**			0.00182	-0.08120**		
	(0.07711)	(0.03586)			(0.07602)	(0.03584)			(0.07602)	(0.03584)		
D(road_1)	-0.24644*	-0.03475			-0.19268	-0.04821			-0.19268	-0.04821		
	(0.13465)	(0.06614)			(0.12846)	(0.06417)			(0.12846)	(0.06417)		
D(road_2)	-0.26585**	0.09581			-0.20665	0.11462**			-0.20665	0.11462**		
	(0.12963)	(0.05881)			(0.12780)	(0.05798)			(0.12780)	(0.05798)		

D(road_3)	0.11654	0.18037		0.17902	0.12951		0.17902	0.12951		
	(0.28464)	(0.12887)		(0.29140)	(0.12928)		(0.29140)	(0.12928)		
distance	-0.00057**	-0.00002		-0.00058**	-0.00003		-0.00058**	-0.00003		
	(0.00028)	(0.00018)		(0.00029)	(0.00018)		(0.00029)	(0.00018)		
P(nowater_dist)	0.42441	-0.17526		0.53204**	-0.11970		0.53204**	-0.11970		
	(0.26075)	(0.14660)		(0.24867)	(0.14239)		(0.24867)	(0.14239)		
P(nodrain_dist)	-0.75845**	-0.23329		-0.95589***	-0.36117*		-0.95589***	-0.36117*		
	(0.36891)	(0.20446)		(0.34665)	(0.19988)		(0.34665)	(0.19988)		
P(noelec_dist)	1.82241**	-1.56153***		2.07547***	-1.58577***		2.07547***	-1.58577***		
	(0.77343)	(0.43466)		(0.74530)	(0.43348)		(0.74530)	(0.43348)		
P(nofloor_dist)	0.11150	0.38133		-0.12444	0.26527		-0.12444	0.26527		
	(0.54539)	(0.24764)		(0.52026)	(0.24438)		(0.52026)	(0.24438)		
P(nowall_dist)	0.16741	0.54437***		0.31744	0.56225***		0.31744	0.56225***		
	(0.35130)	(0.14274)		(0.34226)	(0.14214)		(0.34226)	(0.14214)		
P(noinsu_dist)	0.68895	-0.37573		1.28200**	-0.31394		1.28200**	-0.31394		
	(0.53433)	(0.28064)		(0.52609)	(0.26523)		(0.52609)	(0.26523)		
P(illiterate_dist)	-3.69442	-0.59252		-3.75728	1.58955		-3.75728	1.58955		
	(2.57473)	(1.48331)		(2.49769)	(1.41485)		(2.49769)	(1.41485)		
P(selfwork_dist)	1.25863	0.34805		0.95424	0.20626		0.95424	0.20626		
	(0.85704)	(0.45091)		(0.84814)	(0.44523)		(0.84814)	(0.44523)		
P(unreg_dist)	6.61604	-5.16514**		9.62524**	-5.29650***		9.62524**	-5.29650***		
	(4.07361)	(2.17429)		(3.76598)	(2.04666)		(3.76598)	(2.04666)		
icf_index	-0.83282*	0.21922		-1.11717**	0.35269		-1.11717**	0.35269		
	(0.47451)	(0.27138)		(0.46755)	(0.26136)		(0.46755)	(0.26136)		
P(undernut_dist)	-0.28404	0.66549**		0.05551	0.72415**		0.05551	0.72415**		
	(0.63585)	(0.30054)		(0.62044)	(0.29309)		(0.62044)	(0.29309)		
Mort_distx1000	-0.00733	0.01642		-0.01036	0.01884		-0.01036	0.01884		
	(0.02102)	(0.01671)		(0.01948)	(0.01663)		(0.01948)	(0.01663)		
P(poor_dist)	2.22234***	-1.07257***		2.57002***	-1.07810***		2.57002***	-1.07810***		
	(0.69249)	(0.29593)		(0.62227)	(0.28032)		(0.62227)	(0.28032)		

lifeexp_dist	0.07228	0.04299			0.06327	0.05670			0.06327	0.05670		
	(0.07735)	(0.05358)			(0.07323)	(0.05339)			(0.07323)	(0.05339)		
D(ca_food)_1	0.09636	0.00759			0.07998	0.01147			0.07998	0.01147		
	(0.09985)	(0.05060)			(0.09924)	(0.05050)			(0.09924)	(0.05050)		
D(ca_food)_2	0.00678	0.07465			0.02962	0.07169			0.02962	0.07169		
	(0.10794)	(0.05183)			(0.10721)	(0.05172)			(0.10721)	(0.05172)		
D(ca_food)_3	0.17912	0.00733			0.19070*	0.00564			0.19070*	0.00564		
	(0.11166)	(0.04947)			(0.11247)	(0.04961)			(0.11247)	(0.04961)		
D(ca_food)_4	-0.08162	-0.01916			-0.10697	-0.02248			-0.10697	-0.02248		
	(0.10480)	(0.04801)			(0.10445)	(0.04806)			(0.10445)	(0.04806)		
Agropec_firmPC	-0.03613	-0.02633			-0.05484	-0.03096			-0.05484	-0.03096		
	(0.04539)	(0.01921)			(0.04315)	(0.01899)			(0.04315)	(0.01899)		
D(agro_dist)	0.12782	0.00687			0.13768	0.02670			0.13768	0.02670		
	(0.10460)	(0.04412)			(0.10168)	(0.04389)			(0.10168)	(0.04389)		
D(pec_dist)	-0.07185	-0.04409			-0.06448	-0.05682			-0.06448	-0.05682		
	(0.11747)	(0.05119)			(0.11543)	(0.05052)			(0.11543)	(0.05052)		
D(fish_dist)	0.32519*	0.00576			0.43587**	-0.00207			0.43587**	-0.00207		
	(0.17951)	(0.06219)			(0.18180)	(0.06171)			(0.18180)	(0.06171)		
Prop(workagro_dist)	-1.04719**	-0.82522***			-1.14757**	-0.80565***			-1.14757**	-0.80565***		
	(0.51960)	(0.25149)			(0.50026)	(0.24798)			(0.50026)	(0.24798)		
Prop(workfish_dist)	1.45637	1.77304*			1.22674	2.36604***			1.22674	2.36604***		
	(1.90191)	(0.90921)			(1.77318)	(0.91223)			(1.77318)	(0.91223)		
PHI			-0.18210***	-0.06208***			-0.20214***	-0.05438**			-0.20152***	-0.03924*
			(0.06526)	(0.02395)			(0.06590)	(0.02296)			(0.06584)	(0.02323)
Constant	-16.40651**	-2.49080	0.43080***	0.28277**	-18.60654***	-3.53289	0.42126***	0.27389**	-18.60654***	-3.53289	0.41355***	0.25108**
	(7.69309)	(4.90095)	(0.06398)	(0.11187)	(7.09483)	(4.81140)	(0.06375)	(0.11063)	(7.09483)	(4.81140)	(0.06355)	(0.11050)
Observations	7,513	7,513	7,513		7,513	7,513	7,513		7,513	7,513	7,513	
LogL	-688.412	-4244.802	4091.817		-697.076	-4266.134	4090.41		-697.076	-4266.134	4099.195	
Wald (ident. vars.)	112.55***	224.12***			176.93***	252.03***			176.93***	252.03***		

c) Results for models with UV-3

	Restriction 1				Restriction 2				Restriction 3			
	PROBIT		SUR		PROBIT		SUR		PROBIT		SUR	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expend_Fon)					0.56693***	0.55184***			0.56693***	0.55184***		
					(0.03851)	(0.03062)			(0.03851)	(0.03062)		
O3L(expend_Fon)	0.57590***	0.55286***	-0.00506	-0.00486			-0.00695	-0.00144			-0.00624	0.00024
	(0.03852)	(0.03073)	(0.00715)	(0.01367)			(0.00714)	(0.01349)			(0.00714)	(0.01348)
L(O3_uv_inf)	-1.87149***	0.92472***	-0.23995***	0.04132***			-0.32262***	0.04715***			-0.34501***	0.12296***
	(0.49572)	(0.29765)	(0.04842)	(0.01184)			(0.04595)	(0.01183)			(0.04653)	(0.02402)
L(O3_uv_semi)	1.73766***	0.23860	0.19862***	-0.03980**			0.27547***	-0.01149			0.28500***	-0.07898***
	(0.58066)	(0.28821)	(0.04848)	(0.01765)			(0.04633)	(0.01777)			(0.04558)	(0.02776)
L(O3_uv_for)	0.12557	-0.22710***	0.04132***	-0.00153			0.04715***	-0.03566**			0.06001***	-0.04398***
	(0.12283)	(0.07100)	(0.01184)	(0.01455)			(0.01183)	(0.01435)			(0.01172)	(0.01332)
L(fam_size)	0.50794***	0.14829**	-0.00318	-0.06266***	0.49213***	0.14577**	-0.00537	-0.06047***	0.49213***	0.14577**	-0.00503	-0.06032***
	(0.13644)	(0.05895)	(0.01246)	(0.01344)	(0.13455)	(0.05881)	(0.01245)	(0.01342)	(0.13455)	(0.05881)	(0.01245)	(0.01341)
D(mem_spo)	0.07275	-0.10464*	0.03143**	-0.01729	0.08425	-0.10800*	0.03082**	-0.01745	0.08425	-0.10800*	0.03090**	-0.01742
	(0.12972)	(0.06046)	(0.01259)	(0.01449)	(0.13235)	(0.06034)	(0.01260)	(0.01457)	(0.13235)	(0.06034)	(0.01260)	(0.01448)
D(mem_son)	0.22535*	0.03528	0.01965	-0.03137**	0.22643*	0.03798	0.01875	-0.03247**	0.22643*	0.03798	0.01920	-0.03226**
	(0.13456)	(0.06365)	(0.01374)	(0.01576)	(0.13365)	(0.06352)	(0.01375)	(0.01577)	(0.13365)	(0.06352)	(0.01375)	(0.01572)
D(mem_oth)	-0.00955	0.02455	0.01075	0.00083	-0.00176	0.02727	0.01129	0.00070	-0.00176	0.02727	0.01154	0.00052
	(0.11809)	(0.04529)	(0.00912)	(0.00932)	(0.11806)	(0.04526)	(0.00911)	(0.00934)	(0.11806)	(0.04526)	(0.00911)	(0.00929)
P(men)	-0.31179*	0.07291	-0.07367***	0.01741	-0.30304*	0.06990	-0.07147***	0.01672	-0.30304*	0.06990	-0.07170***	0.01677
	(0.16844)	(0.07889)	(0.01700)	(0.02089)	(0.16909)	(0.07881)	(0.01698)	(0.02089)	(0.16909)	(0.07881)	(0.01698)	(0.02082)
age_head	0.02161	-0.00556	0.00572***	-0.00356*	0.01998	-0.00455	0.00568***	-0.00371*	0.01998	-0.00455	0.00573***	-0.00374**
	(0.01381)	(0.00725)	(0.00157)	(0.00191)	(0.01376)	(0.00722)	(0.00157)	(0.00191)	(0.01376)	(0.00722)	(0.00158)	(0.00191)
age_head-sq	-0.00012	0.00008	-0.00004**	0.00004**	-0.00010	0.00007	-0.00004**	0.00004**	-0.00010	0.00007	-0.00004**	0.00004**
	(0.00014)	(0.00007)	(0.00002)	(0.00002)	(0.00014)	(0.00007)	(0.00002)	(0.00002)	(0.00014)	(0.00007)	(0.00002)	(0.00002)
D(gender_head)	-0.01513	0.13952**	-0.02395*	0.00204	-0.02722	0.14705**	-0.02329*	0.00140	-0.02722	0.14705**	-0.02307*	0.00146

	(0.13632)	(0.06392)	(0.01318)	(0.01589)	(0.13934)	(0.06383)	(0.01318)	(0.01598)	(0.13934)	(0.06383)	(0.01317)	(0.01585)
year_educ_head	-0.01269	0.03511***	-0.00016	0.00594***	-0.01170	0.03421***	-0.00011	0.00617***	-0.01170	0.03421***	-0.00016	0.00629***
	(0.01005)	(0.00433)	(0.00085)	(0.00116)	(0.00988)	(0.00431)	(0.00085)	(0.00114)	(0.00988)	(0.00431)	(0.00085)	(0.00114)
D(Lima)	-0.42125***	0.57133***	0.11404***	0.07001***	-0.57041***	0.66697***	0.11983***	0.06823***	-0.57041***	0.66697***	0.12234***	0.06531***
	(0.16173)	(0.07420)	(0.00955)	(0.01183)	(0.13895)	(0.06615)	(0.00959)	(0.01202)	(0.13895)	(0.06615)	(0.00954)	(0.01185)
Hrs_ff	-0.00105	0.00209**	0.00030*	0.00009	-0.00105	0.00205**	0.00031*	0.00010	-0.00105	0.00205**	0.00031*	0.00010
	(0.00262)	(0.00092)	(0.00017)	(0.00019)	(0.00251)	(0.00091)	(0.00017)	(0.00019)	(0.00251)	(0.00091)	(0.00017)	(0.00019)
Hrs_if	-0.00182	-0.00174	0.00042	0.00048*	-0.00237	-0.00163	0.00043	0.00047*	-0.00237	-0.00163	0.00043	0.00047*
	(0.00355)	(0.00164)	(0.00032)	(0.00028)	(0.00337)	(0.00164)	(0.00032)	(0.00028)	(0.00337)	(0.00164)	(0.00032)	(0.00028)
Hrs_ii	-0.00112	-0.00130*	-0.00002	0.00002	-0.00106	-0.00137*	-0.00002	0.00002	-0.00106	-0.00137*	-0.00003	0.00002
	(0.00171)	(0.00072)	(0.00015)	(0.00014)	(0.00166)	(0.00072)	(0.00015)	(0.00014)	(0.00166)	(0.00072)	(0.00015)	(0.00014)
Hrs_sf	0.00662	0.00126	0.00173***	-0.00003	0.00618	0.00123	0.00171***	-0.00002	0.00618	0.00123	0.00172***	-0.00001
	(0.00504)	(0.00155)	(0.00031)	(0.00032)	(0.00491)	(0.00155)	(0.00031)	(0.00032)	(0.00491)	(0.00155)	(0.00031)	(0.00032)
Hrs_si	0.00239	0.00115*	0.00041***	-0.00009	0.00226	0.00113*	0.00041***	-0.00008	0.00226	0.00113*	0.00041***	-0.00007
	(0.00160)	(0.00060)	(0.00011)	(0.00011)	(0.00162)	(0.00060)	(0.00011)	(0.00011)	(0.00162)	(0.00060)	(0.00011)	(0.00011)
Hrs_fam	-0.00543**	-0.00086	0.00031	0.00061**	-0.00517**	-0.00090	0.00031	0.00060**	-0.00517**	-0.00090	0.00030	0.00061**
	(0.00218)	(0.00111)	(0.00023)	(0.00028)	(0.00223)	(0.00112)	(0.00023)	(0.00028)	(0.00223)	(0.00112)	(0.00023)	(0.00028)
Hrs_second	0.00740	0.00009	0.00006	-0.00026	0.00656	-0.00010	0.00008	-0.00024	0.00656	-0.00010	0.00009	-0.00025
	(0.00486)	(0.00160)	(0.00036)	(0.00031)	(0.00495)	(0.00160)	(0.00036)	(0.00031)	(0.00495)	(0.00160)	(0.00036)	(0.00031)
D(Hrs_ff)	0.16752	-0.04699	0.00956	0.00755	0.17409	-0.04611	0.00893	0.00732	0.17409	-0.04611	0.00878	0.00772
	(0.18254)	(0.06669)	(0.01273)	(0.01447)	(0.17599)	(0.06647)	(0.01270)	(0.01445)	(0.17599)	(0.06647)	(0.01271)	(0.01444)
D(Hrs_if)	0.07383	0.08037	-0.01936	-0.03348*	0.09041	0.07711	-0.01889	-0.03282*	0.09041	0.07711	-0.01886	-0.03309*
	(0.22167)	(0.10161)	(0.01945)	(0.01710)	(0.21040)	(0.10160)	(0.01942)	(0.01716)	(0.21040)	(0.10160)	(0.01941)	(0.01708)
D(Hrs_ii)	0.14945	0.03365	0.01291	-0.01147	0.15223	0.03578	0.01241	-0.01120	0.15223	0.03578	0.01245	-0.01077
	(0.12965)	(0.05587)	(0.01133)	(0.01182)	(0.12696)	(0.05583)	(0.01131)	(0.01182)	(0.12696)	(0.05583)	(0.01131)	(0.01180)
D(Hrs_sf)	-0.26883	0.09119	-0.07591***	-0.00656	-0.26919	0.09041	-0.07466***	-0.00641	-0.26919	0.09041	-0.07526***	-0.00584
	(0.27401)	(0.09877)	(0.02049)	(0.02121)	(0.26821)	(0.09839)	(0.02045)	(0.02093)	(0.26821)	(0.09839)	(0.02046)	(0.02096)
D(Hrs_si)	0.02810	-0.08212*	0.00859	-0.01573	0.03676	-0.08468*	0.00827	-0.01584	0.03676	-0.08468*	0.00812	-0.01636
	(0.11528)	(0.04772)	(0.00941)	(0.01018)	(0.11666)	(0.04771)	(0.00939)	(0.01021)	(0.11666)	(0.04771)	(0.00939)	(0.01018)
D(Hrs_fam)	0.06126	0.00266	-0.00023	0.00215	0.07843	0.00380	-0.00067	0.00254	0.07843	0.00380	-0.00076	0.00264

	(0.15509)	(0.06127)	(0.01230)	(0.01260)	(0.15723)	(0.06134)	(0.01226)	(0.01260)	(0.15723)	(0.06134)	(0.01226)	(0.01258)
D(Hrs_second)	0.09138	0.06912	-0.00894	0.01292	0.10310	0.07213	-0.00979	0.01291	0.10310	0.07213	-0.00967	0.01307
	(0.13267)	(0.05139)	(0.01092)	(0.01135)	(0.13484)	(0.05136)	(0.01090)	(0.01133)	(0.13484)	(0.05136)	(0.01090)	(0.01130)
P(pop_slums_dist)	0.00435	0.04205			-0.04430	0.07969			-0.04430	0.07969		
	(0.12911)	(0.05864)			(0.12734)	(0.05740)			(0.12734)	(0.05740)		
P(migrant_dist)	-0.33160	-0.18028			0.18024	-0.24649			0.18024	-0.24649		
	(0.50340)	(0.25395)			(0.51392)	(0.25037)			(0.51392)	(0.25037)		
P(race_dist)	0.47875	-0.13558			0.42377	-0.28657			0.42377	-0.28657		
	(0.44773)	(0.25739)			(0.42861)	(0.25057)			(0.42861)	(0.25057)		
informal_index	-0.01239	-0.14907**			-0.05393	-0.12182*			-0.05393	-0.12182*		
	(0.16055)	(0.07213)			(0.14623)	(0.07013)			(0.14623)	(0.07013)		
D(conf_gov_1)	0.03660	-0.05564			0.01997	-0.05265			0.01997	-0.05265		
	(0.09435)	(0.03945)			(0.09246)	(0.03936)			(0.09246)	(0.03936)		
D(conf_gov_2)	-0.10364	-0.13331**			-0.10316	-0.13439**			-0.10316	-0.13439**		
	(0.10904)	(0.05477)			(0.10850)	(0.05469)			(0.10850)	(0.05469)		
D(conf_gov_3)	-0.02704	0.13473			0.00501	0.12391			0.00501	0.12391		
	(0.16801)	(0.08557)			(0.16659)	(0.08521)			(0.16659)	(0.08521)		
D(overall_satis)	0.09355	-0.10404***			0.08940	-0.10733***			0.08940	-0.10733***		
	(0.08222)	(0.03661)			(0.08074)	(0.03652)			(0.08074)	(0.03652)		
D(income_satis)	0.09273	-0.08278**			0.10792	-0.08179**			0.10792	-0.08179**		
	(0.08497)	(0.03716)			(0.08444)	(0.03710)			(0.08444)	(0.03710)		
D(income_stab)	0.02039	-0.13682***			0.02924	-0.13535***			0.02924	-0.13535***		
	(0.09370)	(0.04156)			(0.09282)	(0.04148)			(0.09282)	(0.04148)		
D(sub_pov)	0.00478	-0.08594**			0.00182	-0.08120**			0.00182	-0.08120**		
	(0.07729)	(0.03588)			(0.07602)	(0.03584)			(0.07602)	(0.03584)		
D(road_1)	-0.23039*	-0.03566			-0.19268	-0.04821			-0.19268	-0.04821		
	(0.13735)	(0.06515)			(0.12846)	(0.06417)			(0.12846)	(0.06417)		
D(road_2)	-0.26414**	0.12411**			-0.20665	0.11462**			-0.20665	0.11462**		
	(0.13029)	(0.05850)			(0.12780)	(0.05798)			(0.12780)	(0.05798)		
D(road_3)	0.14714	0.09387			0.17902	0.12951			0.17902	0.12951		

	(0.29198)	(0.13173)	(0.29140)	(0.12928)	(0.29140)	(0.12928)
distance	-0.00067**	-0.00004	-0.00058**	-0.00003	-0.00058**	-0.00003
	(0.00030)	(0.00018)	(0.00029)	(0.00018)	(0.00029)	(0.00018)
P(nowater_dist)	0.45894*	-0.08667	0.53204**	-0.11970	0.53204**	-0.11970
	(0.25424)	(0.14485)	(0.24867)	(0.14239)	(0.24867)	(0.14239)
P(nodrain_dist)	-0.68376*	-0.40585**	-0.95589***	-0.36117*	-0.95589***	-0.36117*
	(0.35477)	(0.20316)	(0.34665)	(0.19988)	(0.34665)	(0.19988)
P(noelec_dist)	1.54315*	-1.51536***	2.07547***	-1.58577***	2.07547***	-1.58577***
	(0.79505)	(0.43561)	(0.74530)	(0.43348)	(0.74530)	(0.43348)
P(nofloor_dist)	-0.00764	0.29302	-0.12444	0.26527	-0.12444	0.26527
	(0.52991)	(0.24569)	(0.52026)	(0.24438)	(0.52026)	(0.24438)
P(nowall_dist)	0.11860	0.59131***	0.31744	0.56225***	0.31744	0.56225***
	(0.35396)	(0.14398)	(0.34226)	(0.14214)	(0.34226)	(0.14214)
P(noinsu_dist)	0.90051*	-0.17580	1.28200**	-0.31394	1.28200**	-0.31394
	(0.50944)	(0.27385)	(0.52609)	(0.26523)	(0.52609)	(0.26523)
P(illiterate_dist)	-5.05595**	1.59883	-3.75728	1.58955	-3.75728	1.58955
	(2.55669)	(1.44885)	(2.49769)	(1.41485)	(2.49769)	(1.41485)
P(selfwork_dist)	0.86008	0.15528	0.95424	0.20626	0.95424	0.20626
	(0.83362)	(0.45045)	(0.84814)	(0.44523)	(0.84814)	(0.44523)
P(unreg_dist)	8.06918**	-4.34551**	9.62524**	-5.29650***	9.62524**	-5.29650***
	(3.92492)	(2.13814)	(3.76598)	(2.04666)	(3.76598)	(2.04666)
icf_index	-0.73464	0.15128	-1.11717**	0.35269	-1.11717**	0.35269
	(0.49280)	(0.27225)	(0.46755)	(0.26136)	(0.46755)	(0.26136)
P(undernut_dist)	-0.02108	0.66342**	0.05551	0.72415**	0.05551	0.72415**
	(0.62776)	(0.29516)	(0.62044)	(0.29309)	(0.62044)	(0.29309)
Mort_distx1000	-0.01352	0.01922	-0.01036	0.01884	-0.01036	0.01884
	(0.02280)	(0.01772)	(0.01948)	(0.01663)	(0.01948)	(0.01663)
P(poor_dist)	2.20941***	-0.84373***	2.57002***	-1.07810***	2.57002***	-1.07810***
	(0.69467)	(0.29682)	(0.62227)	(0.28032)	(0.62227)	(0.28032)
lifeexp_dist	0.05121	0.05520	0.06327	0.05670	0.06327	0.05670

	(0.08275)	(0.05653)			(0.07323)	(0.05339)			(0.07323)	(0.05339)		
D(ca_food)_1	0.08140	0.00538			0.07998	0.01147			0.07998	0.01147		
	(0.10018)	(0.05055)			(0.09924)	(0.05050)			(0.09924)	(0.05050)		
D(ca_food)_2	0.01266	0.08188			0.02962	0.07169			0.02962	0.07169		
	(0.10676)	(0.05183)			(0.10721)	(0.05172)			(0.10721)	(0.05172)		
D(ca_food)_3	0.17974	0.01083			0.19070*	0.00564			0.19070*	0.00564		
	(0.11189)	(0.04951)			(0.11247)	(0.04961)			(0.11247)	(0.04961)		
D(ca_food)_4	-0.09157	-0.03135			-0.10697	-0.02248			-0.10697	-0.02248		
	(0.10439)	(0.04803)			(0.10445)	(0.04806)			(0.10445)	(0.04806)		
Agropec_firmPC	-0.04845	-0.03361*			-0.05484	-0.03096			-0.05484	-0.03096		
	(0.04567)	(0.01915)			(0.04315)	(0.01899)			(0.04315)	(0.01899)		
D(agro_dist)	0.08665	0.01637			0.13768	0.02670			0.13768	0.02670		
	(0.10417)	(0.04424)			(0.10168)	(0.04389)			(0.10168)	(0.04389)		
D(pec_dist)	-0.04715	-0.03262			-0.06448	-0.05682			-0.06448	-0.05682		
	(0.11776)	(0.05156)			(0.11543)	(0.05052)			(0.11543)	(0.05052)		
D(fish_dist)	0.40962**	-0.00135			0.43587**	-0.00207			0.43587**	-0.00207		
	(0.18479)	(0.06186)			(0.18180)	(0.06171)			(0.18180)	(0.06171)		
Prop(workagro_dist)	-1.08520**	-0.85901***			-1.14757**	-0.80565***			-1.14757**	-0.80565***		
	(0.50443)	(0.25036)			(0.50026)	(0.24798)			(0.50026)	(0.24798)		
Prop(workfish_dist)	1.55234	2.36456***			1.22674	2.36604***			1.22674	2.36604***		
	(2.06721)	(0.90970)			(1.77318)	(0.91223)			(1.77318)	(0.91223)		
PHI			-0.15120**	-0.07075***			-0.18362***	-0.06449***			-0.17783***	-0.05812**
			(0.06453)	(0.02329)			(0.06559)	(0.02291)			(0.06553)	(0.02297)
Constant	-15.36639*	-4.71809	0.36071***	0.32460***	-18.60654***	-3.53289	0.35369***	0.31885***	-18.60654***	-3.53289	0.33473***	0.32572***
	(8.01587)	(5.15879)	(0.06549)	(0.11123)	(7.09483)	(4.81140)	(0.06451)	(0.11152)	(7.09483)	(4.81140)	(0.06458)	(0.11085)
Observations	7,513	7,513	7,513		7,513	7,513	7,513		7,513	7,513	7,513	
LogL	-685.106	-4256.626	4096.869		-697.076	-4266.134	4096.895		-697.076	-4266.134	4105.179	
Wald (ident. vars.)	114.40***	231.61***			176.93***	252.03***			176.93***	252.03***		

d) Results for models under specification 2

	Restriction 1				Restriction 2				Restriction 3			
	PROBIT		SUR		PROBIT		SUR		PROBIT		SUR	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expend_Fon)					0.56793***	0.54547***			0.56793***	0.54547***		
					(0.03893)	(0.03078)			(0.03893)	(0.03078)		
O1L(expend_Fon)	0.57234***	0.55621***	-0.00403	0.02167			-0.00413	0.02023			-0.00386	0.02592**
	(0.03899)	(0.03113)	(0.00692)	(0.01341)			(0.00689)	(0.01284)			(0.00688)	(0.01285)
L(O1_uv_inf)	-0.43002	0.41347**	-0.14266***	0.07048***			-0.16086***	0.07166***			-0.18210***	0.15595***
	(0.26255)	(0.17283)	(0.02288)	(0.00904)			(0.02198)	(0.00908)			(0.02243)	(0.01813)
L(O1_uv_semi)	0.65799	0.37183*	0.07218***	-0.01598			0.08920***	0.00920			0.10599***	-0.06804***
	(0.41824)	(0.19654)	(0.02338)	(0.01444)			(0.02266)	(0.01199)			(0.02204)	(0.01873)
L(O1_uv_for)	0.10879	-0.47852***	0.07048***	-0.05449***			0.07166***	-0.08086***			0.07611***	-0.08791***
	(0.10966)	(0.06178)	(0.00904)	(0.01296)			(0.00908)	(0.00997)			(0.00885)	(0.00974)
L(fam_size)	0.49312***	0.14936**	0.00610	-0.05164***	0.48937***	0.14426**	0.00526	-0.05152***	0.48937***	0.14426**	0.00517	-0.04992***
	(0.13953)	(0.05926)	(0.01216)	(0.01310)	(0.13752)	(0.05902)	(0.01217)	(0.01304)	(0.13752)	(0.05902)	(0.01217)	(0.01302)
D(mem_spo)	0.11768	-0.08756	0.04101***	-0.01982	0.10531	-0.09983*	0.04140***	-0.01878	0.10531	-0.09983*	0.04170***	-0.02031
	(0.13382)	(0.06073)	(0.01239)	(0.01412)	(0.13492)	(0.06041)	(0.01239)	(0.01405)	(0.13492)	(0.06041)	(0.01238)	(0.01391)
D(mem_son)	0.22754*	0.03167	0.02109	-0.03129**	0.22602*	0.03926	0.02137	-0.03297**	0.22602*	0.03926	0.02170	-0.03306**
	(0.13605)	(0.06399)	(0.01346)	(0.01518)	(0.13525)	(0.06370)	(0.01347)	(0.01514)	(0.13525)	(0.06370)	(0.01346)	(0.01505)
D(mem_oth)	0.00575	0.02217	0.01311	0.00139	0.00105	0.02582	0.01355	0.00118	0.00105	0.02582	0.01405	0.00041
	(0.11976)	(0.04562)	(0.00889)	(0.00907)	(0.11869)	(0.04544)	(0.00890)	(0.00906)	(0.11869)	(0.04544)	(0.00889)	(0.00899)
P(men)	-0.33054*	0.07431	-0.07637***	0.02105	-0.31244*	0.06263	-0.07572***	0.02083	-0.31244*	0.06263	-0.07568***	0.02116
	(0.17103)	(0.07977)	(0.01666)	(0.02019)	(0.17178)	(0.07927)	(0.01665)	(0.02024)	(0.17178)	(0.07927)	(0.01665)	(0.02010)
age_head	0.01515	-0.00594	0.00629***	-0.00323*	0.01654	-0.00487	0.00632***	-0.00341*	0.01654	-0.00487	0.00632***	-0.00338*
	(0.01391)	(0.00732)	(0.00154)	(0.00189)	(0.01386)	(0.00722)	(0.00155)	(0.00188)	(0.01386)	(0.00722)	(0.00155)	(0.00186)
age_head-sq	-0.00005	0.00007	-0.00004***	0.00003*	-0.00006	0.00006	-0.00004***	0.00003*	-0.00006	0.00006	-0.00004***	0.00003*
	(0.00014)	(0.00007)	(0.00001)	(0.00002)	(0.00014)	(0.00007)	(0.00001)	(0.00002)	(0.00014)	(0.00007)	(0.00001)	(0.00002)
D(gender_head)	-0.03094	0.13387**	-0.02365*	0.00759	-0.03568	0.15777**	-0.02385*	0.00561	-0.03568	0.15777**	-0.02390*	0.00792

	(0.14013)	(0.06448)	(0.01296)	(0.01545)	(0.14191)	(0.06403)	(0.01297)	(0.01540)	(0.14191)	(0.06403)	(0.01295)	(0.01523)
year_educ_head	-0.01129	0.03157***	-0.00186**	0.00654***	-0.01179	0.02991***	-0.00186**	0.00655***	-0.01179	0.02991***	-0.00184**	0.00686***
	(0.01010)	(0.00440)	(0.00085)	(0.00114)	(0.01008)	(0.00438)	(0.00085)	(0.00111)	(0.01008)	(0.00438)	(0.00085)	(0.00110)
Hrs_ff	-0.00123	0.00220**	0.00019	0.00018	-0.00121	0.00205**	0.00020	0.00019	-0.00121	0.00205**	0.00020	0.00020
	(0.00247)	(0.00092)	(0.00017)	(0.00019)	(0.00247)	(0.00091)	(0.00017)	(0.00019)	(0.00247)	(0.00091)	(0.00017)	(0.00018)
Hrs_if	-0.00310	-0.00155	0.00035	0.00044	-0.00290	-0.00151	0.00035	0.00045	-0.00290	-0.00151	0.00035	0.00044
	(0.00308)	(0.00165)	(0.00032)	(0.00028)	(0.00318)	(0.00165)	(0.00032)	(0.00028)	(0.00318)	(0.00165)	(0.00032)	(0.00028)
Hrs_ii	-0.00129	-0.00110	-0.00009	0.00010	-0.00112	-0.00122*	-0.00008	0.00010	-0.00112	-0.00122*	-0.00009	0.00009
	(0.00162)	(0.00072)	(0.00015)	(0.00014)	(0.00162)	(0.00072)	(0.00015)	(0.00014)	(0.00162)	(0.00072)	(0.00015)	(0.00014)
Hrs_sf	0.00676	0.00133	0.00152***	0.00008	0.00649	0.00110	0.00151***	0.00011	0.00649	0.00110	0.00151***	0.00011
	(0.00519)	(0.00158)	(0.00030)	(0.00030)	(0.00499)	(0.00156)	(0.00030)	(0.00031)	(0.00499)	(0.00156)	(0.00030)	(0.00031)
Hrs_si	0.00219	0.00129**	0.00030***	0.00001	0.00227	0.00116*	0.00030***	0.00002	0.00227	0.00116*	0.00031***	0.00003
	(0.00167)	(0.00059)	(0.00011)	(0.00011)	(0.00168)	(0.00059)	(0.00011)	(0.00011)	(0.00168)	(0.00059)	(0.00011)	(0.00011)
Hrs_fam	-0.00482**	-0.00065	0.00024	0.00046*	-0.00496**	-0.00079	0.00024	0.00048*	-0.00496**	-0.00079	0.00024	0.00047*
	(0.00229)	(0.00112)	(0.00022)	(0.00028)	(0.00225)	(0.00112)	(0.00022)	(0.00028)	(0.00225)	(0.00112)	(0.00022)	(0.00028)
Hrs_second	0.00755	0.00002	0.00000	-0.00023	0.00742	-0.00025	0.00001	-0.00021	0.00742	-0.00025	0.00002	-0.00025
	(0.00516)	(0.00161)	(0.00036)	(0.00030)	(0.00503)	(0.00160)	(0.00036)	(0.00030)	(0.00503)	(0.00160)	(0.00036)	(0.00030)
D(Hrs_ff)	0.17373	-0.04439	0.01079	0.00872	0.18640	-0.04484	0.01064	0.00835	0.18640	-0.04484	0.01021	0.00994
	(0.17576)	(0.06688)	(0.01242)	(0.01382)	(0.17508)	(0.06656)	(0.01242)	(0.01378)	(0.17508)	(0.06656)	(0.01242)	(0.01371)
D(Hrs_if)	0.15532	0.07651	-0.02133	-0.02907*	0.12424	0.07294	-0.02109	-0.02868*	0.12424	0.07294	-0.02076	-0.02936*
	(0.20119)	(0.10245)	(0.01922)	(0.01693)	(0.20245)	(0.10228)	(0.01922)	(0.01684)	(0.20245)	(0.10228)	(0.01921)	(0.01676)
D(Hrs_ii)	0.17513	0.03541	0.01347	-0.01425	0.16764	0.03480	0.01317	-0.01332	0.16764	0.03480	0.01312	-0.01265
	(0.12668)	(0.05595)	(0.01121)	(0.01151)	(0.12632)	(0.05593)	(0.01121)	(0.01151)	(0.12632)	(0.05593)	(0.01121)	(0.01147)
D(Hrs_sf)	-0.26682	0.08637	-0.05433***	-0.00832	-0.25699	0.09116	-0.05379***	-0.00905	-0.25699	0.09116	-0.05389***	-0.00795
	(0.27817)	(0.09980)	(0.01965)	(0.02041)	(0.26972)	(0.09882)	(0.01966)	(0.02066)	(0.26972)	(0.09882)	(0.01966)	(0.02060)
D(Hrs_si)	0.03738	-0.07696	0.00658	-0.01629	0.03545	-0.08259*	0.00652	-0.01546	0.03545	-0.08259*	0.00620	-0.01616
	(0.11783)	(0.04791)	(0.00924)	(0.00995)	(0.11785)	(0.04774)	(0.00923)	(0.00997)	(0.11785)	(0.04774)	(0.00923)	(0.00995)
D(Hrs_fam)	0.09423	-0.00736	0.00634	0.00159	0.09397	0.00717	0.00621	0.00193	0.09397	0.00717	0.00606	0.00241
	(0.16043)	(0.06145)	(0.01196)	(0.01246)	(0.15924)	(0.06152)	(0.01194)	(0.01232)	(0.15924)	(0.06152)	(0.01195)	(0.01233)
D(Hrs_second)	0.07048	0.07028	-0.00344	0.01590	0.08182	0.07719	-0.00341	0.01537	0.08182	0.07719	-0.00335	0.01562

	(0.13714)	(0.05179)	(0.01079)	(0.01102)	(0.13490)	(0.05154)	(0.01078)	(0.01096)	(0.13490)	(0.05154)	(0.01078)	(0.01090)
L(pop_urb_dist)	0.18155	0.25047	0.18784***	0.21534***	0.16208	0.01340	0.18181***	0.23042***	0.16208	0.01340	0.18209***	0.23033***
	(0.38129)	(0.21676)	(0.03829)	(0.04899)	(0.36378)	(0.21323)	(0.03843)	(0.04778)	(0.36378)	(0.21323)	(0.03850)	(0.04704)
L(pop_urb_dist)-sq	0.00072	-0.00830	-0.00709***	-0.01050***	0.00220	0.00333	-0.00685***	-0.01123***	0.00220	0.00333	-0.00688***	-0.01112***
	(0.01790)	(0.00993)	(0.00180)	(0.00227)	(0.01723)	(0.00978)	(0.00181)	(0.00223)	(0.01723)	(0.00978)	(0.00181)	(0.00220)
pop_den_distx1000	0.02628***	0.01803***	-0.00008	-0.00214**	0.02547***	0.01776***	-0.00006	-0.00243***	0.02547***	0.01776***	-0.00003	-0.00235***
	(0.00963)	(0.00524)	(0.00086)	(0.00086)	(0.00947)	(0.00523)	(0.00086)	(0.00087)	(0.00947)	(0.00523)	(0.00086)	(0.00087)
D(not_slum)	0.08483	0.22584***	0.03067***	0.04485***	0.08143	0.21698***	0.03052***	0.04461***	0.08143	0.21698***	0.02994***	0.04837***
	(0.07943)	(0.03520)	(0.00669)	(0.00759)	(0.07956)	(0.03503)	(0.00669)	(0.00736)	(0.07956)	(0.03503)	(0.00670)	(0.00736)
D(Lima)	-1.00387***	0.38016***	0.05232***	0.17512***	-1.02283***	0.42623***	0.05340***	0.17398***	-1.02283***	0.42623***	0.05466***	0.17270***
	(0.22359)	(0.10003)	(0.01810)	(0.01968)	(0.22161)	(0.09726)	(0.01811)	(0.01958)	(0.22161)	(0.09726)	(0.01806)	(0.01933)
D(mid_city)	-0.25502**	-0.02502	-0.05375***	0.03760***	-0.19727	-0.04928	-0.05405***	0.03948***	-0.19727	-0.04928	-0.05471***	0.04119***
	(0.12291)	(0.04868)	(0.00771)	(0.00740)	(0.12197)	(0.04826)	(0.00770)	(0.00738)	(0.12197)	(0.04826)	(0.00772)	(0.00738)
D(border)	0.58006***	0.05490	0.08556***	-0.00017	0.52497***	0.03690	0.08752***	-0.00222	0.52497***	0.03690	0.08769***	-0.00651
	(0.17063)	(0.06714)	(0.01037)	(0.01072)	(0.16726)	(0.06660)	(0.01037)	(0.01044)	(0.16726)	(0.06660)	(0.01035)	(0.01045)
P(pop_slums_dist)	0.13348	0.04729			0.07333	0.12714**			0.07333	0.12714**		
	(0.13105)	(0.06044)			(0.12932)	(0.05889)			(0.12932)	(0.05889)		
P(migrant_dist)	0.04317	-0.36816			0.16872	-0.35268			0.16872	-0.35268		
	(0.54610)	(0.25589)			(0.52679)	(0.25371)			(0.52679)	(0.25371)		
P(race_dist)	0.46426	0.77796***			0.65923	-0.03784			0.65923	-0.03784		
	(0.48850)	(0.28465)			(0.45177)	(0.26058)			(0.45177)	(0.26058)		
informal_index	0.14026	-0.13760*			0.06315	-0.11752			0.06315	-0.11752		
	(0.15287)	(0.07376)			(0.14959)	(0.07266)			(0.14959)	(0.07266)		
D(conf_gov_1)	0.04823	-0.05086			0.03788	-0.05635			0.03788	-0.05635		
	(0.09415)	(0.03970)			(0.09351)	(0.03962)			(0.09351)	(0.03962)		
D(conf_gov_2)	-0.08143	-0.11766**			-0.08303	-0.12564**			-0.08303	-0.12564**		
	(0.11022)	(0.05529)			(0.10982)	(0.05481)			(0.10982)	(0.05481)		
D(conf_gov_3)	-0.01495	0.13814			0.01211	0.13716			0.01211	0.13716		
	(0.16799)	(0.08649)			(0.16916)	(0.08539)			(0.16916)	(0.08539)		
D(overall_satis)	0.08208	-0.10720***			0.08264	-0.11250***			0.08264	-0.11250***		

	(0.08202)	(0.03694)	(0.08151)	(0.03674)	(0.08151)	(0.03674)
D(income_satis)	0.10153	-0.09087**	0.09875	-0.08747**	0.09875	-0.08747**
	(0.08523)	(0.03741)	(0.08470)	(0.03726)	(0.08470)	(0.03726)
D(income_stab)	0.05501	-0.13181***	0.04372	-0.12945***	0.04372	-0.12945***
	(0.09433)	(0.04198)	(0.09406)	(0.04170)	(0.09406)	(0.04170)
D(sub_pov)	0.03189	-0.08428**	0.03239	-0.07904**	0.03239	-0.07904**
	(0.07659)	(0.03617)	(0.07668)	(0.03607)	(0.07668)	(0.03607)
D(road_1)	-0.17669	0.01420	-0.16088	-0.01611	-0.16088	-0.01611
	(0.13517)	(0.06758)	(0.13192)	(0.06587)	(0.13192)	(0.06587)
D(road_2)	-0.23954*	0.09331	-0.20670	0.09804*	-0.20670	0.09804*
	(0.13404)	(0.05941)	(0.13173)	(0.05903)	(0.13173)	(0.05903)
D(road_3)	0.13603	0.16339	0.14675	0.09831	0.14675	0.09831
	(0.27725)	(0.12998)	(0.28032)	(0.13038)	(0.28032)	(0.13038)
distance	-0.00038	0.00015	-0.00061**	0.00002	-0.00061**	0.00002
	(0.00031)	(0.00019)	(0.00030)	(0.00018)	(0.00030)	(0.00018)
P(nowater_dist)	0.37646	-0.05611	0.45152*	-0.04571	0.45152*	-0.04571
	(0.24508)	(0.15083)	(0.24143)	(0.14545)	(0.24143)	(0.14545)
P(nodrain_dist)	-0.40866	-0.10745	-0.54722	-0.17560	-0.54722	-0.17560
	(0.35377)	(0.21429)	(0.34254)	(0.20594)	(0.34254)	(0.20594)
P(noelec_dist)	2.00830***	-1.39756***	2.11415***	-1.47745***	2.11415***	-1.47745***
	(0.74533)	(0.43898)	(0.70396)	(0.43697)	(0.70396)	(0.43697)
P(nofloor_dist)	-0.23299	0.42066	-0.35804	0.21996	-0.35804	0.21996
	(0.52876)	(0.25606)	(0.51860)	(0.25088)	(0.51860)	(0.25088)
P(nowall_dist)	0.62711*	0.59360***	0.66623*	0.66285***	0.66623*	0.66285***
	(0.35189)	(0.14941)	(0.34564)	(0.14814)	(0.34564)	(0.14814)
P(noinsu_dist)	0.91573	-0.20988	0.92120*	-0.49536*	0.92120*	-0.49536*
	(0.57069)	(0.27913)	(0.55375)	(0.26721)	(0.55375)	(0.26721)
P(illiterate_dist)	-3.38387	-1.63750	-3.56749	0.94056	-3.56749	0.94056
	(2.52850)	(1.47813)	(2.48691)	(1.41620)	(2.48691)	(1.41620)
P(selfwork_dist)	-0.06616	0.05098	0.25366	0.04043	0.25366	0.04043

	(0.89582)	(0.47198)	(0.88597)	(0.46140)	(0.88597)	(0.46140)
P(unreg_dist)	8.77300**	-4.26241*	9.99294***	-4.91631**	9.99294***	-4.91631**
	(3.85954)	(2.20736)	(3.78599)	(2.09061)	(3.78599)	(2.09061)
icf_index	-0.81775*	0.00606	-1.15007**	0.30619	-1.15007**	0.30619
	(0.49072)	(0.28829)	(0.48443)	(0.27289)	(0.48443)	(0.27289)
P(undernut_dist)	0.01472	0.97452***	0.40261	0.68232**	0.40261	0.68232**
	(0.64082)	(0.30330)	(0.61900)	(0.29479)	(0.61900)	(0.29479)
Mort_distx1000	-0.01695	0.00143	-0.02533	0.01261	-0.02533	0.01261
	(0.02076)	(0.01807)	(0.02178)	(0.01633)	(0.02178)	(0.01633)
P(poor_dist)	2.24291***	-0.96634***	2.57869***	-0.93299***	2.57869***	-0.93299***
	(0.67626)	(0.30884)	(0.62557)	(0.28818)	(0.62557)	(0.28818)
lifeexp_dist	0.04698	0.00457	0.02094	0.04426	0.02094	0.04426
	(0.07501)	(0.05756)	(0.07931)	(0.05244)	(0.07931)	(0.05244)
D(ca_food)_1	0.07524	0.00236	0.07092	0.01058	0.07092	0.01058
	(0.10058)	(0.05081)	(0.09993)	(0.05064)	(0.09993)	(0.05064)
D(ca_food)_2	-0.01976	0.07819	0.01006	0.07087	0.01006	0.07087
	(0.11097)	(0.05211)	(0.10944)	(0.05194)	(0.10944)	(0.05194)
D(ca_food)_3	0.21353*	0.00847	0.20609*	0.00322	0.20609*	0.00322
	(0.11284)	(0.04962)	(0.11324)	(0.04973)	(0.11324)	(0.04973)
D(ca_food)_4	-0.11771	-0.03176	-0.14236	-0.03230	-0.14236	-0.03230
	(0.10437)	(0.04831)	(0.10367)	(0.04829)	(0.10367)	(0.04829)
Agropec_firmPC	-0.06710*	-0.02740	-0.07744*	-0.03276*	-0.07744*	-0.03276*
	(0.03981)	(0.01936)	(0.03959)	(0.01957)	(0.03959)	(0.01957)
D(agro_dist)	0.11622	0.07236	0.13781	0.07198	0.13781	0.07198
	(0.10039)	(0.04625)	(0.10044)	(0.04607)	(0.10044)	(0.04607)
D(pec_dist)	0.03277	-0.05006	0.03172	-0.03075	0.03172	-0.03075
	(0.10961)	(0.05185)	(0.11195)	(0.05106)	(0.11195)	(0.05106)
D(fish_dist)	0.43862**	0.05437	0.52917***	0.02677	0.52917***	0.02677
	(0.19019)	(0.06499)	(0.19065)	(0.06401)	(0.19065)	(0.06401)
Prop(workagro_dist)	-0.75432	-0.72507**	-0.67198	-0.61843**	-0.67198	-0.61843**

	(0.59870)	(0.30289)			(0.57666)	(0.29199)			(0.57666)	(0.29199)		
Prop(workfish_dist)	0.68483	2.22590**			1.09289	2.78119***			1.09289	2.78119***		
	(1.50168)	(0.92672)			(1.48801)	(0.92916)			(1.48801)	(0.92916)		
PHI			0.00838	0.02003			0.00372	0.01480			0.00455	0.03488
			(0.06445)	(0.02535)			(0.06364)	(0.02343)			(0.06361)	(0.02338)
Constant	-18.02501**	-1.70452	-0.84322***	-1.03329***	-17.48736**	-3.40827	-0.81110***	-1.07363***	-17.48736**	-3.40827	-0.82102***	-1.12134***
	(7.92250)	(5.42370)	(0.21207)	(0.30136)	(7.93998)	(5.05107)	(0.21262)	(0.28616)	(7.93998)	(5.05107)	(0.21269)	(0.28160)
Observations	7,513	7,513	7,513		7,513	7,513	7,513		7,513	7,513	7,513	
LogL	-680.086	-4197.159	4367.198		-684.824	-4235.236	4373.806		-684.824	-4235.236	4394.748	
Wald (ident. vars.)	88.05***	166.90***			141.31***	197.68***			141.31***	197.68***		

e) Results for models under specification 3

	Restriction 1				Restriction 2				Restriction 3			
	PROBIT		SUR		PROBIT		SUR		PROBIT		SUR	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expend_Fon)					0.56935***	0.55005***			0.56935***	0.55005***		
					(0.03887)	(0.03065)			(0.03887)	(0.03065)		
O1L(expend_Fon)	0.57274***	0.56194***	-0.00604	0.00013			-0.00686	-0.00088			-0.00714	0.00267
	(0.03896)	(0.03101)	(0.00718)	(0.01340)			(0.00716)	(0.01326)			(0.00715)	(0.01324)
L(O1_uv_inf)	-0.32654	0.36301**	-0.13773***	0.07148***			-0.16374***	0.07403***			-0.18492***	0.15421***
	(0.27029)	(0.17142)	(0.02529)	(0.00977)			(0.02431)	(0.00975)			(0.02490)	(0.01952)
L(O1_uv_semi)	0.65088	0.41853**	0.06625***	-0.05267***			0.08972***	-0.01466			0.10964***	-0.08832***
	(0.42096)	(0.19344)	(0.02517)	(0.01309)			(0.02438)	(0.01229)			(0.02391)	(0.01963)
L(O1_uv_for)	0.10340	-0.45633***	0.07148***	-0.01881*			0.07403***	-0.05937***			0.07528***	-0.06589***
	(0.11234)	(0.06107)	(0.00977)	(0.01111)			(0.00975)	(0.00990)			(0.00953)	(0.00978)
L(fam_size)	0.47660***	0.15974***	-0.00374	-0.05943***	0.47685***	0.15659***	-0.00508	-0.05989***	0.47685***	0.15659***	-0.00543	-0.05834***
	(0.13703)	(0.05922)	(0.01238)	(0.01343)	(0.13496)	(0.05896)	(0.01237)	(0.01347)	(0.13496)	(0.05896)	(0.01237)	(0.01345)
D(mem_spo)	0.10314	-0.08620	0.02963**	-0.01487	0.08886	-0.09822	0.03012**	-0.01416	0.08886	-0.09822	0.03049**	-0.01569
	(0.13086)	(0.06103)	(0.01257)	(0.01435)	(0.13179)	(0.06071)	(0.01257)	(0.01435)	(0.13179)	(0.06071)	(0.01256)	(0.01420)
D(mem_son)	0.23283*	0.02750	0.02378*	-0.03295**	0.23110*	0.03218	0.02385*	-0.03450**	0.23110*	0.03218	0.02392*	-0.03472**

	(0.13511)	(0.06370)	(0.01372)	(0.01565)	(0.13415)	(0.06343)	(0.01373)	(0.01569)	(0.13415)	(0.06343)	(0.01372)	(0.01559)
D(mem_oth)	0.01391	0.01329	0.01552*	-0.00114	0.00789	0.01583	0.01587*	-0.00081	0.00789	0.01583	0.01621*	-0.00151
	(0.11834)	(0.04556)	(0.00905)	(0.00922)	(0.11746)	(0.04542)	(0.00906)	(0.00924)	(0.11746)	(0.04542)	(0.00905)	(0.00916)
P(men)	-0.31984*	0.07751	-0.07081***	0.01700	-0.30816*	0.06664	-0.06985***	0.01669	-0.30816*	0.06664	-0.06960***	0.01640
	(0.16781)	(0.07947)	(0.01691)	(0.02049)	(0.16856)	(0.07900)	(0.01691)	(0.02061)	(0.16856)	(0.07900)	(0.01690)	(0.02047)
age_head	0.01734	-0.00454	0.00573***	-0.00353*	0.01810	-0.00293	0.00574***	-0.00384**	0.01810	-0.00293	0.00572***	-0.00376**
	(0.01387)	(0.00733)	(0.00157)	(0.00190)	(0.01383)	(0.00723)	(0.00158)	(0.00189)	(0.01383)	(0.00723)	(0.00158)	(0.00188)
age_head-sq	-0.00007	0.00007	-0.00004**	0.00004**	-0.00008	0.00006	-0.00004**	0.00004**	-0.00008	0.00006	-0.00004**	0.00004**
	(0.00014)	(0.00007)	(0.00002)	(0.00002)	(0.00014)	(0.00007)	(0.00002)	(0.00002)	(0.00014)	(0.00007)	(0.00002)	(0.00002)
D(gender_head)	-0.02795	0.13669**	-0.02471*	0.00393	-0.03340	0.16042**	-0.02445*	0.00079	-0.03340	0.16042**	-0.02454*	0.00206
	(0.13793)	(0.06459)	(0.01313)	(0.01574)	(0.13928)	(0.06414)	(0.01314)	(0.01579)	(0.13928)	(0.06414)	(0.01312)	(0.01561)
year_educ_head	-0.01075	0.03173***	0.00032	0.00623***	-0.01033	0.02980***	0.00026	0.00642***	-0.01033	0.02980***	0.00025	0.00668***
	(0.01022)	(0.00446)	(0.00087)	(0.00116)	(0.01018)	(0.00443)	(0.00087)	(0.00114)	(0.01018)	(0.00443)	(0.00087)	(0.00113)
Hrs_ff	-0.00086	0.00219**	0.00024	0.00013	-0.00083	0.00205**	0.00024	0.00014	-0.00083	0.00205**	0.00024	0.00014
	(0.00258)	(0.00092)	(0.00017)	(0.00019)	(0.00257)	(0.00091)	(0.00017)	(0.00019)	(0.00257)	(0.00091)	(0.00017)	(0.00019)
Hrs_if	-0.00240	-0.00172	0.00040	0.00047*	-0.00223	-0.00170	0.00040	0.00049*	-0.00223	-0.00170	0.00040	0.00049*
	(0.00322)	(0.00164)	(0.00032)	(0.00028)	(0.00334)	(0.00163)	(0.00032)	(0.00028)	(0.00334)	(0.00163)	(0.00032)	(0.00028)
Hrs_ii	-0.00122	-0.00126*	-0.00007	0.00005	-0.00100	-0.00134*	-0.00007	0.00004	-0.00100	-0.00134*	-0.00008	0.00004
	(0.00166)	(0.00072)	(0.00015)	(0.00014)	(0.00166)	(0.00072)	(0.00015)	(0.00014)	(0.00166)	(0.00072)	(0.00015)	(0.00014)
Hrs_sf	0.00638	0.00160	0.00159***	-0.00004	0.00631	0.00138	0.00157***	-0.00000	0.00631	0.00138	0.00157***	-0.00000
	(0.00507)	(0.00157)	(0.00031)	(0.00030)	(0.00491)	(0.00156)	(0.00031)	(0.00031)	(0.00491)	(0.00156)	(0.00031)	(0.00031)
Hrs_si	0.00206	0.00135**	0.00036***	-0.00004	0.00214	0.00121**	0.00036***	-0.00003	0.00214	0.00121**	0.00036***	-0.00003
	(0.00160)	(0.00060)	(0.00011)	(0.00011)	(0.00161)	(0.00060)	(0.00011)	(0.00011)	(0.00161)	(0.00060)	(0.00011)	(0.00011)
Hrs_fam	-0.00504**	-0.00072	0.00028	0.00058**	-0.00511**	-0.00086	0.00028	0.00061**	-0.00511**	-0.00086	0.00028	0.00061**
	(0.00225)	(0.00111)	(0.00023)	(0.00029)	(0.00224)	(0.00111)	(0.00023)	(0.00028)	(0.00224)	(0.00111)	(0.00023)	(0.00028)
Hrs_second	0.00699	0.00004	0.00004	-0.00025	0.00697	-0.00022	0.00006	-0.00024	0.00697	-0.00022	0.00007	-0.00028
	(0.00508)	(0.00161)	(0.00036)	(0.00031)	(0.00497)	(0.00160)	(0.00036)	(0.00031)	(0.00497)	(0.00160)	(0.00036)	(0.00031)
D(Hrs_ff)	0.14725	-0.03745	0.00838	0.00904	0.15051	-0.03682	0.00837	0.00807	0.15051	-0.03682	0.00801	0.00947
	(0.18032)	(0.06690)	(0.01269)	(0.01435)	(0.17907)	(0.06655)	(0.01269)	(0.01433)	(0.17907)	(0.06655)	(0.01269)	(0.01425)
D(Hrs_if)	0.10708	0.08221	-0.01819	-0.03366**	0.07325	0.07851	-0.01786	-0.03325*	0.07325	0.07851	-0.01749	-0.03407**

	(0.20669)	(0.10180)	(0.01933)	(0.01701)	(0.20906)	(0.10156)	(0.01936)	(0.01700)	(0.20906)	(0.10156)	(0.01934)	(0.01690)
D(Hrs_ii)	0.16415	0.04099	0.01103	-0.01159	0.15463	0.03881	0.01066	-0.01000	0.15463	0.03881	0.01058	-0.00925
	(0.12726)	(0.05594)	(0.01131)	(0.01164)	(0.12691)	(0.05590)	(0.01130)	(0.01166)	(0.12691)	(0.05590)	(0.01130)	(0.01161)
D(Hrs_sf)	-0.28429	0.08558	-0.06878***	-0.00413	-0.28542	0.09135	-0.06772***	-0.00563	-0.28542	0.09135	-0.06752***	-0.00507
	(0.27595)	(0.09939)	(0.02036)	(0.02043)	(0.26902)	(0.09857)	(0.02033)	(0.02063)	(0.26902)	(0.09857)	(0.02032)	(0.02059)
D(Hrs_si)	0.04311	-0.08095*	0.00635	-0.01719*	0.04137	-0.08505*	0.00618	-0.01594	0.04137	-0.08505*	0.00599	-0.01611
	(0.11648)	(0.04797)	(0.00938)	(0.01010)	(0.11643)	(0.04779)	(0.00938)	(0.01016)	(0.11643)	(0.04779)	(0.00938)	(0.01014)
D(Hrs_fam)	0.07891	-0.00435	0.00079	0.00205	0.07557	0.00867	0.00102	0.00167	0.07557	0.00867	0.00097	0.00181
	(0.15719)	(0.06107)	(0.01225)	(0.01292)	(0.15672)	(0.06119)	(0.01221)	(0.01277)	(0.15672)	(0.06119)	(0.01222)	(0.01275)
D(Hrs_second)	0.08277	0.06613	-0.00474	0.01352	0.08972	0.07422	-0.00477	0.01269	0.08972	0.07422	-0.00467	0.01246
	(0.13599)	(0.05168)	(0.01094)	(0.01132)	(0.13404)	(0.05146)	(0.01092)	(0.01131)	(0.13404)	(0.05146)	(0.01092)	(0.01126)
D(Lima)	-0.51712***	0.56033***	0.10834***	0.07737***	-0.56207***	0.66815***	0.10942***	0.06584***	-0.56207***	0.66815***	0.11023***	0.06635***
	(0.14888)	(0.07042)	(0.00936)	(0.01209)	(0.13833)	(0.06646)	(0.00936)	(0.01239)	(0.13833)	(0.06646)	(0.00925)	(0.01227)
D(tongue)	0.13061	-0.21094***	0.02217***	-0.02021**	0.15001	-0.22238***	0.02033**	-0.01133	0.15001	-0.22238***	0.01901**	-0.00900
	(0.10461)	(0.04836)	(0.00827)	(0.00927)	(0.10313)	(0.04789)	(0.00826)	(0.00942)	(0.10313)	(0.04789)	(0.00826)	(0.00941)
D(migrant)	-0.06584	-0.04891	0.03740***	-0.02026**	-0.06929	-0.04594	0.03744***	-0.01980**	-0.06929	-0.04594	0.03763***	-0.02023**
	(0.08020)	(0.03597)	(0.00717)	(0.00837)	(0.08024)	(0.03579)	(0.00717)	(0.00836)	(0.08024)	(0.03579)	(0.00717)	(0.00832)
D(social)	0.08367	-0.01509	-0.01629**	0.00215	0.08980	-0.01657	-0.01607**	0.00273	0.08980	-0.01657	-0.01621**	0.00299
	(0.08014)	(0.03402)	(0.00682)	(0.00770)	(0.07997)	(0.03384)	(0.00682)	(0.00771)	(0.07997)	(0.03384)	(0.00682)	(0.00766)
P(pop_slums_dist)	0.02661	-0.01002			-0.02090	0.06406			-0.02090	0.06406		
	(0.13043)	(0.05910)			(0.12961)	(0.05753)			(0.12961)	(0.05753)		
P(migrant_dist)	0.25249	-0.14649			0.29641	-0.13813			0.29641	-0.13813		
	(0.53836)	(0.25655)			(0.52501)	(0.25528)			(0.52501)	(0.25528)		
P(race_dist)	-0.00593	0.92684***			0.13590	0.15126			0.13590	0.15126		
	(0.47995)	(0.28750)			(0.45315)	(0.26625)			(0.45315)	(0.26625)		
informal_index	0.00619	-0.15230**			-0.05669	-0.12819*			-0.05669	-0.12819*		
	(0.14919)	(0.07197)			(0.14603)	(0.07030)			(0.14603)	(0.07030)		
D(conf_gov_1)	0.02614	-0.04743			0.01783	-0.05181			0.01783	-0.05181		
	(0.09280)	(0.03943)			(0.09223)	(0.03935)			(0.09223)	(0.03935)		
D(conf_gov_2)	-0.10471	-0.12829**			-0.10280	-0.13527**			-0.10280	-0.13527**		

	(0.10939)	(0.05522)		(0.10882)	(0.05473)	(0.10882)	(0.05473)
D(conf_gov_3)	-0.00884	0.12469		0.00832	0.12268	0.00832	0.12268
	(0.16531)	(0.08644)		(0.16633)	(0.08532)	(0.16633)	(0.08532)
D(overall_satis)	0.08290	-0.09559***		0.08606	-0.09930***	0.08606	-0.09930***
	(0.08121)	(0.03672)		(0.08040)	(0.03655)	(0.08040)	(0.03655)
D(income_satis)	0.10862	-0.08103**		0.10518	-0.07951**	0.10518	-0.07951**
	(0.08450)	(0.03726)		(0.08398)	(0.03712)	(0.08398)	(0.03712)
D(income_stab)	0.03417	-0.13639***		0.02519	-0.13448***	0.02519	-0.13448***
	(0.09213)	(0.04178)		(0.09169)	(0.04151)	(0.09169)	(0.04151)
D(sub_pov)	-0.00222	-0.08188**		-0.00245	-0.07753**	-0.00245	-0.07753**
	(0.07623)	(0.03599)		(0.07590)	(0.03593)	(0.07590)	(0.03593)
D(road_1)	-0.20805	-0.01283		-0.21174*	-0.04253	-0.21174*	-0.04253
	(0.13045)	(0.06572)		(0.12838)	(0.06440)	(0.12838)	(0.06440)
D(road_2)	-0.23729*	0.10538*		-0.21288*	0.11527**	-0.21288*	0.11527**
	(0.12793)	(0.05849)		(0.12664)	(0.05806)	(0.12664)	(0.05806)
D(road_3)	0.14849	0.20617		0.17162	0.14109	0.17162	0.14109
	(0.28445)	(0.12905)		(0.29018)	(0.12928)	(0.29018)	(0.12928)
distance	-0.00035	0.00010		-0.00058**	-0.00002	-0.00058**	-0.00002
	(0.00030)	(0.00019)		(0.00029)	(0.00018)	(0.00029)	(0.00018)
P(nowater_dist)	0.45512*	-0.14798		0.51363**	-0.12200	0.51363**	-0.12200
	(0.25110)	(0.14801)		(0.24766)	(0.14248)	(0.24766)	(0.14248)
P(nodrain_dist)	-0.83622**	-0.27861		-0.93488***	-0.37271*	-0.93488***	-0.37271*
	(0.36214)	(0.20713)		(0.35071)	(0.20044)	(0.35071)	(0.20044)
P(noelec_dist)	1.96053**	-1.53848***		2.11163***	-1.62208***	2.11163***	-1.62208***
	(0.79237)	(0.43248)		(0.75374)	(0.43280)	(0.75374)	(0.43280)
P(nofloor_dist)	-0.02236	0.43434*		-0.11451	0.25801	-0.11451	0.25801
	(0.52908)	(0.24889)		(0.51810)	(0.24468)	(0.51810)	(0.24468)
P(nowall_dist)	0.31195	0.51127***		0.32975	0.57217***	0.32975	0.57217***
	(0.34755)	(0.14313)		(0.34395)	(0.14267)	(0.34395)	(0.14267)
P(noinsu_dist)	1.21636**	-0.07307		1.30655**	-0.28715	1.30655**	-0.28715

	(0.53959)	(0.27434)		(0.52468)	(0.26403)		(0.52468)	(0.26403)		
P(illiterate_dist)	-3.31319	-1.26675		-3.49123	1.23989		-3.49123	1.23989		
	(2.50925)	(1.46526)		(2.50453)	(1.40348)		(2.50453)	(1.40348)		
P(selfwork_dist)	0.66417	0.36068		0.82326	0.30423		0.82326	0.30423		
	(0.85468)	(0.45652)		(0.85083)	(0.44627)		(0.85083)	(0.44627)		
P(unreg_dist)	8.56601**	-4.01111*		9.24571**	-4.59968**		9.24571**	-4.59968**		
	(3.84533)	(2.14641)		(3.80244)	(2.05922)		(3.80244)	(2.05922)		
icf_index	-0.84992*	0.05859		-1.15167**	0.36345		-1.15167**	0.36345		
	(0.47072)	(0.27628)		(0.46509)	(0.26024)		(0.46509)	(0.26024)		
P(undernut_dist)	-0.25605	0.97962***		0.09325	0.71826**		0.09325	0.71826**		
	(0.63877)	(0.30207)		(0.61913)	(0.29343)		(0.61913)	(0.29343)		
Mort_distx1000	-0.00431	0.00848		-0.01072	0.01870		-0.01072	0.01870		
	(0.01899)	(0.01871)		(0.01933)	(0.01640)		(0.01933)	(0.01640)		
P(poor_dist)	2.35220***	-1.11691***		2.60998***	-1.08303***		2.60998***	-1.08303***		
	(0.66837)	(0.29621)		(0.62221)	(0.27948)		(0.62221)	(0.27948)		
lifeexp_dist	0.08512	0.01946		0.06364	0.05576		0.06364	0.05576		
	(0.07069)	(0.05961)		(0.07242)	(0.05277)		(0.07242)	(0.05277)		
D(ca_food)_1	0.07990	-0.00183		0.08117	0.00597		0.08117	0.00597		
	(0.09939)	(0.05074)		(0.09898)	(0.05057)		(0.09898)	(0.05057)		
D(ca_food)_2	0.00720	0.08356		0.02463	0.07695		0.02463	0.07695		
	(0.10891)	(0.05206)		(0.10728)	(0.05190)		(0.10728)	(0.05190)		
D(ca_food)_3	0.19555*	0.00916		0.19231*	0.00409		0.19231*	0.00409		
	(0.11272)	(0.04963)		(0.11246)	(0.04975)		(0.11246)	(0.04975)		
D(ca_food)_4	-0.07864	-0.02783		-0.09884	-0.03077		-0.09884	-0.03077		
	(0.10519)	(0.04814)		(0.10428)	(0.04813)		(0.10428)	(0.04813)		
Agropec_firmPC	-0.04288	-0.02375		-0.05686	-0.03056		-0.05686	-0.03056		
	(0.04346)	(0.01894)		(0.04265)	(0.01898)		(0.04265)	(0.01898)		
D(agro_dist)	0.12606	0.04278		0.13601	0.03129		0.13601	0.03129		
	(0.10257)	(0.04395)		(0.10205)	(0.04388)		(0.10205)	(0.04388)		
D(pec_dist)	-0.07091	-0.07543		-0.07636	-0.05673		-0.07636	-0.05673		

	(0.11158)	(0.05156)			(0.11462)	(0.05068)			(0.11462)	(0.05068)		
D(fish_dist)	0.34831*	0.02511			0.43121**	0.00766			0.43121**	0.00766		
	(0.17996)	(0.06258)			(0.18217)	(0.06179)			(0.18217)	(0.06179)		
Prop(workagro_dist)	-1.13021**	-0.94245***			-1.17074**	-0.81337***			-1.17074**	-0.81337***		
	(0.52980)	(0.25840)			(0.50151)	(0.24859)			(0.50151)	(0.24859)		
Prop(workfish_dist)	1.18677	1.58013*			1.42053	2.20632**			1.42053	2.20632**		
	(1.77733)	(0.90802)			(1.81788)	(0.91131)			(1.81788)	(0.91131)		
PHI			-0.14306**	-0.04731**			-0.15328**	-0.05418**			-0.15539**	-0.04326*
			(0.06704)	(0.02344)			(0.06612)	(0.02330)			(0.06608)	(0.02335)
Constant	-19.10400***	-1.46978	0.34381***	0.29445***	-18.24321***	-4.16662	0.34279***	0.33826***	-18.24321***	-4.16662	0.33923***	0.32533***
	(7.22901)	(5.31686)	(0.06373)	(0.10843)	(7.05045)	(4.76887)	(0.06331)	(0.10981)	(7.05045)	(4.76887)	(0.06318)	(0.10919)
Observations	7,513	7,513	7,513		7,513	7,513	7,513		7,513	7,513	7,513	
LogL	-691.436	-4216.255	4157.312		-695.351	-4252.393	4165.933		-695.351	-4252.393	4184.239	
Wald (ident. vars.)	99.46***	228.57***			165.70***	231.97***			165.70***	231.97***		

f) Results for models under specification 4

	Restriction 1				Restriction 2				Restriction 3			
	PROBIT		SUR		PROBIT		SUR		PROBIT		SUR	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
L(expend_Fon)					0.56833***	0.55535***			0.56833***	0.55535***		
					(0.03870)	(0.03073)			(0.03870)	(0.03073)		
O1L(expend_Fon)	0.57167***	0.56737***	-0.00588	0.00717			-0.00587	0.00940			-0.00634	0.01224
	(0.03885)	(0.03108)	(0.00719)	(0.01359)			(0.00717)	(0.01330)			(0.00716)	(0.01331)
L(O1_uv_inf)	-0.28682	0.26844	-0.15092***	0.06607***			-0.17553***	0.06817***			-0.19127***	0.13079***
	(0.27504)	(0.17472)	(0.02501)	(0.00899)			(0.02419)	(0.00901)			(0.02464)	(0.01815)
L(O1_uv_semi)	0.60173	0.42150**	0.08485***	-0.05175***			0.10736***	-0.01738			0.12745***	-0.07644***
	(0.43955)	(0.19429)	(0.02525)	(0.01282)			(0.02459)	(0.01179)			(0.02410)	(0.01885)
L(O1_uv_for)	0.06141	-0.46611***	0.06607***	-0.01432			0.06817***	-0.05079***			0.06383***	-0.05435***
	(0.11015)	(0.06177)	(0.00899)	(0.01079)			(0.00901)	(0.00925)			(0.00886)	(0.00899)
L(fam_size)	0.49108***	0.15798***	-0.00495	-0.04857***	0.49147***	0.15480***	-0.00534	-0.04891***	0.49147***	0.15480***	-0.00578	-0.04769***

	(0.13674)	(0.05914)	(0.01238)	(0.01330)	(0.13482)	(0.05891)	(0.01238)	(0.01333)	(0.13482)	(0.05891)	(0.01238)	(0.01332)
D(mem_spo)	0.09145	-0.09220	0.03573***	-0.01806	0.07854	-0.10476*	0.03659***	-0.01803	0.07854	-0.10476*	0.03693***	-0.01932
	(0.13197)	(0.06076)	(0.01250)	(0.01428)	(0.13327)	(0.06042)	(0.01251)	(0.01432)	(0.13327)	(0.06042)	(0.01250)	(0.01421)
D(mem_son)	0.21459	0.03343	0.02275*	-0.02859*	0.21121	0.03949	0.02317*	-0.03029**	0.21121	0.03949	0.02308*	-0.03075**
	(0.13471)	(0.06382)	(0.01365)	(0.01532)	(0.13380)	(0.06356)	(0.01366)	(0.01536)	(0.13380)	(0.06356)	(0.01366)	(0.01531)
D(mem_oth)	-0.01425	0.02256	0.01444	0.00098	-0.02227	0.02568	0.01479	0.00132	-0.02227	0.02568	0.01503*	0.00068
	(0.11953)	(0.04551)	(0.00904)	(0.00903)	(0.11868)	(0.04535)	(0.00905)	(0.00906)	(0.11868)	(0.04535)	(0.00904)	(0.00900)
P(men)	-0.31754*	0.07355	-0.07209***	0.01756	-0.30377*	0.06287	-0.07174***	0.01836	-0.30377*	0.06287	-0.07142***	0.01814
	(0.17043)	(0.07944)	(0.01691)	(0.02025)	(0.17157)	(0.07895)	(0.01691)	(0.02041)	(0.17157)	(0.07895)	(0.01691)	(0.02033)
age_head	0.02063	-0.00687	0.00597***	-0.00386**	0.02174	-0.00559	0.00593***	-0.00402**	0.02174	-0.00559	0.00588***	-0.00394**
	(0.01375)	(0.00731)	(0.00156)	(0.00184)	(0.01374)	(0.00724)	(0.00157)	(0.00184)	(0.01374)	(0.00724)	(0.00157)	(0.00183)
age_head-sq	-0.00010	0.00009	-0.00004**	0.00004**	-0.00011	0.00008	-0.00004**	0.00004**	-0.00011	0.00008	-0.00004**	0.00004**
	(0.00014)	(0.00007)	(0.00001)	(0.00002)	(0.00014)	(0.00007)	(0.00002)	(0.00002)	(0.00014)	(0.00007)	(0.00002)	(0.00002)
D(gender_head)	-0.02231	0.11605*	-0.02369*	0.00448	-0.02626	0.14045**	-0.02397*	0.00329	-0.02626	0.14045**	-0.02432*	0.00429
	(0.14025)	(0.06438)	(0.01309)	(0.01571)	(0.14196)	(0.06393)	(0.01311)	(0.01579)	(0.14196)	(0.06393)	(0.01309)	(0.01566)
year_educ_head	-0.01231	0.03677***	-0.00038	0.00653***	-0.01281	0.03489***	-0.00041	0.00682***	-0.01281	0.03489***	-0.00038	0.00704***
	(0.01000)	(0.00435)	(0.00084)	(0.00114)	(0.00998)	(0.00432)	(0.00085)	(0.00111)	(0.00998)	(0.00432)	(0.00085)	(0.00111)
Hrs_ff	-0.00098	0.00219**	0.00026	0.00012	-0.00094	0.00206**	0.00025	0.00015	-0.00094	0.00206**	0.00026	0.00015
	(0.00257)	(0.00092)	(0.00017)	(0.00019)	(0.00257)	(0.00091)	(0.00017)	(0.00019)	(0.00257)	(0.00091)	(0.00017)	(0.00019)
Hrs_if	-0.00226	-0.00166	0.00041	0.00040	-0.00209	-0.00163	0.00040	0.00044	-0.00209	-0.00163	0.00040	0.00043
	(0.00331)	(0.00164)	(0.00032)	(0.00027)	(0.00343)	(0.00164)	(0.00032)	(0.00027)	(0.00343)	(0.00164)	(0.00032)	(0.00027)
Hrs_ii	-0.00126	-0.00135*	-0.00005	0.00000	-0.00108	-0.00142**	-0.00005	-0.00002	-0.00108	-0.00142**	-0.00005	-0.00002
	(0.00166)	(0.00072)	(0.00015)	(0.00014)	(0.00166)	(0.00072)	(0.00015)	(0.00014)	(0.00166)	(0.00072)	(0.00015)	(0.00014)
Hrs_sf	0.00617	0.00142	0.00174***	-0.00002	0.00605	0.00120	0.00172***	0.00004	0.00605	0.00120	0.00172***	0.00003
	(0.00525)	(0.00156)	(0.00031)	(0.00031)	(0.00512)	(0.00155)	(0.00031)	(0.00031)	(0.00512)	(0.00155)	(0.00031)	(0.00031)
Hrs_si	0.00239	0.00123**	0.00039***	-0.00007	0.00250	0.00108*	0.00039***	-0.00005	0.00250	0.00108*	0.00039***	-0.00005
	(0.00163)	(0.00060)	(0.00011)	(0.00011)	(0.00164)	(0.00060)	(0.00011)	(0.00011)	(0.00164)	(0.00060)	(0.00011)	(0.00011)
Hrs_fam	-0.00530**	-0.00072	0.00029	0.00059**	-0.00535**	-0.00088	0.00029	0.00061**	-0.00535**	-0.00088	0.00030	0.00060**
	(0.00228)	(0.00111)	(0.00023)	(0.00029)	(0.00228)	(0.00112)	(0.00023)	(0.00028)	(0.00228)	(0.00112)	(0.00023)	(0.00028)
Hrs_second	0.00646	0.00003	0.00009	-0.00029	0.00640	-0.00018	0.00010	-0.00028	0.00640	-0.00018	0.00011	-0.00031

	(0.00507)	(0.00160)	(0.00037)	(0.00030)	(0.00498)	(0.00160)	(0.00037)	(0.00030)	(0.00498)	(0.00160)	(0.00037)	(0.00030)
D(Hrs_ff)	0.17321	-0.05171	0.00783	0.00488	0.18042	-0.05240	0.00753	0.00469	0.18042	-0.05240	0.00716	0.00599
	(0.18063)	(0.06688)	(0.01270)	(0.01405)	(0.17979)	(0.06650)	(0.01272)	(0.01404)	(0.17979)	(0.06650)	(0.01272)	(0.01401)
D(Hrs_if)	0.12376	0.08193	-0.01951	-0.03161*	0.09795	0.07707	-0.01893	-0.03166*	0.09795	0.07707	-0.01857	-0.03239*
	(0.21090)	(0.10185)	(0.01930)	(0.01691)	(0.21341)	(0.10164)	(0.01932)	(0.01695)	(0.21341)	(0.10164)	(0.01930)	(0.01687)
D(Hrs_ii)	0.16577	0.04140	0.01221	-0.00937	0.15978	0.03914	0.01202	-0.00797	0.15978	0.03914	0.01198	-0.00745
	(0.12838)	(0.05596)	(0.01128)	(0.01135)	(0.12823)	(0.05595)	(0.01129)	(0.01138)	(0.12823)	(0.05595)	(0.01128)	(0.01136)
D(Hrs_sf)	-0.25425	0.09009	-0.07479***	-0.00518	-0.25157	0.09581	-0.07368***	-0.00636	-0.25157	0.09581	-0.07348***	-0.00580
	(0.28387)	(0.09930)	(0.02038)	(0.02024)	(0.27854)	(0.09852)	(0.02034)	(0.02040)	(0.27854)	(0.09852)	(0.02034)	(0.02039)
D(Hrs_si)	0.03236	-0.07859	0.00514	-0.01477	0.02979	-0.08181*	0.00515	-0.01448	0.02979	-0.08181*	0.00514	-0.01469
	(0.11735)	(0.04795)	(0.00938)	(0.00996)	(0.11749)	(0.04777)	(0.00938)	(0.01003)	(0.11749)	(0.04777)	(0.00938)	(0.01002)
D(Hrs_fam)	0.10273	-0.01044	0.00092	-0.00010	0.10290	0.00227	0.00111	-0.00006	0.10290	0.00227	0.00099	0.00021
	(0.15965)	(0.06128)	(0.01225)	(0.01278)	(0.15995)	(0.06143)	(0.01222)	(0.01257)	(0.15995)	(0.06143)	(0.01223)	(0.01257)
D(Hrs_second)	0.09704	0.06506	-0.00796	0.01575	0.10587	0.07283	-0.00788	0.01525	0.10587	0.07283	-0.00792	0.01519
	(0.13740)	(0.05164)	(0.01095)	(0.01114)	(0.13579)	(0.05143)	(0.01094)	(0.01112)	(0.13579)	(0.05143)	(0.01094)	(0.01110)
D(Lima)	-0.52473***	0.57420***	0.10925***	0.08696***	-0.56484***	0.67136***	0.11041***	0.07975***	-0.56484***	0.67136***	0.10999***	0.08032***
	(0.14993)	(0.07092)	(0.00951)	(0.01170)	(0.13994)	(0.06687)	(0.00951)	(0.01189)	(0.13994)	(0.06687)	(0.00943)	(0.01180)
Pn_comisa_dist	-0.06153	0.21181	-0.13223***	-0.01248	-0.02846	0.22108	-0.13080***	-0.01295	-0.02846	0.22108	-0.13071***	-0.01065
	(0.24035)	(0.16734)	(0.02910)	(0.03584)	(0.24188)	(0.15911)	(0.02885)	(0.03704)	(0.24188)	(0.15911)	(0.02871)	(0.03694)
Pn_persmuni_dist	-0.00587	0.01816***	-0.00070	0.00435***	-0.00828*	0.01715***	-0.00067	0.00456***	-0.00828*	0.01715***	-0.00060	0.00442***
	(0.00515)	(0.00393)	(0.00095)	(0.00109)	(0.00496)	(0.00371)	(0.00096)	(0.00113)	(0.00496)	(0.00371)	(0.00096)	(0.00113)
LPCppto_dist	-0.19860***	-0.05244	0.00380	0.03330***	-0.18971***	-0.02973	0.00471	0.03124***	-0.18971***	-0.02973	0.00501	0.02954***
	(0.07065)	(0.03455)	(0.00573)	(0.00697)	(0.06965)	(0.03431)	(0.00574)	(0.00704)	(0.06965)	(0.03431)	(0.00573)	(0.00700)
d_paahh_dist	-0.10462	-0.08933	-0.02870*	-0.00411	-0.11849	-0.03212	-0.02648	-0.00886	-0.11849	-0.03212	-0.02598	-0.01301
	(0.22005)	(0.09511)	(0.01634)	(0.01626)	(0.21878)	(0.09475)	(0.01634)	(0.01610)	(0.21878)	(0.09475)	(0.01636)	(0.01617)
P(pop_slums_dist)	0.04063	0.01054			-0.00350	0.07298			-0.00350	0.07298		
	(0.13303)	(0.06134)			(0.13139)	(0.05996)			(0.13139)	(0.05996)		
P(migrant_dist)	0.03613	-0.42878*			0.13069	-0.36905			0.13069	-0.36905		
	(0.54942)	(0.25640)			(0.53275)	(0.25383)			(0.53275)	(0.25383)		
P(race_dist)	0.44967	0.51111*			0.55618	-0.29977			0.55618	-0.29977		

	(0.47269)	(0.27591)		(0.44156)	(0.25186)		(0.44156)	(0.25186)	
informal_index	0.06820	-0.17895**		0.01588	-0.15734**		0.01588	-0.15734**	
	(0.15462)	(0.07278)		(0.15165)	(0.07142)		(0.15165)	(0.07142)	
D(conf_gov_1)	0.02586	-0.04768		0.01858	-0.05206		0.01858	-0.05206	
	(0.09342)	(0.03954)		(0.09323)	(0.03945)		(0.09323)	(0.03945)	
D(conf_gov_2)	-0.11323	-0.11555**		-0.11603	-0.12344**		-0.11603	-0.12344**	
	(0.10833)	(0.05536)		(0.10800)	(0.05489)		(0.10800)	(0.05489)	
D(conf_gov_3)	-0.00965	0.11747		0.00703	0.11568		0.00703	0.11568	
	(0.16555)	(0.08654)		(0.16681)	(0.08548)		(0.16681)	(0.08548)	
D(overall_satis)	0.07951	-0.09790***		0.08100	-0.10165***		0.08100	-0.10165***	
	(0.08252)	(0.03677)		(0.08191)	(0.03661)		(0.08191)	(0.03661)	
D(income_satis)	0.11776	-0.08089**		0.11413	-0.08023**		0.11413	-0.08023**	
	(0.08538)	(0.03727)		(0.08491)	(0.03714)		(0.08491)	(0.03714)	
D(income_stab)	0.01299	-0.12962***		0.00124	-0.12894***		0.00124	-0.12894***	
	(0.09606)	(0.04178)		(0.09577)	(0.04151)		(0.09577)	(0.04151)	
D(sub_pov)	-0.00052	-0.09251**		-0.00024	-0.08738**		-0.00024	-0.08738**	
	(0.07737)	(0.03598)		(0.07720)	(0.03593)		(0.07720)	(0.03593)	
D(road_1)	-0.27185**	-0.05315		-0.27415**	-0.07214		-0.27415**	-0.07214	
	(0.13421)	(0.06815)		(0.13389)	(0.06697)		(0.13389)	(0.06697)	
D(road_2)	-0.32280**	0.09716		-0.30358**	0.12583**		-0.30358**	0.12583**	
	(0.13101)	(0.06148)		(0.13075)	(0.06111)		(0.13075)	(0.06111)	
D(road_3)	0.12023	0.19779		0.13071	0.13687		0.13071	0.13687	
	(0.28793)	(0.13093)		(0.29362)	(0.13149)		(0.29362)	(0.13149)	
distance	-0.00033	0.00013		-0.00055*	-0.00002		-0.00055*	-0.00002	
	(0.00030)	(0.00019)		(0.00029)	(0.00018)		(0.00029)	(0.00018)	
P(nowater_dist)	0.62412**	-0.20246		0.68926***	-0.15928		0.68926***	-0.15928	
	(0.25902)	(0.14907)		(0.25547)	(0.14316)		(0.25547)	(0.14316)	
P(nodrain_dist)	-1.11949***	-0.18958		-1.23297***	-0.30218		-1.23297***	-0.30218	
	(0.36531)	(0.20997)		(0.35346)	(0.20268)		(0.35346)	(0.20268)	
P(noelec_dist)	1.62253**	-1.57376***		1.74773**	-1.61465***		1.74773**	-1.61465***	

	(0.81398)	(0.45804)		(0.77567)	(0.46365)		(0.77567)	(0.46365)	
P(nofloor_dist)	-0.09927	0.44952*		-0.17063	0.25318		-0.17063	0.25318	
	(0.51855)	(0.25142)		(0.50840)	(0.24689)		(0.50840)	(0.24689)	
P(nowall_dist)	0.36876	0.49557***		0.37761	0.54305***		0.37761	0.54305***	
	(0.35440)	(0.14394)		(0.34967)	(0.14372)		(0.34967)	(0.14372)	
P(noinsu_dist)	0.86656	-0.02698		0.87405	-0.18640		0.87405	-0.18640	
	(0.55000)	(0.27853)		(0.54150)	(0.26904)		(0.54150)	(0.26904)	
P(illiterate_dist)	-3.15256	-1.44704		-3.07462	1.40917		-3.07462	1.40917	
	(2.45357)	(1.51131)		(2.42321)	(1.44169)		(2.42321)	(1.44169)	
P(selfwork_dist)	0.42952	0.60805		0.55569	0.48967		0.55569	0.48967	
	(0.87474)	(0.46541)		(0.86945)	(0.45538)		(0.86945)	(0.45538)	
P(unreg_dist)	9.84563***	-4.25891**		10.39691***	-4.91951**		10.39691***	-4.91951**	
	(3.78492)	(2.16610)		(3.75052)	(2.08318)		(3.75052)	(2.08318)	
icf_index	-0.68200	0.21475		-0.98212**	0.44290		-0.98212**	0.44290	
	(0.49800)	(0.28787)		(0.48780)	(0.27181)		(0.48780)	(0.27181)	
P(undernut_dist)	-0.43098	0.97546***		-0.13400	0.74983**		-0.13400	0.74983**	
	(0.63180)	(0.30427)		(0.61012)	(0.29578)		(0.61012)	(0.29578)	
Mort_distx1000	-0.00678	0.01095		-0.01307	0.02152		-0.01307	0.02152	
	(0.01994)	(0.01970)		(0.02086)	(0.01723)		(0.02086)	(0.01723)	
P(poor_dist)	2.00302***	-1.26953***		2.27912***	-1.10684***		2.27912***	-1.10684***	
	(0.67561)	(0.30958)		(0.62602)	(0.29024)		(0.62602)	(0.29024)	
lifeexp_dist	0.05629	0.03066		0.03458	0.06972		0.03458	0.06972	
	(0.07541)	(0.06272)		(0.07894)	(0.05533)		(0.07894)	(0.05533)	
D(ca_food)_1	0.07748	0.00574		0.07910	0.01361		0.07910	0.01361	
	(0.09975)	(0.05077)		(0.09935)	(0.05060)		(0.09935)	(0.05060)	
D(ca_food)_2	0.02820	0.07479		0.04459	0.06797		0.04459	0.06797	
	(0.11003)	(0.05201)		(0.10857)	(0.05184)		(0.10857)	(0.05184)	
D(ca_food)_3	0.17802	0.01452		0.17197	0.01077		0.17197	0.01077	
	(0.11563)	(0.04958)		(0.11573)	(0.04967)		(0.11573)	(0.04967)	
D(ca_food)_4	-0.06289	-0.02259		-0.08173	-0.02719		-0.08173	-0.02719	

	(0.10764)	(0.04819)			(0.10708)	(0.04818)			(0.10708)	(0.04818)		
Agropec_firmPC	-0.01762	-0.02086			-0.03267	-0.03126			-0.03267	-0.03126		
	(0.04604)	(0.01934)			(0.04473)	(0.01935)			(0.04473)	(0.01935)		
D(agro_dist)	0.13239	0.02946			0.14000	0.02279			0.14000	0.02279		
	(0.10230)	(0.04442)			(0.10144)	(0.04442)			(0.10144)	(0.04442)		
D(pec_dist)	-0.10934	-0.03706			-0.11172	-0.02870			-0.11172	-0.02870		
	(0.12407)	(0.05592)			(0.12644)	(0.05506)			(0.12644)	(0.05506)		
D(fish_dist)	0.30187*	0.02980			0.37514**	0.01056			0.37514**	0.01056		
	(0.17906)	(0.06319)			(0.18058)	(0.06245)			(0.18058)	(0.06245)		
Prop(workagro_dist)	-0.81870	-1.05053***			-0.83160*	-0.91776***			-0.83160*	-0.91776***		
	(0.52985)	(0.26283)			(0.50275)	(0.25282)			(0.50275)	(0.25282)		
Prop(workfish_dist)	1.39896	1.27838			1.72177	2.03494**			1.72177	2.03494**		
	(1.85393)	(0.92575)			(1.88503)	(0.92593)			(1.88503)	(0.92593)		
PHI			-0.13726**	-0.02104			-0.14029**	-0.01544			-0.14469**	-0.00643
			(0.06673)	(0.02302)			(0.06600)	(0.02199)			(0.06594)	(0.02212)
Constant	-16.54432**	-1.82627	0.35539***	0.02443	-15.63869**	-4.96872	0.34394***	0.04136	-15.63869**	-4.96872	0.34521***	0.03840
	(7.45083)	(5.52327)	(0.06931)	(0.11861)	(7.38198)	(4.93610)	(0.06898)	(0.11858)	(7.38198)	(4.93610)	(0.06880)	(0.11785)
Observations	7,513	7,513	7,513		7,513	7,513	7,513		7,513	7,513	7,513	
LogL	-686.014	-4218.244	4313.524		-689.148	-4255.895	4320.728		-689.148	-4255.895	4328.787	
Wald (ident. vars.)	107.10***	227.24***			173.58***	247.39***			173.58***	247.39***		

Appendix 8: Definition of variables and basic tabulations (models in Chapter 8)

a) Models with aggregate markets (non-linear SUR–GLS) and IV models

Variable	Definition	Non-linear SUR–GLS		Linear IV (aggregate)		Linear IV (Fon_inf)	
		2,222 obs.		2,160 obs.		2,142 obs.	
		Mean	s.d.	Mean	s.d.	Mean	s.d.
Dependant variables							
Informal	Informal consumption as a share of total expenditure (purchased), three-market definition (see Chapter 3)	0.404	0.17961	0.403	0.17868		
Semi-Formal	Semi-formal consumption as a share of total expenditure (purchased), three-market definition (see Chapter 3)	0.335	0.18368	0.335	0.18245		
Formal	Formal consumption as a share of total expenditure (purchased), three-market definition (see Chapter 3)	0.261	0.15701	0.263	0.15663		
Market_1	Consumption done in market 1 as a share of total expenditure (purchased), five-market definition (see Chapter 3)	0.176	0.13073				
Market_2	Consumption done in Market 2 as a share of total expenditure (purchased), five-market definition (see Chapter 3)	0.228	0.17202				
Market_3	Consumption done in Market 3 as a share of total expenditure (purchased), five-market definition (see Chapter 3)	0.335	0.18368				
Market_4	Consumption done in Market 4 as a share of total expenditure (purchased), five-market definition (see Chapter 3)	0.082	0.08702				
Market_5	Consumption done in Market 5 as a share of total expenditure (purchased), five-market definition (see Chapter 3)	0.179	0.12874				
Fon_inf	Consumption of Fon done at informal markets as a share of total Fon expenditure					0.602	0.28459
Fon_semi	Consumption of Fon done at semi-formal markets as a share of total Fon expenditure					0.333	0.27943
Fon_for	Consumption of Fon done at formal markets as a share of total Fon expenditure					0.065	0.14230
Independant variables: main regressions							
M	Total household expenditure (purchased) in annual new *soles*	14991.810	11526.95000				
L(expenditure)	Natural logarithm of total household expenditure (purchased) in annual new *soles*			9.421	0.63892		
L(expenditure_Fon)	Natural logarithm of total household expenditure (purchased) in annual new *soles* - Fon					8.404	0.67194
θ	Income of the husband over the sum of the income of the husband and wife	0.761	0.27930	0.761	0.27817	0.761	0.27763
L(fam_size)	Natural logarithm of the number of family members	1.369	0.31920	1.370	0.31355	1.372	0.31141

D(mem_son)	Dummy variable: 1 = household has a son or daughter, 0 = otherwise	0.893	0.30875	0.900	0.30069	0.902	0.29744
D(mem_oth)	Dummy variable: 1 = household has other member, 0 = otherwise	0.059	0.23560	0.057	0.23179	0.057	0.23270
P(men)	Proportion of men in the household	0.507	0.16123	0.508	0.16191	0.508	0.16226
D(gender_head)	Dummy variable: 1 = household head is male, 0 = otherwise	0.967	0.17947	0.967	0.17834	0.967	0.17906
age_male	Age of the male parent of the household (in years)	41.012	9.48395				
age_female	Age of the female parent of the household (in years) squared	37.102	9.07936				
year_educ_male	Number of years of education of the male parent of the household	11.851	3.96380				
year_educ_female	Number of years of education of the female parent of the household	10.773	4.52228				
D(chronic_male)	Dummy variable: 1 =male parent has a chronic illness, 0 = otherwise	0.236	0.42461				
D(chronic_female)	Dummy variable: 1 =female parent has a chronic illness, 0 = otherwise	0.282	0.44994				
Hrs_male	Sum of the weekly hours of work of the male parent including all labour options	53.067	21.95015				
Hrs_female	Sum of the weekly hours of work of the female parent including all labour options	31.277	28.74865				
D(Hrs_male)	Dummy variable: 1 =if the male parent works, 0 = otherwise	0.947	0.22429				
D(Hrs_female)	Dummy variable: 1 =if the female parent works, 0 = otherwise	0.688	0.46337				
Hrs_both	Dummy variable: 1 =if either of the parents works, 0 = otherwise			84.561	37.92940	84.521	37.83645
D(Hrs_both)	Dummy variable: 1 = household lives in Lima, 0 = otherwise			0.983	0.12805	0.983	0.12858
D(Lima)	Dummy variable: 1 = household lives in Lima, 0 = otherwise	0.165	0.37101	0.165	0.37110	0.165	0.37151
Pcoh_1	Proportion of household members under 3 years old	0.075	0.12077	0.076	0.12183	0.077	0.12195
Pcoh_2	Proportion of household members between 4 and 9 years old	0.163	0.15938	0.165	0.15912	0.165	0.15912
Pcoh_3	Proportion of household members between 10 and 13 years old	0.110	0.14025	0.110	0.13977	0.110	0.13990
Pcoh_4	Proportion of household members between 14 and 18 years old	0.076	0.12976	0.076	0.12971	0.076	0.13001
Pcoh_5	Proportion of household members between 19 and 24 years old	0.050	0.13155	0.050	0.13179	0.050	0.13189
Pcoh_6	Proportion of household members between 25 and 34 years old	0.174	0.22488	0.177	0.22548	0.176	0.22381

Pcoh_7	Proportion of household members between 35 and 40 years old	0.130	0.17724	0.132	0.17844	0.132	0.17766
Pcoh_8	Proportion of household members between 41 and 64 years old	0.218	0.28531	0.211	0.27906	0.211	0.27918
Pcoh_9	Proportion of household members over 64 years old	0.003	0.02835	0.004	0.02853	0.004	0.02865
L(pop_urb_dist)	Natural logarithm of the population size of the district where the household lives	10.955	1.32787	10.979	1.30586	10.985	1.30121
L(pop_urb_dist)-sq	Natural logarithm of the population size of the district where the household lives squared	121.771	28.38226	122.240	28.02395	122.370	27.96144
pop_den_distx1000	Population density: inhabitants per 1,000 km^2	2.611	4.98810	2.626	4.99287	2.638	5.00317
D(not_slum)	Dummy variable: 1 = household does nott live in a slum, 0 = otherwise	0.428	0.49483	0.431	0.49527	0.430	0.49519
D(mid_city)	Dummy variable: 1 = household lives in a departmental capital (except Lima), 0 = otherwise	0.482	0.49979	0.486	0.49991	0.486	0.49993
D(border)	Dummy variable: 1 = household lives in a border city (Tumbes, Tacna or Puno), 0 = otherwise	0.117	0.32204	0.118	0.32220	0.117	0.32171
D(tongue)	Dummy variable: 1 = either of the two parents of the household has indigenous mother tongue , 0 = otherwise	0.257	0.43681	0.245	0.43013	0.245	0.42997
D(migrant)	Dummy variable: 1 = either of the two parents of the household is a migrant, 0 = otherwise	0.691	0.46207	0.689	0.46287	0.690	0.46241
D(social)	Dummy variable: 1 = any of the household members pertain to social organizations (unions, social clubs, social programs, etc.) , 0 = otherwise	0.457	0.49824	0.456	0.49818	0.456	0.49815
Station_distx1000	Number of police stations in the district per 1,000 inhabitants	0.065	0.12995	0.065	0.13132	0.065	0.13162
Muni_persx1000	Number of workers in the local government of the district per 1,000 inhabitants	3.444	3.91409	3.455	3.95910	3.465	3.97148
L(budget_dist_pc)	Natural logarithm of the *per-capita* budget of the local government, in thousands of new *soles*	5.088	0.63421	5.089	0.63545	5.091	0.63525
D(sdp_dist)	Dummy variable: 1 = the district has a slum development plan, 0 = otherwise	0.038	0.19185	0.038	0.19115	0.038	0.19192
Instruments							
theta_educ_prima	Ratio of years of education of the male parent over years of education of the female parent			1.229	0.64824	1.231	0.65026
L(asset)	Natural logarithm of the value of assets owned by the household			7.487	1.46548	7.491	1.46633
year_educ_female	Number of years of education of the female parent of the household			11.071	4.21378	11.074	4.21487
year_educ_female-sq	Number of years of education of the female parent of the household (squared)			140.321	90.93251	140.395	90.96803
D(risk7)	Dummy variable 1 = if the family has experienced robbery or other crime against it in the last year, 0 = otherwise			0.036	0.18661	0.036	0.18736
D(house1)	Dummy variable 1 = if the family house is overcrowded (INEI, methodology), 0 = otherwise			0.091	0.28731	0.091	0.28773

Additional covariates: reduced form exercise					
M	Total household expenditure (purchased) in annual new *soles*		15191.830	11576.390	
θ_educ	Years of education of the male parent over the sum of years of education of both parents		0.526	0.09784	
asset	Value of assets in new *soles*		4446.628	8155.50100	
age_male	Age of the male parent of the household (in years)		40.752	9.34443	
age_female	Age of the female parent of the household (in years) squared		36.789	8.82229	
year_educ_male	Number of years of education of the male parent of the household		12.028	3.82081	
year_educ_female	Number of years of education of the female parent of the household		11.071	4.21378	
D(chronic_male)	Dummy variable: 1 =male parent has a chronic illness, 0 = otherwise		0.232	0.42247	
D(chronic_female)	Dummy variable: 1 =female parent has a chronic illness, 0 = otherwise		0.274	0.44615	
Hrs_male	Sum of the weekly hours of work of the male parent including all labour options		53.263	21.94676	
Hrs_female	Sum of the weekly hours of work of the female parent including all labour options		31.298	28.74015	
D(Hrs_male)	Dummy variable: 1 =if the male parent works, 0 = otherwise		0.948	0.22271	
D(Hrs_female)	Dummy variable: 1 =if the female parent works, 0 = otherwise		0.689	0.46287	

b) Models with disaggregated markets (consumption groups)

Variable	Definition	Groups	
		2,105 obs.	
		Mean	**s.d.**
Dependant variables			
Fon	Consumption done in food to be consumed within the household as a share of total expenditure	0.400	0.15278
Foff	Consumption done in food to be consumed outside the household as a share of total expenditure	0.090	0.09919
CC	Consumption done in clothing and personal care as a share of total expenditure	0.093	0.06157

HEA	Consumption done in health goods and services as a share of total expenditure	0.047	0.07111
TC	Consumption done in transportation and communication as a share of total expenditure	0.122	0.08609
ED	Consumption done in entertainment, education and cultural goods and services as a share of total expenditure	0.069	0.07473
OT	Consumption done in other or non-classified goods as a share of total expenditure	0.179	0.09575
D(Fon)	Dummy variable: 1 = household purchase Fon, 0 = otherwise	0.993	0.08413
D(Foff)	Dummy variable: 1 = household purchase Foff, 0 = otherwise	0.852	0.35493
D(CC)	Dummy variable: 1 = household purchase CC, 0 = otherwise	0.996	0.06155
D(HEA)	Dummy variable: 1 = household purchase HEA, 0 = otherwise	0.823	0.38193
D(TC)	Dummy variable: 1 = household purchase TC, 0 = otherwise	0.955	0.20660
D(ED)	Dummy variable: 1 = household purchase ED, 0 = otherwise	0.949	0.22067
D(OT)	Dummy variable: 1 = household purchase OT, 0 = otherwise	1.000	0.00000
Independant variables: SUR model			
M	Total household expenditure (purchased) in annual new *soles*	14812.390	10708.40000
θ	Income of the husband over the sum of the income of the husband and wife	0.762	0.27769
L(fam_size)	Natural logarithm of the number of family members	1.382	0.30825
D(mem_son)	Dummy variable: 1 = household has a son or daughter, 0 = otherwise	0.911	0.28457
D(mem_oth)	Dummy variable: 1 = household has other member, 0 = otherwise	0.058	0.23372
P(men)	Proportion of men in the household	0.507	0.16227
D(gender_head)	Dummy variable: 1 = household head is male, 0 = otherwise	0.971	0.16911
age_male	Age of the male parent of the household (in years)	40.842	9.39173
age_female	Age of the female parent of the household (in years) squared	36.904	8.91651
year_educ_male	Number of years of education of the male parent of the household	11.838	3.95257
year_educ_female	Number of years of education of the female parent of the household	10.759	4.51841

D(chronic_male)	Dummy variable: 1 =male parent has a chronic illness, 0 = otherwise	0.235	0.42420
D(chronic_female)	Dummy variable: 1 =female parent has a chronic illness, 0 = otherwise	0.279	0.44855
Hrs_male	Sum of the weekly hours of work of the male parent including all labour options	53.256	21.70760
Hrs_female	Sum of the weekly hours of work of the female parent including all labour options	31.222	28.71036
D(Hrs_male)	Dummy variable: 1 =if the male parent works, 0 = otherwise	0.950	0.21873
D(Hrs_female)	Dummy variable: 1 =if the female parent works, 0 = otherwise	0.687	0.46385
D(Lima)	Dummy variable: 1 = household lives in Lima, 0 = otherwise	0.158	0.36501
Pcoh_1	Proportion of household members under 3 years old	0.077	0.12183
Pcoh_2	Proportion of household members between 4 and 9 years old	0.168	0.15947
Pcoh_3	Proportion of household members between 10 and 13 years old	0.113	0.14139
Pcoh_4	Proportion of household members between 14 and 18 years old	0.077	0.12968
Pcoh_5	Proportion of household members between 19 and 24 years old	0.049	0.12812
Pcoh_6	Proportion of household members between 25 and 34 years old	0.175	0.22200
Pcoh_7	Proportion of household members between 35 and 40 years old	0.131	0.17549
Pcoh_8	Proportion of household members between 41 and 64 years old	0.209	0.27620
Pcoh_9	Proportion of household members over 64 years old	0.003	0.02674
Independant variables: PROBIT model (change of variables)			
L(expenditure)	Natural logarithm of total expenditure (purchased) in annual new *soles*	9.394	0.64954
age_average	Average age of the male and female parents (in years)	38.873	8.71466
year_educ_average	Average years of education of the male and female parents	11.298	3.85377
D(chronic_any)	Dummy variable: 1 = either of the parents has a chronic illness, 0 = otherwise	0.419	0.49343
Hrs_both	Sum of the weekly hours of work of the male and female parents including all labour options	84.477	37.96372
D(Hrs_both)	Dummy variable: 1 =if either of the parents works, 0 = otherwise	0.983	0.12968

Independant variables: PROBIT model (new variables)			
Hotel_firmPC	Number of tourism (hotel and restaurants) firms per 1,000 habitants (province level)	3.611	1.42221
D(rest_dist)	Dummy variable: 1 = restaurant activity is important in the district, 0 = otherwise (declared by the mayor)	0.747	0.43468
D(tour_dist)	Dummy variable: 1 = tourism activity is important in the district, 0 = otherwise (declared by the mayor)	0.209	0.40637
Prop(workrest_dist)	Proportion of urban workers in restaurant and hotel sector in the district	0.058	0.01500
D(ca_health)_1	Dummy variable: 1 = the household considers it has an adequate consumption of medicines, 0 = otherwise	0.437	0.49608
D(ca_health)_2	Dummy variable: 1 = the household considers it has an adequate consumption of soaps, 0 = otherwise	0.523	0.49961
n_eventual	Number of family members present a health episode	2.231	1.44082
Health1_distx1000	Number of health centres per 1,000 habitants (district level)	0.746	0.77491
Health2_distx1000	Number of hospitals per 1,000 habitants (district level)	0.028	0.04693
Health3_distx1000	Number of health micro-centres per 1,000 habitants (district level)	0.113	0.28048
Health4_distx1000	Number of other health providers per 1,000 habitants (district level)	0.555	0.64424
Health_firmPC	Number of health firms per 1,000 habitants (province level)	0.609	0.36250
Prop(workhealth_dist)	Proportion of urban workers in health sector in the district	0.027	0.01182

c) Models with disaggregated markets (markets within groups) 1

Variable	Definition	Fon		Foff		CC	
		2,090 obs.		1,794 obs.		2,097 obs.	
		Mean	s.d.	Mean	s.d.	Mean	s.d.
Dependent variables							
Fon_inf	Consumption of Fon done at informal markets as a share of total Fon expenditure	0.600	0.28433				
Fon_semi	Consumption of Fon done at semi-formal markets as a share of total Fon expenditure	0.338	0.27956				
Fon_for	Consumption of Fon done at formal markets as a share of total Fon expenditure	0.062	0.13680				

D(Fon_inf)	Dummy variable: 1 = household purchases from Fon_inf, 0 = otherwise	0.983	0.12835				
D(Fon_semi)	Dummy variable: 1 = household purchases from Fon_semi, 0 = otherwise	0.936	0.24416				
D(Fon_for)	Dummy variable: 1 = household purchases from Fon_for, 0 = otherwise	0.487	0.49994				
Foff_inf	Consumption of Foff done at informal markets as a share of total Foff expenditure			0.289	0.38728		
Foff_semi	Consumption of Foff done at semi-formal markets as a share of total Foff expenditure			0.167	0.33237		
Foff_for	Consumption of Foff done at formal markets as a share of total Foff expenditure			0.544	0.42859		
D(Foff_inf)	Dummy variable: 1 = household purchases from Foff_inf, 0 = otherwise			0.550	0.49762		
D(Foff_semi)	Dummy variable: 1 = household purchases from Foff_semi, 0 = otherwise			0.231	0.42144		
D(Foff_for)	Dummy variable: 1 = household purchases from Foff_for, 0 = otherwise			0.735	0.44133		
CC_inf	Consumption of CC done at informal markets as a share of total CC expenditure					0.442	0.32654
CC_semi	Consumption of CC done at semi-formal markets as a share of total CC expenditure					0.495	0.31734
CC_for	Consumption of CC done at formal markets as a share of total CC expenditure					0.063	0.13750
D(CC_inf)	Dummy variable: 1 = household purchases from CC_inf, 0 = otherwise					0.855	0.35215
D(CC_semi)	Dummy variable: 1 = household purchases from CC_semi, 0 = otherwise					0.958	0.20164
D(CC_for)	Dummy variable: 1 = household purchases from CC_for, 0 = otherwise					0.323	0.46767
Independent variables: SUR model							
Expenditure_Fon	Total household expenditure (purchased) in annual new *soles* - Fon	5277.114	2878.78600				
Expenditure_Foff	Total household expenditure (purchased) in annual new *soles* - Foff			1540.758	1734.03700		
Expenditure_CC	Total household expenditure (purchased) in annual new *soles* - CC					1404.166	1436.82800
θ	Income of the husband over the sum of the income of the husband and wife	0.762	0.27752	0.759	0.27208	0.761	0.27793
L(fam_size)	Natural logarithm of the number of family members	1.384	0.30618	1.380	0.30447	1.384	0.30652
D(mem_son)	Dummy variable: 1 = household has a son or daughter, 0 = otherwise	0.913	0.28132	0.914	0.28103	0.914	0.28089
D(mem_oth)	Dummy variable: 1 = household has other member, 0 = otherwise	0.058	0.23450	0.054	0.22511	0.058	0.23414

P(men)	Proportion of men in the household	0.507	0.16256	0.508	0.16359	0.507	0.16242
D(gender_head)	Dummy variable: 1 = household head is male, 0 = otherwise	0.970	0.16970	0.971	0.16781	0.970	0.16943
age_male	Age of the male parent of the household (in years)	40.868	9.38621	40.624	9.15184	40.791	9.34065
age_female	Age of the female parent of the household (in years) squared	36.916	8.92542	36.747	8.70405	36.858	8.87404
year_educ_male	Number of years of education of the male parent of the household	11.838	3.93943	12.101	3.80796	11.865	3.93184
year_educ_female	Number of years of education of the female parent of the household	10.756	4.52233	11.022	4.42538	10.785	4.50286
D(chronic_male)	Dummy variable: 1 =male parent has a chronic illness, 0 = otherwise	0.234	0.42376	0.233	0.42286	0.234	0.42326
D(chronic_female)	Dummy variable: 1 =female parent has a chronic illness, 0 = otherwise	0.278	0.44788	0.282	0.45012	0.278	0.44789
Hrs_male	Sum of the weekly hours of work of the male parent including all labour options	53.214	21.71216	54.350	21.31183	53.271	21.72827
Hrs_female	Sum of the weekly hours of work of the female parent including all labour options	31.185	28.67835	32.196	28.61971	31.197	28.70000
D(Hrs_male)	Dummy variable: 1 =if the male parent works, 0 = otherwise	0.949	0.21947	0.957	0.20274	0.949	0.21913
D(Hrs_female)	Dummy variable: 1 =if the female parent works, 0 = otherwise	0.687	0.46399	0.702	0.45760	0.687	0.46376
D(Lima)	Dummy variable: 1 = household lives in Lima, 0 = otherwise	0.158	0.36518	0.163	0.36976	0.159	0.36558
Pcoh_1	Proportion of household members under 3 years old	0.077	0.12191	0.077	0.12253	0.077	0.12195
Pcoh_2	Proportion of household members between 4 and 9 years old	0.168	0.15956	0.166	0.15806	0.168	0.15944
Pcoh_3	Proportion of household members between 10 and 13 years old	0.113	0.14148	0.110	0.14088	0.113	0.14136
Pcoh_4	Proportion of household members between 14 and 18 years old	0.077	0.12993	0.081	0.13258	0.077	0.12984
Pcoh_5	Proportion of household members between 19 and 24 years old	0.049	0.12814	0.051	0.13191	0.048	0.12781
Pcoh_6	Proportion of household members between 25 and 34 years old	0.173	0.22023	0.174	0.22538	0.175	0.22208
Pcoh_7	Proportion of household members between 35 and 40 years old	0.131	0.17558	0.134	0.17728	0.131	0.17566
Pcoh_8	Proportion of household members between 41 and 64 years old	0.209	0.27618	0.205	0.26943	0.207	0.27378
Pcoh_9	Proportion of household members over 64 years old	0.003	0.02684	0.003	0.02551	0.003	0.02679
Independent variables: PROBIT model (change of variables)							

L(expenditure_Fon)	Natural logarithm of total expenditure (purchased) in annual new *soles* - Fon	8.394	0.67229				
L(expenditure_Foff)	Natural logarithm of total expenditure (purchased) in annual new *soles* - Foff			6.737	1.23559		
L(expenditure_CC)	Natural logarithm of total expenditure (purchased) in annual new *soles* - CC					6.806	1.00638
age_average	Average age of the male and female parents (in years)	38.892	8.71776	38.686	8.48887	38.825	8.66641
year_educ_average	Average years of education of the male and female parents	11.297	3.84775	11.562	3.73132	11.325	3.83344
D(chronic_any)	Dummy variable: 1 = either of the parents has a chronic illness, 0 = otherwise	0.417	0.49314	0.420	0.49365	0.417	0.49322
Hrs_both	Sum of the weekly hours of work of the male and female parent including all labour options	84.400	37.86125	86.546	37.40526	84.468	37.96785
D(Hrs_both)	Dummy variable: 1 =if either of the parents works, 0 = otherwise	0.983	0.13014	0.987	0.11492	0.983	0.12993
Independent variables: PROBIT model (new variables)							
D(ca_food)_1	Dummy variable: 1 = the household considers it has an adequate consumption of food, 0 = otherwise	0.766	0.42346				
D(ca_food)_2	Dummy variable: 1 = the household considers it has an adequate consumption of tubers and cereals, 0 = otherwise	0.648	0.47761				
D(ca_food)_3	Dummy variable: 1 = the household considers it has an adequate consumption of fruits and vegetables, 0 = otherwise	0.569	0.49528				
D(ca_food)_4	Dummy variable: 1 = the household considers it has an adequate consumption of meat, fish and chicken, 0 = otherwise	0.560	0.49653				
Agropec_firmPC	Number of agropecuarian firms per 1,000 habitants (province level)	0.519	1.38545				
D(agro_dist)	Dummy variable: 1 = agrarian activity is important in the district, 0 = otherwise (declared by the mayor)	0.607	0.48860				
D(pec_dist)	Dummy variable: 1 = pecuarian activity is important in the district, 0 = otherwise (declared by the mayor)	0.313	0.46399				
D(fish_dist)	Dummy variable: 1 = fishing activity is important in the district, 0 = otherwise (declared by the mayor)	0.110	0.31302				
Prop(workagro_dist)	Proportion of urban workers in agropecuarian sector in the district	0.103	0.14231				
Prop(workfish_dist)	Proportion of urban workers in fishing sector in the district	0.011	0.02947				
Hotel_firmPC	Number of tourism (hotel and restaurants) firms per 1,000 habitants (province level)			3.703	1.41693		
D(rest_dist)	Dummy variable: 1 = restaurant activity is important in the district, 0 = otherwise (declared by the mayor)			0.764	0.42461		
D(tour_dist)	Dummy variable: 1 = tourism activity is important in the district, 0 = otherwise (declared by the mayor)			0.212	0.40910		
Prop(workrest_dist)	Proportion of urban workers in restaurant and hotel sector in the district			0.059	0.01449		

Com_firmPC	Number of commercial firms per 1,000 habitants (province level)					23.029	7.00569
Prop(workcom_dist)	Proportion of urban workers in commercial sector in the district					0.225	0.05624
D(ca_cloth)_1	Dummy variable: 1 = the household considers it has an adequate consumption of clothes, 0 = otherwise					0.420	0.49362
D(ca_cloth)_2	Dummy variable: 1 = the household considers it has an adequate consumption of shoes, 0 = otherwise					0.403	0.49061
P(pop_slums_dist)	Proportion of people in the district living in a slum	0.334	0.34640	0.345	0.34797	0.335	0.34622
P(migrant_dist)	Proportion of migrants living in the district (urban places)	0.316	0.10812	0.321	0.10466	0.316	0.10813
P(race_dist)	Proportion of indigenous people (by declared race) living in the district (urban places)	0.097	0.12619	0.095	0.11599	0.096	0.12536
informal_index	Index of tax informality at the province level, varies from 1 (high informality) to 0 (no informality) Index = (Tax of the IVA - (IVA payments/Value Added))/Tax of the IVA	0.305	0.31147	0.296	0.31056	0.305	0.31125
D(conf_gov_1)	Dummy variable: 1 = the head of the household considers that the public administration performance is good or very good , 0 = otherwise	0.311	0.46302	0.303	0.45954	0.312	0.46356
D(conf_gov_2)	Dummy variable: 1 = the head of the household considers that the performance of the government is not an obstacle for development , 0 = otherwise	0.197	0.39793	0.186	0.38891	0.197	0.39815
D(conf_gov_3)	Dummy variable: 1 = if the head of the household has a positive opinion in all responses related to the government , 0 = otherwise	0.089	0.28549	0.083	0.27604	0.090	0.28575
D(overall_satis)	Dummy variable: 1 = the head of the household considers his/her family as non-poor, 0 = otherwise	0.589	0.49205	0.570	0.49518	0.588	0.49232
D(income_satis)	Dummy variable: 1 = the head of the household is satisfied with his/her income, 0 = otherwise	0.327	0.46915	0.309	0.46213	0.329	0.46998
D(income_stab)	Dummy variable: 1 = the head of the household considers that his/her income is stable, 0 = otherwise	0.793	0.40538	0.784	0.41143	0.793	0.40557
D(sub_pov)	Dummy variable: 1 = if income is below the minimum income question, 0 = otherwise	0.400	0.49002	0.386	0.48703	0.401	0.49013
D(road_1)	Dummy variable: 1 = main road of the district is a non-asphalted highway, 0 = otherwise	0.205	0.40399	0.196	0.39681	0.206	0.40488
D(road_2)	Dummy variable: 1 = main road of the district is an asphalted highway, 0 = otherwise	0.663	0.47291	0.674	0.46891	0.662	0.47301
D(road_3)	Dummy variable: 1 = main road of the district is any other type, 0 = otherwise	0.025	0.15580	0.019	0.13639	0.025	0.15554
distance	Distance from the district to the capital of the province in minutes	83.752	163.17620	81.945	164.12380	83.577	162.74860
P(nowater_dist)	Proportion of households without water service in the district (urban places)	0.225	0.21029	0.215	0.20242	0.224	0.20954
P(nodrain_dist)	Proportion of houses without drain service in the district (urban places)	0.302	0.22440	0.288	0.21614	0.303	0.22447
P(noelec_dist)	Proportion of houses without electricity in the district (urban places)	0.119	0.08698	0.113	0.08154	0.119	0.08753
P(nofloor_dist)	Proportion of houses with floor of inadequate materials in the district (urban places)	0.308	0.16765	0.298	0.16046	0.308	0.16750

P(nowall_dist)	Proportion of houses with walls of inadequate materials in the district (urban places)	0.432	0.23726	0.419	0.23323	0.433	0.23725
P(noinsu_dist)	Proportion of people without insurance in the district (urban places)	0.582	0.09933	0.582	0.09903	0.582	0.09932
P(illiterate_dist)	Proportion of illiterate people in the district (urban places)	0.040	0.03129	0.038	0.02752	0.040	0.03125
P(selfwork_dist)	Proportion of self-employers in the district (urban places)	0.385	0.07573	0.383	0.07368	0.385	0.07563
P(unreg_dist)	Proportion of unregistered people in the district (urban places)	0.973	0.01523	0.974	0.01477	0.973	0.01524
P(undernut_dist)	Proportion of under-nourished children below 5 years old in the district	0.187	0.12998	0.180	0.12383	0.188	0.13010
Mort_distx1000	Child mortality, number per 1,000 born alive	17.175	6.61441	16.946	6.41947	17.167	6.59815
P(poor_dist)	Proportion of poor people in the district	0.308	0.16866	0.296	0.15948	0.308	0.16879
icf_index	Targeting index (between 0 and 1) of the government for social programs (district level)	0.147	0.18272	0.134	0.16713	0.148	0.18319
lifeexp_dist	Average life expectancy of the district	73.679	2.11702	73.746	2.06219	73.681	2.11264

d) Models with disaggregated markets (markets within groups) 2

Variable	Definition	HEA		TC		ED	
		1,732 obs.		2,011 obs.		1,997 obs.	
		Mean	s.d.	Mean	s.d.	Mean	s.d.
Dependent variables							
HEA_inf	Consumption of HEA done at informal markets as a share of total HEA expenditure	0.075	0.21179				
HEA_semi	Consumption of HEA done at semi-formal markets as a share of total HEA expenditure	0.158	0.31198				
HEA_for	Consumption of HEA done at formal markets as a share of total HEA expenditure	0.767	0.35820				
D(HEA_inf)	Dummy variable: 1 = household purchases from HEA_inf, 0 = otherwise	0.193	0.39464				
D(HEA_semi)	Dummy variable: 1 = household purchases from HEA_semi, 0 = otherwise	0.374	0.48404				
D(HEA_for)	Dummy variable: 1 = household purchases from HEA_for, 0 = otherwise	0.887	0.31688				
TC_inf	Consumption of TC done at informal markets as a share of total TC expenditure			0.466	0.33220		
TC_semi	Consumption of TC done at semi-formal markets as a share of total TC expenditure			0.075	0.14524		
TC_for	Consumption of TC done at formal markets as a share of total TC expenditure			0.459	0.33387		
D(TC_inf)	Dummy variable: 1 = household purchases from TC_inf, 0 = otherwise			0.878	0.32717		
D(TC_semi)	Dummy variable: 1 = household purchases from TC_semi, 0 = otherwise			0.569	0.49536		
D(TC_for)	Dummy variable: 1 = household purchases from TC_for, 0 = otherwise			0.824	0.38094		

ED_inf	Consumption of ED done at informal markets as a share of total ED expenditure					0.264	0.29928
ED_semi	Consumption of ED done at semi-formal markets as a share of total ED expenditure					0.283	0.27569
ED_for	Consumption of ED done at formal markets as a share of total ED expenditure					0.453	0.29667
D(ED_inf)	Dummy variable: 1 = household purchases from ED_inf, 0 = otherwise					0.757	0.42922
D(ED_semi)	Dummy variable: 1 = household purchases from ED_semi, 0 = otherwise					0.763	0.42557
D(ED_for)	Dummy variable: 1 = household purchases from ED_for, 0 = otherwise					0.947	0.22325
Independent variables: SUR model							
Expenditure_HEA	Total household expenditure (purchased) in annual new *soles* - HEA	944.095	1886.59800				
Expenditure_TC	Total household expenditure (purchased) in annual new *soles* - TC			2177.255	2519.43400		
Expenditure_ED	Total household expenditure (purchased) in annual new *soles* - ED					1385.675	2584.36300
θ	Income of the husband over the sum of the income of the husband and wife	0.762	0.27641	0.759	0.27678	0.762	0.27476
L(fam_size)	Natural logarithm of the number of family members	1.392	0.30274	1.383	0.30465	1.407	0.28750
D(mem_son)	Dummy variable: 1 = household has a son or daughter, 0 = otherwise	0.919	0.27265	0.915	0.27900	0.940	0.23771
D(mem_oth)	Dummy variable: 1 = household has other member, 0 = otherwise	0.059	0.23549	0.058	0.23320	0.060	0.23678
P(men)	Proportion of men in the household	0.503	0.16339	0.507	0.16301	0.508	0.16444
D(gender_head)	Dummy variable: 1 = household head is male, 0 = otherwise	0.968	0.17693	0.970	0.17155	0.970	0.16937
age_male	Age of the male parent of the household (in years)	40.865	9.36338	40.904	9.32613	40.529	8.97328
age_female	Age of the female parent of the household (in years) squared	36.931	8.85788	36.934	8.83357	36.582	8.43719
year_educ_male	Number of years of education of the male parent of the household	11.975	3.88831	12.020	3.85723	11.992	3.87692
year_educ_female	Number of years of education of the female parent of the household	10.932	4.42402	10.938	4.43568	10.886	4.48596
D(chronic_male)	Dummy variable: 1 =male parent has a chronic illness, 0 = otherwise	0.244	0.42975	0.237	0.42516	0.232	0.42212
D(chronic_female)	Dummy variable: 1 =female parent has a chronic illness, 0 = otherwise	0.296	0.45645	0.278	0.44811	0.273	0.44557
Hrs_male	Sum of the weekly hours of work of the male parent including all labour options	53.734	21.87513	53.436	21.86137	53.472	21.55967
Hrs_female	Sum of the weekly hours of work of the female parent including all labour options	30.976	28.83244	31.676	28.83767	31.409	28.67560
D(Hrs_male)	Dummy variable: 1 =if the male parent works, 0 = otherwise	0.950	0.21848	0.949	0.22050	0.952	0.21291
D(Hrs_female)	Dummy variable: 1 =if the female parent works, 0 = otherwise	0.682	0.46566	0.691	0.46232	0.691	0.46239
D(Lima)	Dummy variable: 1 = household lives in Lima, 0 = otherwise	0.171	0.37704	0.164	0.37046	0.161	0.36785
Pcoh_1	Proportion of household members under 3 years old	0.079	0.12336	0.075	0.12165	0.076	0.11950
Pcoh_2	Proportion of household members between 4 and 9 years old	0.169	0.15961	0.167	0.15927	0.175	0.15904
Pcoh_3	Proportion of household members between 10 and 13 years old	0.115	0.14135	0.114	0.14200	0.119	0.14258
Pcoh_4	Proportion of household members between 14 and 18 years old	0.076	0.12929	0.078	0.13071	0.081	0.13179
Pcoh_5	Proportion of household members between 19 and 24 years old	0.048	0.12616	0.048	0.12612	0.047	0.12364

Pcoh_6	Proportion of household members between 25 and 34 years old	0.175	0.21941	0.174	0.22274	0.174	0.21860
Pcoh_7	Proportion of household members between 35 and 40 years old	0.126	0.17095	0.132	0.17534	0.134	0.17342
Pcoh_8	Proportion of household members between 41 and 64 years old	0.209	0.27287	0.208	0.27254	0.191	0.25092
Pcoh_9	Proportion of household members over 64 years old	0.003	0.02742	0.003	0.02654	0.003	0.02745
Independent variables: PROBIT model (change of variables)							
L(expenditure_HEA)	Natural logarithm of total expenditure (purchased) in annual new *soles* - HEA	5.864	1.51231				
L(expenditure_TC)	Natural logarithm of total expenditure (purchased) in annual new *soles* - TC			7.125	1.17207		
L(expenditure_ED)	Natural logarithm of total expenditure (purchased) in annual new *soles* - ED					6.339	1.33424
age_average	Average age of the male and female parents (in years)	38.898	8.67263	38.919	8.63106	38.556	8.24450
year_educ_average	Average years of education of the male and female parents	11.454	3.76709	11.479	3.76058	11.439	3.79584
D(chronic_any)	Dummy variable: 1 = either of the parents has a chronic illness, 0 = otherwise	0.442	0.49673	0.419	0.49355	0.412	0.49234
Hrs_both	Sum of the weekly hours of work of the male and female parents including all labour options	84.710	38.32369	85.111	38.14834	84.880	37.93108
D(Hrs_both)	Dummy variable: 1 =if either of the parents works, 0 = otherwise	0.981	0.13675	0.983	0.13080	0.985	0.12167
Independent variables: PROBIT model (new variables)							
D(ca_health)_1	Dummy variable: 1 = the household considers it has an adequate consumption of medicines, 0 = otherwise	0.442	0.49673				
D(ca_health)_2	Dummy variable: 1 = the household considers it has an adequate consumption of soaps, 0 = otherwise	0.538	0.49873				
n_eventual	Number family members with a chronic illness	2.436	1.38447				
Health1_distx1000	Number of health centres per 1,000 habitants (district level)	0.741	0.77835				
Health2_distx1000	Number of hospitals per 1,000 habitants (district level)	0.027	0.04466				
Health3_distx1000	Number of health micro-centres per 1,000 habitants (district level)	0.105	0.26380				
Health4_distx1000	Number of other health providers per 1,000 habitants (district level)	0.560	0.65469				
Health_firmPC	Number of health firms per 1,000 habs. (province level)	0.620	0.36229				
Prop(workhealth_dist)	Proportion of urban workers in health sector in the district	0.027	0.01181				
D(ca_transport)	Dummy variable: 1 = the household considers it has an adequate consumption of transport, 0 = otherwise			0.529	0.49928		
comm1_distx1000	Number of communal phones per 1,000 habitants (district level)			0.170	0.31255		
comm2_distx1000	Number of telephone centrals per 1,000 habitants (district level)			0.022	0.08685		
comm3_distx1000	Number of post offices per 1,000 habitants (district level)			0.023	0.05928		
comm4_distx1000	Number of satellite dishes per 1,000 habitants (district level)			0.073	0.21245		
comm5_distx1000	Number of internet cabins per 1,000 habitants (district level)			0.829	1.22915		
comm6_distx1000	Number of other communication infrastructure per 1,000 habitants (district level)			0.012	0.06788		
D(comm_dist)	Dummy variable: 1 = if the district has other transportation and communication infrastructure , 0 = otherwise (declared by the mayor)			0.672	0.46967		

D(transp_dist)	Dummy variable: 1 = transportation activity is important in the district, 0 = otherwise (declared by the mayor)			0.912	0.28339		
transp_distx1000	Number of transport lines per 1,000 habitants (district level)			0.309	0.37557		
D(good_roads_dist)	Dummy variable: 1 =the roads in the district are good, 0 = otherwise (declared by the mayor)			0.914	0.28047		
Transport_firmPC	Number of transport firms per 1,000 habitants (province level)			0.821	0.42971		
D(TC_dist)	Dummy variable: 1 = there are workers in transport and communication in the district, 0 = otherwise (declared by the mayor)			0.127	0.33339		
Prop(worktransp_dist)	Proportion of urban workers in transport and communication sector in the district			0.102	0.02732		
Commun_firmPC	Number of communication firms per 1,000 habitants (province level)			1.756	0.72512		
D(ca_educ)	Dummy variable: 1 = the household considers it has an adequate consumption of education, 0 = otherwise					0.567	0.49564
D(ca_entert)	Dummy variable: 1 = the household considers it has an adequate consumption of entertainment, 0 = otherwise					0.149	0.35640
school_distx1000	Number of school rooms per 1,000 habitants (district level)					4.726	1.79051
Educa_firmPC	Number of educational firms per 1,000 habitants (province level)					0.521	0.32609
Entert_firmPC	Number of entertainment firms per 1,000 habitants (province level)					0.247	0.09741
Prop(workeduca_dist)	Proportion of urban workers in educational sector in the district					0.080	0.03886
P(pop_slums_dist)	Proportion of people in the district living in a slum	0.337	0.34386	0.337	0.34521	0.336	0.34545
P(migrant_dist)	Proportion of migrants living in the district (urban places)	0.318	0.10702	0.318	0.10510	0.316	0.10682
P(race_dist)	Proportion of indigenous people (by declared race) living in the district (urban places)	0.094	0.12104	0.094	0.11667	0.096	0.12262
informal_index	Index of tax informality at the province level, varies from 1 (high informality) to 0 (no informality) Index = (Tax of the IVA - (IVA payments/Value Added))/Tax of the IVA	0.289	0.30684	0.300	0.30912	0.303	0.31057
D(conf_gov_1)	Dummy variable: 1 = the head of the household considers that the public administration performance is good or very good , 0 = otherwise	0.301	0.45899	0.311	0.46293	0.309	0.46218
D(conf_gov_2)	Dummy variable: 1 = the head of the household considers that the performance of the government is not an obstacle for development , 0 = otherwise	0.191	0.39283	0.188	0.39078	0.195	0.39614
D(conf_gov_3)	Dummy variable: 1 = if the head of the household has a positive opinion in all responses related to the government , 0 = otherwise	0.087	0.28218	0.088	0.28266	0.089	0.28501
D(overall_satis)	Dummy variable: 1 = the head of the household considers his/her family as non-poor, 0 = otherwise	0.582	0.49337	0.576	0.49434	0.581	0.49346
D(income_satis)	Dummy variable: 1 = the head of the household is satisfied with his/her income, 0 = otherwise	0.330	0.47044	0.321	0.46707	0.327	0.46942
D(income_stab)	Dummy variable: 1 = the head of the household considers that his/her income is stable, 0 = otherwise	0.792	0.40589	0.787	0.40976	0.788	0.40870
D(sub_pov)	Dummy variable: 1 = if income is below the minimum income question, 0 = otherwise	0.387	0.48717	0.395	0.48904	0.402	0.49045

D(road_1)	Dummy variable: 1 = main road of the district is a non-asphalted highway, 0 = otherwise	0.195	0.39643	0.197	0.39815	0.207	0.40512
D(road_2)	Dummy variable: 1 = main road of the district is an asphalted highway, 0 = otherwise	0.679	0.46700	0.669	0.47058	0.663	0.47263
D(road_3)	Dummy variable: 1 = main road of the district is any other type, 0 = otherwise	0.021	0.14271	0.024	0.15422	0.026	0.15779
distance	Distance from the district to the capital of the province in minutes	81.236	161.92620	81.303	156.38560	83.619	161.84560
P(nowater_dist)	Proportion of households without water service in the district (urban places)	0.220	0.20423	0.219	0.20504	0.221	0.20554
P(nodrain_dist)	Proportion of houses without drain service in the district (urban places)	0.294	0.21857	0.293	0.21725	0.298	0.22094
P(noelec_dist)	Proportion of houses without electricity in the district (urban places)	0.116	0.08514	0.115	0.08159	0.119	0.08703
P(nofloor_dist)	Proportion of houses with floor of inadequate materials in the district (urban places)	0.303	0.16540	0.302	0.16067	0.306	0.16554
P(nowall_dist)	Proportion of houses with walls of inadequate materials in the district (urban places)	0.425	0.23420	0.425	0.23399	0.429	0.23592
P(noinsu_dist)	Proportion of people without insurance in the district (urban places)	0.583	0.09908	0.582	0.09807	0.581	0.09872
P(illiterate_dist)	Proportion of illiterate people in the district (urban places)	0.039	0.02907	0.038	0.02747	0.039	0.03043
P(selfwork_dist)	Proportion of self-employers in the district (urban places)	0.384	0.07328	0.384	0.07466	0.384	0.07537
P(unreg_dist)	Proportion of unregistered people in the district (urban places)	0.974	0.01407	0.974	0.01501	0.974	0.01461
P(undernut_dist)	Proportion of under-nourished children below 5 years old in the district	0.183	0.12684	0.183	0.12673	0.186	0.12972
Mort_distx1000	Child mortality, number per 1,000 born alive	16.904	6.47019	17.014	6.51228	17.120	6.60302
P(poor_dist)	Proportion of poor people in the district	0.302	0.16474	0.301	0.16334	0.305	0.16703
icf_index	Targeting index (between 0 and 1) of the government for social programs (district level)	0.140	0.17602	0.139	0.17304	0.144	0.17901
lifeexp_dist	Average life expectancy of the district	73.766	2.07355	73.730	2.09164	73.706	2.11957

Appendix 9: Full regression results (models in Chapter 8)

a) Aggregate model (three-market definition)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
θ	-0.18973	-0.07715	-0.04667	-0.08721	-0.16462	-0.07486	-0.18738	-0.07391
	(0.18485)	(0.15310)	(0.17807)	(0.14998)	(0.18372)	(0.15417)	(0.18469)	(0.15166)
$\theta \ln(\theta M)$	-0.03248***	0.07278***	-0.04710***	0.07239***	-0.03290***	0.07235***	-0.03169***	0.07152***
	(0.00906)	(0.00791)	(0.00894)	(0.00777)	(0.00904)	(0.00796)	(0.00906)	(0.00789)
$(1-\theta)\ln[(1-\theta)M]$	-0.05555***	0.06595***	-0.05635***	0.06484***	-0.05353***	0.06577***	-0.05444***	0.06507***
	(0.01569)	(0.01313)	(0.01506)	(0.01299)	(0.01563)	(0.01320)	(0.01564)	(0.01299)
L(fam_size)	-0.03624	0.01902	-0.03251	0.01879	-0.03820	0.01981	-0.03671	0.02146
	(0.03166)	(0.02339)	(0.03052)	(0.02315)	(0.03180)	(0.02347)	(0.03171)	(0.02339)
D(mem_son)	-0.01181	0.00102	-0.02099	0.00441	-0.00966	0.00080	-0.01214	0.00149
	(0.01865)	(0.01522)	(0.01812)	(0.01504)	(0.01866)	(0.01529)	(0.01864)	(0.01519)
D(mem_oth)	0.01501	-0.01432	0.01860	-0.01682	0.01774	-0.01460	0.01458	-0.01412
	(0.02076)	(0.01312)	(0.02073)	(0.01341)	(0.02064)	(0.01311)	(0.02078)	(0.01316)
P(men)	-0.01304	-0.04202**	-0.00677	-0.04133**	-0.01081	-0.04188**	-0.01161	-0.04255**
	(0.02227)	(0.01656)	(0.02168)	(0.01645)	(0.02222)	(0.01659)	(0.02224)	(0.01652)
D(gender_head)	-0.02360	0.00828	-0.02431	0.00846	-0.02322	0.00842	-0.02406	0.00587
	(0.02315)	(0.01501)	(0.02215)	(0.01513)	(0.02353)	(0.01500)	(0.02309)	(0.01503)
age_male	-0.00039	0.00137**	-0.00018	0.00127*	-0.00054	0.00135**	-0.00040	0.00133**
	(0.00086)	(0.00066)	(0.00084)	(0.00066)	(0.00086)	(0.00066)	(0.00086)	(0.00066)
age_female	0.00110	0.00151*	0.00139	0.00130*	0.00113	0.00154**	0.00114	0.00130*
	(0.00097)	(0.00078)	(0.00094)	(0.00077)	(0.00097)	(0.00078)	(0.00097)	(0.00078)
year_educ_male	-0.00079	0.00607***	-0.00143	0.00589***	-0.00089	0.00610***	-0.00089	0.00617***
	(0.00136)	(0.00096)	(0.00133)	(0.00097)	(0.00137)	(0.00097)	(0.00136)	(0.00096)
year_educ_female	-0.00351***	0.00459***	-0.00317***	0.00401***	-0.00239**	0.00456***	-0.00345***	0.00449***
	(0.00116)	(0.00086)	(0.00114)	(0.00087)	(0.00117)	(0.00087)	(0.00116)	(0.00086)
D(chronic_male)	0.00205	0.00274	-0.00281	0.00296	0.00236	0.00286	0.00131	0.00337
	(0.00906)	(0.00684)	(0.00878)	(0.00673)	(0.00905)	(0.00688)	(0.00906)	(0.00683)
D(chronic_female)	-0.00056	0.01370**	-0.00519	0.01495**	-0.00034	0.01381**	-0.00108	0.01374**
	(0.00850)	(0.00665)	(0.00834)	(0.00666)	(0.00848)	(0.00668)	(0.00847)	(0.00663)
Hrs_male	0.00032	-0.00011	0.00013	-0.00006	0.00032	-0.00012	0.00029	-0.00010
	(0.00021)	(0.00017)	(0.00021)	(0.00017)	(0.00021)	(0.00017)	(0.00022)	(0.00017)
Hrs_female	0.00012	0.00069***	0.00013	0.00069***	0.00007	0.00069***	0.00013	0.00069***
	(0.00020)	(0.00016)	(0.00019)	(0.00016)	(0.00020)	(0.00016)	(0.00020)	(0.00016)
D(Hrs_male)	-0.06673***	0.02467	-0.05562***	0.02291	-0.06764***	0.02517	-0.06438***	0.02347
	(0.02134)	(0.01898)	(0.02080)	(0.01867)	(0.02132)	(0.01892)	(0.02131)	(0.01897)
D(Hrs_female)	-0.01444	0.00105	-0.01170	0.00019	-0.01142	0.00145	-0.01570	0.00067
	(0.01225)	(0.00898)	(0.01186)	(0.00892)	(0.01226)	(0.00905)	(0.01228)	(0.00899)
D(Lima)	0.10529***	0.01989**	0.11961***	-0.00163	0.10266***	0.01937**	0.10175***	0.01949**
	(0.01005)	(0.00803)	(0.01790)	(0.01339)	(0.01013)	(0.00812)	(0.01025)	(0.00817)
Pcoh_1	0.02581	-0.04956	-0.06917	-0.05340	0.03734	-0.05028	0.01947	-0.04776
	(0.15386)	(0.11620)	(0.15947)	(0.11680)	(0.15705)	(0.11656)	(0.15331)	(0.11640)
Pcoh_2	0.07402	-0.12471	-0.00358	-0.12913	0.08948	-0.12526	0.07116	-0.12507
	(0.15115)	(0.11323)	(0.15697)	(0.11360)	(0.15455)	(0.11362)	(0.15053)	(0.11334)
Pcoh_3	0.14447	-0.12622	0.05555	-0.12752	0.15201	-0.12659	0.14211	-0.12492
	(0.15269)	(0.11351)	(0.15856)	(0.11410)	(0.15588)	(0.11398)	(0.15212)	(0.11364)
Pcoh_4	0.09980	-0.06660	0.02809	-0.07306	0.10775	-0.06768	0.09579	-0.06375
	(0.15341)	(0.11473)	(0.15883)	(0.11532)	(0.15650)	(0.11507)	(0.15290)	(0.11478)
Pcoh_5	-0.01180	0.04152	-0.09293	0.03678	-0.00440	0.04139	-0.01457	0.04534
	(0.15034)	(0.11226)	(0.15645)	(0.11294)	(0.15322)	(0.11243)	(0.14964)	(0.11228)

Pcoh_6	-0.02421	0.02124	-0.11822	0.02137	-0.01580	0.02106	-0.03074	0.02688
	(0.15369)	(0.11415)	(0.15943)	(0.11507)	(0.15651)	(0.11431)	(0.15307)	(0.11411)
Pcoh_7	0.00550	0.00873	-0.09138	0.00802	0.01493	0.00826	0.00043	0.01672
	(0.15711)	(0.11710)	(0.16260)	(0.11818)	(0.15973)	(0.11723)	(0.15660)	(0.11711)
Pcoh_8	-0.02423	0.01622	-0.12072	0.01588	-0.01056	0.01550	-0.03061	0.02815
	(0.16064)	(0.11933)	(0.16671)	(0.12035)	(0.16331)	(0.11949)	(0.16009)	(0.11928)
L(pop_urb_dist)			0.19329***	-0.03656				
			(0.04264)	(0.03599)				
L(pop_urb_dist)-sq			-0.00797***	0.00163				
			(0.00202)	(0.00169)				
pop_den_distx1000			-0.00289***	0.00150*				
			(0.00099)	(0.00082)				
D(not_slum)			-0.00396	0.02725***				
			(0.00754)	(0.00596)				
D(mid_city)			-0.00490	-0.00902				
			(0.00845)	(0.00653)				
D(border)			0.05692***	-0.00202				
			(0.01085)	(0.00906)				
D(tongue)					0.03387***	-0.00235		
					(0.00887)	(0.00697)		
D(migrant)					0.01118	0.00263		
					(0.00806)	(0.00616)		
D(social)					-0.00923	-0.00179		
					(0.00762)	(0.00587)		
Station_distx1000							-0.06530***	0.01154
							(0.02245)	(0.01393)
Muni_persx1000							-0.00084	0.00232***
							(0.00115)	(0.00061)
L(budget_dist_pc)							0.00567	-0.00727
							(0.00657)	(0.00488)
D(sdp_dist)							-0.00959	-0.03224**
							(0.01830)	(0.01283)
Constant	1.00976***	-0.58901***	-0.06600	-0.36283	0.95982***	-0.58910***	0.98431***	-0.54795***
	(0.22202)	(0.17348)	(0.32285)	(0.24467)	(0.22381)	(0.17400)	(0.22181)	(0.17332)
Observations	2,222		2,222		2,222		2,222	
LogL	2348.952		2429.140		2360.635		2359.235	

b) Aggregate model (five-market definition)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.13513	-0.04445	0.14510	-0.04415	0.13882	-0.04584	0.13908	-0.04199
	(0.14335)	(0.12500)	(0.14196)	(0.12352)	(0.13834)	(0.12589)	(0.14265)	(0.12414)
$\theta\ln(\theta M)$	-0.02651***	0.05196***	-0.02738***	0.05013***	-0.02357***	0.05168***	-0.02836***	0.05069***
	(0.00663)	(0.00630)	(0.00656)	(0.00629)	(0.00643)	(0.00639)	(0.00659)	(0.00630)
$(1-\theta)\ln[(1-\theta)M]$	-0.01534	0.04750***	-0.01519	0.04591***	-0.01229	0.04706***	-0.01677	0.04652***
	(0.01265)	(0.01114)	(0.01236)	(0.01097)	(0.01219)	(0.01120)	(0.01263)	(0.01106)
L(fam_size)	-0.04575**	0.02862	-0.04122*	0.02925	-0.05390**	0.02922	-0.04307**	0.03093*
	(0.02187)	(0.01842)	(0.02132)	(0.01833)	(0.02207)	(0.01848)	(0.02115)	(0.01842)
D(mem_son)	-0.00871	0.00698	-0.01344	0.00905	-0.00397	0.00723	-0.00941	0.00739
	(0.01366)	(0.01142)	(0.01306)	(0.01133)	(0.01360)	(0.01142)	(0.01345)	(0.01137)
D(mem_oth)	0.01905	-0.00815	0.02604**	-0.00938	0.02415*	-0.00836	0.01856	-0.00806
	(0.01350)	(0.01015)	(0.01324)	(0.01054)	(0.01327)	(0.01015)	(0.01350)	(0.01016)
P(men)	-0.00721	-0.04638***	-0.00416	-0.04526***	-0.00632	-0.04652***	-0.00251	-0.04610***

	(0.01601)	(0.01381)	(0.01529)	(0.01369)	(0.01559)	(0.01382)	(0.01587)	(0.01381)
D(gender_head)	-0.03088*	0.01741	-0.03279**	0.01740	-0.03230**	0.01710	-0.02892*	0.01537
	(0.01584)	(0.01234)	(0.01549)	(0.01217)	(0.01591)	(0.01236)	(0.01523)	(0.01228)
age_male	-0.00081	0.00145***	-0.00058	0.00140***	-0.00082	0.00142***	-0.00084	0.00140***
	(0.00062)	(0.00052)	(0.00061)	(0.00052)	(0.00062)	(0.00052)	(0.00062)	(0.00052)
age_female	0.00090	0.00219***	0.00100	0.00199***	0.00066	0.00220***	0.00105	0.00203***
	(0.00071)	(0.00059)	(0.00067)	(0.00059)	(0.00070)	(0.00059)	(0.00070)	(0.00059)
year_educ_male	-0.00012	0.00454***	0.00002	0.00430***	-0.00055	0.00453***	-0.00014	0.00463***
	(0.00095)	(0.00079)	(0.00094)	(0.00080)	(0.00094)	(0.00080)	(0.00094)	(0.00079)
year_educ_female	-0.00118	0.00506***	-0.00049	0.00457***	0.00035	0.00510***	-0.00130	0.00496***
	(0.00082)	(0.00068)	(0.00080)	(0.00069)	(0.00082)	(0.00070)	(0.00081)	(0.00068)
D(chronic_male)	0.00696	-0.00120	0.00321	-0.00204	0.00592	-0.00138	0.00841	-0.00051
	(0.00668)	(0.00577)	(0.00647)	(0.00566)	(0.00666)	(0.00579)	(0.00662)	(0.00576)
D(chronic_female)	0.00734	0.00809	0.00368	0.00823	0.00635	0.00795	0.00592	0.00786
	(0.00630)	(0.00564)	(0.00626)	(0.00560)	(0.00622)	(0.00565)	(0.00620)	(0.00563)
Hrs_male	-0.00023	-0.00044***	-0.00033**	-0.00042***	-0.00017	-0.00044***	-0.00024	-0.00043***
	(0.00017)	(0.00014)	(0.00016)	(0.00014)	(0.00016)	(0.00014)	(0.00016)	(0.00014)
Hrs_female	0.00084***	0.00039***	0.00080***	0.00038***	0.00080***	0.00039***	0.00084***	0.00040***
	(0.00016)	(0.00013)	(0.00015)	(0.00013)	(0.00016)	(0.00013)	(0.00016)	(0.00013)
D(Hrs_male)	-0.00104	0.02002	0.00294	0.02002	-0.00608	0.02028	0.00007	0.01930
	(0.01596)	(0.01666)	(0.01526)	(0.01642)	(0.01574)	(0.01654)	(0.01570)	(0.01666)
D(Hrs_female)	-0.00535	0.01066	-0.00517	0.01014	-0.00566	0.01071	-0.00556	0.01008
	(0.00880)	(0.00742)	(0.00858)	(0.00731)	(0.00875)	(0.00750)	(0.00869)	(0.00744)
D(Lima)	-0.02356***	0.02184***	-0.00580	0.01195	-0.02236***	0.02143***	-0.01780***	0.02193***
	(0.00591)	(0.00711)	(0.01193)	(0.01185)	(0.00597)	(0.00717)	(0.00614)	(0.00726)
Pcoh_1	0.10168	-0.01658	0.09542	-0.02710	0.11350	-0.02070	0.07385	-0.01903
	(0.10132)	(0.10538)	(0.10027)	(0.10478)	(0.10019)	(0.10538)	(0.10232)	(0.10527)
Pcoh_2	0.12649	-0.04448	0.12884	-0.05326	0.14130	-0.04836	0.10390	-0.04797
	(0.09898)	(0.10305)	(0.09790)	(0.10223)	(0.09800)	(0.10306)	(0.09998)	(0.10283)
Pcoh_3	0.10385	-0.04341	0.10569	-0.04988	0.10832	-0.04744	0.08253	-0.04505
	(0.09961)	(0.10345)	(0.09825)	(0.10276)	(0.09849)	(0.10342)	(0.10066)	(0.10327)
Pcoh_4	0.15285	-0.00839	0.15504	-0.01966	0.16453	-0.01222	0.12196	-0.01050
	(0.10092)	(0.10463)	(0.09966)	(0.10387)	(0.10003)	(0.10448)	(0.10197)	(0.10442)
Pcoh_5	0.05645	0.06326	0.05991	0.05273	0.06281	0.06171	0.03618	0.06384
	(0.09922)	(0.10380)	(0.09769)	(0.10305)	(0.09779)	(0.10358)	(0.09998)	(0.10358)
Pcoh_6	0.04339	0.03571	0.04578	0.02939	0.05077	0.03384	0.01788	0.03669
	(0.10271)	(0.10590)	(0.10092)	(0.10519)	(0.10125)	(0.10563)	(0.10312)	(0.10561)
Pcoh_7	0.06968	0.02854	0.06399	0.02043	0.08116	0.02692	0.04079	0.03142
	(0.10558)	(0.10812)	(0.10356)	(0.10745)	(0.10407)	(0.10781)	(0.10575)	(0.10785)
Pcoh_8	0.04346	0.02256	0.04483	0.01598	0.06090	0.02045	0.01065	0.02829
	(0.10895)	(0.11007)	(0.10672)	(0.10937)	(0.10743)	(0.10974)	(0.10896)	(0.10978)
L(pop_urb_dist)			-0.01697	-0.01165				
			(0.03283)	(0.02896)				
L(pop_urb_dist)-sq			0.00098	0.00045				
			(0.00153)	(0.00137)				
pop_den_distx1000			-0.00226***	0.00115				
			(0.00063)	(0.00072)				
D(not_slum)			-0.01475***	0.02695***				
			(0.00572)	(0.00480)				
D(mid_city)			-0.01072	-0.00651				
			(0.00679)	(0.00523)				
D(border)			0.08883***	0.01858**				
			(0.00984)	(0.00775)				
D(tongue)					0.05619***	-0.00062		

					(0.00749)	(0.00571)		
D(migrant)					-0.00304	0.00524		
					(0.00586)	(0.00503)		
D(social)					0.00869	0.00136		
					(0.00557)	(0.00489)		
Station_distx1000							-0.03321*	0.00022
							(0.01779)	(0.01321)
Muni_persx1000							0.00049	0.00200***
							(0.00084)	(0.00064)
L(budget_dist_pc)							0.02937***	-0.00140
							(0.00505)	(0.00409)
D(sdp_dist)							0.02688*	-0.02814***
							(0.01520)	(0.00912)
Constant	0.31792**	-0.56283***	0.36726	-0.46467**	0.26949*	-0.56057***	0.19661	-0.54660***
	(0.15941)	(0.15108)	(0.24207)	(0.21199)	(0.15566)	(0.15102)	(0.16023)	(0.15090)
Observations	2,222		2,222		2,222		2,222	
LogL	3270.640		3364.251		3311.910		3308.613	

c) IV models with labour supply (OLS version)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.00113	-0.00432	0.00618	-0.00462	0.00147	-0.00399	0.00141	-0.00420
	(0.01376)	(0.01072)	(0.01334)	(0.01066)	(0.01381)	(0.01071)	(0.01375)	(0.01068)
L(expenditure)	-0.05396***	0.10977***	-0.06624***	0.10511***	-0.04996***	0.10907***	-0.05356***	0.10860***
	(0.00632)	(0.00567)	(0.00626)	(0.00572)	(0.00643)	(0.00576)	(0.00633)	(0.00566)
L(fam_size)	-0.01288	0.02753	-0.00453	0.02438	-0.02013	0.02826	-0.01271	0.02593
	(0.02636)	(0.01988)	(0.02533)	(0.01961)	(0.02654)	(0.02001)	(0.02633)	(0.01976)
D(mem_son)	-0.01114	0.01594	-0.01622	0.01805	-0.00831	0.01573	-0.01148	0.01621
	(0.01897)	(0.01556)	(0.01854)	(0.01530)	(0.01899)	(0.01561)	(0.01899)	(0.01554)
D(mem_oth)	0.01795	-0.01994	0.02093	-0.02228	0.02009	-0.02028	0.01713	-0.01972
	(0.02154)	(0.01372)	(0.02162)	(0.01417)	(0.02140)	(0.01371)	(0.02159)	(0.01371)
P(men)	-0.02277	-0.02928*	-0.01488	-0.03052*	-0.01897	-0.03046*	-0.02087	-0.03014*
	(0.02266)	(0.01719)	(0.02196)	(0.01695)	(0.02259)	(0.01720)	(0.02264)	(0.01716)
D(gender_head)	-0.02617	0.01883	-0.02773	0.01713	-0.02435	0.01806	-0.02650	0.01629
	(0.02325)	(0.01521)	(0.02226)	(0.01543)	(0.02375)	(0.01531)	(0.02322)	(0.01535)
Pcoh_1	0.07905	-0.15401	-0.02489	-0.14690	0.08942	-0.15832	0.06827	-0.14737
	(0.15190)	(0.10932)	(0.15751)	(0.10906)	(0.15540)	(0.11035)	(0.15156)	(0.10955)
Pcoh_2	0.13839	-0.23019**	0.05125	-0.22336**	0.15113	-0.23526**	0.13165	-0.22708**
	(0.14924)	(0.10637)	(0.15511)	(0.10573)	(0.15294)	(0.10745)	(0.14885)	(0.10645)
Pcoh_3	0.22007	-0.23948**	0.12644	-0.22959**	0.22155	-0.24257**	0.21388	-0.23498**
	(0.15130)	(0.10757)	(0.15696)	(0.10705)	(0.15470)	(0.10868)	(0.15095)	(0.10770)
Pcoh_4	0.17873	-0.18780*	0.09890	-0.18153*	0.17760	-0.18936*	0.17093	-0.18343*
	(0.15212)	(0.10906)	(0.15762)	(0.10864)	(0.15550)	(0.11015)	(0.15185)	(0.10918)
Pcoh_5	0.06917	-0.05505	-0.01144	-0.05228	0.06751	-0.05586	0.06311	-0.05062
	(0.14835)	(0.10628)	(0.15425)	(0.10579)	(0.15175)	(0.10735)	(0.14784)	(0.10636)
Pcoh_6	0.06121	-0.04981	-0.02912	-0.04618	0.06085	-0.05145	0.05122	-0.04615
	(0.15121)	(0.10820)	(0.15658)	(0.10789)	(0.15453)	(0.10923)	(0.15079)	(0.10829)
Pcoh_7	0.09336	-0.01958	0.00717	-0.02220	0.09380	-0.02116	0.08514	-0.01604
	(0.15247)	(0.11001)	(0.15741)	(0.10982)	(0.15572)	(0.11102)	(0.15211)	(0.11016)
Pcoh_8	0.09390	0.01228	0.01263	0.00691	0.09443	0.01012	0.08414	0.01601
	(0.15092)	(0.10846)	(0.15670)	(0.10805)	(0.15429)	(0.10955)	(0.15050)	(0.10858)
Hrs_both	0.00007	0.00019**	0.00006	0.00021**	0.00005	0.00019**	0.00006	0.00020**
	(0.00011)	(0.00009)	(0.00011)	(0.00009)	(0.00011)	(0.00009)	(0.00011)	(0.00009)

D(Hrs_both)	-0.03056	-0.03721	-0.02267	-0.03240	-0.02855	-0.03774	-0.03089	-0.03678
	(0.03078)	(0.02880)	(0.03165)	(0.02847)	(0.03025)	(0.02873)	(0.03090)	(0.02899)
D(Lima)	0.11012***	0.00622	0.12330***	-0.00745	0.10606***	0.00751	0.10665***	0.00592
	(0.01009)	(0.00838)	(0.01834)	(0.01414)	(0.01023)	(0.00850)	(0.01032)	(0.00852)
L(pop_urb_dist)			0.20417***	-0.05340				
			(0.04409)	(0.03960)				
L(pop_urb_dist)-sq			-0.00841***	0.00234				
			(0.00208)	(0.00185)				
pop_den_distx1000			-0.00283***	0.00096				
			(0.00100)	(0.00083)				
D(not_slum)			-0.00500	0.03822***				
			(0.00749)	(0.00607)				
D(mid_city)			-0.00549	-0.00582				
			(0.00864)	(0.00678)				
D(border)			0.05807***	-0.00416				
			(0.01093)	(0.00944)				
D(tongue)					0.03634***	-0.00580		
					(0.00896)	(0.00728)		
D(migrant)					0.01215	-0.00171		
					(0.00806)	(0.00640)		
D(social)					-0.01229	0.00617		
					(0.00754)	(0.00592)		
Station_distx1000							-0.07084***	0.01492
							(0.02256)	(0.01539)
Muni_persx1000							-0.00079	0.00209***
							(0.00122)	(0.00066)
L(budget_dist_pc)							0.00640	-0.00562
							(0.00675)	(0.00504)
D(sdp_dist)							-0.00820	-0.03837***
							(0.01864)	(0.01294)
Constant	0.86656***	-0.70077***	-0.16038	-0.37387	0.81839***	-0.69169***	0.84705***	-0.66821***
	(0.17088)	(0.12843)	(0.28862)	(0.23263)	(0.17436)	(0.12938)	(0.17213)	(0.12964)
Observations	2,160	2,160	2,160	2,160	2,160	2,160	2,160	2,160
LogL	758.621	1275.637	822.280	1301.578	770.401	1276.560	762.252	1282.255

d) First-stage regression for IV models with labour supply

	L(expenditure)				θ			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
L(fam_size)	0.3314865***	0.3274483***	0.3498183***	0.343997***	-0.019627	-0.0191199	-0.0109274	-0.0166853
	(0.06829)	(0.06566)	(0.06790)	(0.06783)	(0.04015)	(0.04013)	(0.04042)	(0.04028)
D(mem_son)	0.1016992**	0.0996555**	0.0968152**	0.095471**	-0.0055221	-0.0044957	-0.0083805	-0.0073572
	(0.04895)	(0.04847)	(0.04888)	(0.04851)	(0.03073)	(0.03056)	(0.03087)	(0.03081)
D(mem_oth)	-0.0199442	-0.0209157	-0.024051	-0.0214981	-0.0256886	-0.0267965	-0.0286654	-0.0257683
	(0.04782)	(0.04563)	(0.04821)	(0.04764)	(0.03299)	(0.03304)	(0.03307)	(0.03313)
P(men)	0.0678127	0.0798221	0.0712308	0.0764876	0.0040099	0.0041602	0.0038984	0.0050857
	(0.05762)	(0.05652)	(0.05719)	(0.05750)	(0.03387)	(0.03404)	(0.03391)	(0.03401)
D(gender_head)	-0.1129027**	-0.100262**	-0.111196**	-0.111451***	0.1336476***	0.1322544***	0.1338373***	0.134119***
	(0.04431)	(0.04300)	(0.04349)	(0.04423)	(0.03632)	(0.03636)	(0.03663)	(0.03628)
Pcoh_1	-0.9607172***	-1.108212***	-0.9679774***	-1.026393***	0.1680528	0.1776305	0.1604404	0.158449
	(0.31901)	(0.32519)	(0.32381)	(0.31992)	(0.30243)	(0.29557)	(0.30613)	(0.30325)

Pcoh_2	-0.8247098***	-0.9631616***	-0.8227427***	-0.8778704***	-0.0000143	0.0118838	-0.0060389	-0.0075777
	(0.31061)	(0.31755)	(0.31536)	(0.31210)	(0.30194)	(0.29475)	(0.30587)	(0.30270)
Pcoh_3	-0.7421377**	-0.8855057***	-0.7336148**	-0.7939493**	-0.0029513	0.0100085	-0.0014236	-0.011198
	(0.31611)	(0.32247)	(0.32037)	(0.31718)	(0.30097)	(0.29393)	(0.30453)	(0.30170)
Pcoh_4	-0.4231997	-0.5681839*	-0.434056	-0.4842875	0.0292428	0.0368332	0.023862	0.0228387
	(0.31942)	(0.32495)	(0.32397)	(0.32026)	(0.30198)	(0.29480)	(0.30573)	(0.30255)
Pcoh_5	-0.3473459	-0.4937201	-0.3383481	-0.3921458	0.0136673	0.0234119	0.0153098	0.0073704
	(0.31490)	(0.32141)	(0.31947)	(0.31541)	(0.29951)	(0.29238)	(0.30297)	(0.30023)
Pcoh_6	-0.4227694	-0.578925*	-0.408033	-0.4732332	-0.030795	-0.0155911	-0.0275096	-0.0364642
	(0.32368)	(0.33063)	(0.32828)	(0.32479)	(0.30245)	(0.29543)	(0.30587)	(0.30328)
Pcoh_7	-0.3675913	-0.5225369	-0.3560038	-0.426148	-0.0920957	-0.0830867	-0.090108	-0.1001198
	(0.32664)	(0.33347)	(0.33110)	(0.32774)	(0.30630)	(0.29891)	(0.30971)	(0.30715)
Pcoh_8	-0.2934018	-0.4630356	-0.2842821	-0.358195	-0.0984544	-0.0873359	-0.0984918	-0.1071681
	(0.32083)	(0.32827)	(0.32531)	(0.32222)	(0.30252)	(0.29527)	(0.30614)	(0.30338)
Hrs_both	0.0009732***	0.0009344***	0.0009811***	0.0009769***	-0.0016842***	-0.0016844***	-0.0016694***	-0.0016831***
	(0.00030)	(0.00030)	(0.00030)	(0.00030)	(0.00018)	(0.00018)	(0.00018)	(0.00018)
D(Hrs_both)	-0.0587325	-0.0087809	-0.0518329	-0.0707466	0.1089451*	0.1069518*	0.1100467*	0.1080064*
	(0.07295)	(0.07102)	(0.07350)	(0.07278)	(0.05876)	(0.05933)	(0.05870)	(0.05874)
D(Lima)	0.2461642***	0.1665346***	0.2313829***	0.2549116***	-0.0211301	-0.035964	-0.0248415	-0.0194062
	(0.02602)	(0.04671)	(0.02634)	(0.02586)	(0.01504)	(0.02991)	(0.01515)	(0.01526)
theta_educ_prima	0.167793***	0.1463916***	0.1688955***	0.1685395***	0.0398507***	0.0424543***	0.0409871***	0.0399192***
	(0.02162)	(0.02086)	(0.02185)	(0.02148)	(0.01238)	(0.01255)	(0.01243)	(0.01247)
year_educ_female	0.0423834***	0.0327692**	0.0396936***	0.0441876***	0.0571744***	0.058449***	0.0550632***	0.0572829***
	(0.01484)	(0.01445)	(0.01493)	(0.01477)	(0.00876)	(0.00880)	(0.00879)	(0.00882)
year_educ_female-sq	0.0011286*	0.0012735**	0.0012608*	0.0010182	-0.0029823***	-0.0030225***	-0.0029031***	-0.0029902***
	(0.00064)	(0.00063)	(0.00065)	(0.00064)	(0.00037)	(0.00038)	(0.00038)	(0.00038)
D(risk7)	0.1612495***	0.1605618***	0.1561633***	0.1672851***	0.045552*	0.0446508*	0.0443377*	0.045974*
	(0.05041)	(0.04966)	(0.04934)	(0.05002)	(0.02667)	(0.02661)	(0.02665)	(0.02676)
L(asset)	0.1871591***	0.1818678***	0.1850781***	0.186949***	-0.0043055	-0.0037449	-0.0054448	-0.0044358
	(0.00796)	(0.00767)	(0.00797)	(0.00792)	(0.00444)	(0.00446)	(0.00441)	(0.00444)
L(pop_urb_dist)		0.4129241***				-0.0599243		
		(0.11724)				(0.06292)		
L(pop_urb_dist)-sq		-0.0170243***				0.0024628		
		(0.00553)				(0.00303)		
pop_den_distx1000		0.0068171***				0.0008579		
		(0.00256)				(0.00162)		
D(not_slum)		0.0817529***				0.0075363		
		(0.02041)				(0.01234)		
D(mid_city)		0.0291951				-0.0248511*		
		(0.02359)				(0.01355)		
D(border)		0.0812128***				0.0103931		
		(0.03009)				(0.01843)		
D(tongue)			-0.0465811*				-0.036889**	
			(0.02362)				(0.01425)	
D(migrant)			0.0654324***				0.0163802	
			(0.02119)				(0.01233)	

D(social)			-0.0438832**				-0.0119492	
			(0.01994)				(0.01153)	
Station_distx1000				-0.0750482				-0.0041782
				(0.05114)				(0.04877)
Muni_persx1000				0.008172***				0.0020579*
				(0.00265)				(0.00124)
L(budget_dist_pc)				0.0285633*				-0.0014451
				(0.01628)				(0.00997)
D(sdp_dist)				0.0482501				0.0182764
				(0.05392)				(0.02868)
Constant	7.170149***	4.901359***	7.146818***	7.045006***	0.4920025	0.8311044*	0.5044984	0.4973546
	(0.36550)	(0.68447)	(0.36948)	(0.37131)	(0.32028)	(0.42403)	(0.32373)	(0.32775)
Observations	2,160	2,160	2,160	2,160	2,160	2,160	2,160	2,160
R-squared	0.5119	0.5367	0.5159	0.5168	0.1312	0.1353	0.1351	0.1322

e) Second-stage regression for IV models with labour supply (2SLS version)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.22711***	-0.08066	0.18594***	-0.05413	0.22732***	-0.08037	0.23379***	-0.07844
	(0.06979)	(0.05300)	(0.06579)	(0.05193)	(0.07109)	(0.05410)	(0.06983)	(0.05258)
L(expenditure)	-0.04847***	0.15254***	-0.06364***	0.15344***	-0.03895***	0.15310***	-0.04681***	0.15153***
	(0.01184)	(0.00861)	(0.01166)	(0.00886)	(0.01238)	(0.00904)	(0.01194)	(0.00858)
L(fam_size)	-0.01142	0.01997	-0.00347	0.01664	-0.02203	0.01840	-0.01226	0.01815
	(0.02826)	(0.02092)	(0.02668)	(0.02045)	(0.02860)	(0.02115)	(0.02838)	(0.02075)
D(mem_son)	-0.00644	-0.00268	-0.01235	0.00045	-0.00475	-0.00238	-0.00649	-0.00224
	(0.02072)	(0.01644)	(0.01973)	(0.01609)	(0.02076)	(0.01647)	(0.02084)	(0.01640)
D(mem_oth)	0.02334	-0.02003	0.02533	-0.02123	0.02618	-0.01968	0.02272	-0.01976
	(0.02239)	(0.01439)	(0.02194)	(0.01463)	(0.02230)	(0.01444)	(0.02247)	(0.01435)
P(men)	-0.02441	-0.03199*	-0.01634	-0.03398**	-0.02139	-0.03212*	-0.02283	-0.03313*
	(0.02390)	(0.01752)	(0.02282)	(0.01725)	(0.02384)	(0.01757)	(0.02394)	(0.01749)
D(gender_head)	-0.05666**	0.03483**	-0.05200**	0.02953*	-0.05441**	0.03483**	-0.05798**	0.03222*
	(0.02611)	(0.01727)	(0.02429)	(0.01714)	(0.02657)	(0.01746)	(0.02610)	(0.01737)
Pcoh_1	0.04027	-0.08290	-0.05782	-0.06368	0.05728	-0.08145	0.03191	-0.07498
	(0.16402)	(0.12158)	(0.16130)	(0.11527)	(0.16786)	(0.12135)	(0.16468)	(0.12175)
Pcoh_2	0.13777	-0.17462	0.04824	-0.15289	0.15578	-0.17377	0.13376	-0.16966
	(0.16151)	(0.11863)	(0.15841)	(0.11211)	(0.16567)	(0.11841)	(0.16216)	(0.11870)
Pcoh_3	0.21189	-0.17587	0.11752	-0.15222	0.21809	-0.17569	0.20882	-0.17020
	(0.16317)	(0.11960)	(0.16028)	(0.11312)	(0.16694)	(0.11941)	(0.16386)	(0.11973)
Pcoh_4	0.16110	-0.13653	0.08363	-0.11955	0.16519	-0.13526	0.15445	-0.12958
	(0.16335)	(0.12072)	(0.16043)	(0.11442)	(0.16709)	(0.12049)	(0.16398)	(0.12075)
Pcoh_5	0.05777	-0.01464	-0.02267	0.00217	0.05891	-0.01485	0.05353	-0.00906
	(0.16019)	(0.11889)	(0.15737)	(0.11203)	(0.16374)	(0.11857)	(0.16061)	(0.11888)
Pcoh_6	0.06561	-0.01662	-0.02903	0.00547	0.06694	-0.01719	0.05736	-0.01114
	(0.16314)	(0.12065)	(0.15963)	(0.11413)	(0.16664)	(0.12035)	(0.16367)	(0.12067)
Pcoh_7	0.12307	-0.01355	0.02852	0.00547	0.12342	-0.01404	0.11746	-0.00740

	(0.16552)	(0.12233)	(0.16125)	(0.11596)	(0.16894)	(0.12202)	(0.16621)	(0.12239)
Pcoh_8	0.12168	0.02884	0.03182	0.04577	0.12299	0.02852	0.11477	0.03531
	(0.16344)	(0.12118)	(0.15995)	(0.11464)	(0.16701)	(0.12093)	(0.16407)	(0.12126)
Hrs_both	0.00043***	-0.00001	0.00034**	0.00005	0.00039**	-0.00001	0.00043***	-0.00000
	(0.00016)	(0.00012)	(0.00015)	(0.00012)	(0.00016)	(0.00012)	(0.00016)	(0.00012)
D(Hrs_both)	-0.05226	-0.02542	-0.03907	-0.02794	-0.05045	-0.02594	-0.05325	-0.02420
	(0.03479)	(0.02981)	(0.03424)	(0.02968)	(0.03425)	(0.02985)	(0.03515)	(0.02999)
D(Lima)	0.11116***	-0.00590	0.12881***	-0.02063	0.10723***	-0.00508	0.10717***	-0.00698
	(0.01089)	(0.00831)	(0.01955)	(0.01444)	(0.01102)	(0.00841)	(0.01122)	(0.00847)
L(pop_urb_dist)			0.21308***	-0.09547**				
			(0.04536)	(0.04074)				
L(pop_urb_dist)-sq			-0.00878***	0.00413**				
			(0.00214)	(0.00190)				
pop_den_distx1000			-0.00301***	0.00085				
			(0.00107)	(0.00083)				
D(not_slum)			-0.00243	0.02351***				
			(0.00809)	(0.00663)				
D(mid_city)			0.00081	-0.01141				
			(0.00924)	(0.00731)				
D(border)			0.05669***	-0.00754				
			(0.01144)	(0.00973)				
D(tongue)					0.04299***	0.00407		
					(0.01012)	(0.00802)		
D(migrant)					0.00737	-0.00412		
					(0.00876)	(0.00677)		
D(social)					-0.00645	0.00256		
					(0.00812)	(0.00621)		
Station_distx1000							-0.06920***	0.01157
							(0.02304)	(0.01645)
Muni_persx1000							-0.00119	0.00157**
							(0.00113)	(0.00065)
L(budget_dist_pc)							0.00800	-0.00794
							(0.00705)	(0.00514)
D(sdp_dist)							-0.01637	-0.03431**
							(0.01961)	(0.01365)
Constant	0.65278***	-1.06303***	-0.36903	-0.56563**	0.55537**	-1.06569***	0.60984***	-1.02225***
	(0.23116)	(0.16762)	(0.30961)	(0.25717)	(0.23700)	(0.16986)	(0.23518)	(0.16757)
Observations	2,160	2,160	2,160	2,160	2,160	2,160	2,160	2,160
LogL	631.231	1207.603	735.605	1242.857	642.179	1207.132	627.711	1215.471

f) Second-stage regression for IV models with labour supply (LIML version)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.24249***	-0.08205	0.19895***	-0.05451	0.23442***	-0.08189	0.24971***	-0.07999
	(0.07523)	(0.05419)	(0.07104)	(0.05251)	(0.07362)	(0.05546)	(0.07536)	(0.05394)

L(expenditure)	-0.04713***	0.15250***	-0.06250***	0.15346***	-0.03827***	0.15306***	-0.04541***	0.15149***
	(0.01225)	(0.00868)	(0.01203)	(0.00889)	(0.01259)	(0.00913)	(0.01236)	(0.00866)
L(fam_size)	-0.01148	0.01997	-0.00353	0.01664	-0.02217	0.01841	-0.01239	0.01814
	(0.02850)	(0.02093)	(0.02686)	(0.02045)	(0.02871)	(0.02117)	(0.02864)	(0.02076)
D(mem_son)	-0.00649	-0.00271	-0.01239	0.00043	-0.00476	-0.00241	-0.00649	-0.00228
	(0.02091)	(0.01645)	(0.01987)	(0.01609)	(0.02085)	(0.01648)	(0.02105)	(0.01641)
D(mem_oth)	0.02375	-0.02007	0.02569	-0.02124	0.02639	-0.01972	0.02314	-0.01980
	(0.02254)	(0.01439)	(0.02203)	(0.01463)	(0.02237)	(0.01444)	(0.02263)	(0.01435)
P(men)	-0.02460	-0.03198*	-0.01652	-0.03398**	-0.02148	-0.03210*	-0.02304	-0.03312*
	(0.02408)	(0.01753)	(0.02295)	(0.01725)	(0.02392)	(0.01758)	(0.02413)	(0.01749)
D(gender_head)	-0.05861**	0.03502**	-0.05365**	0.02958*	-0.05530**	0.03504**	-0.06002**	0.03243*
	(0.02652)	(0.01734)	(0.02462)	(0.01717)	(0.02676)	(0.01754)	(0.02653)	(0.01745)
Pcoh_1	0.03889	-0.08267	-0.05877	-0.06358	0.05673	-0.08120	0.03069	-0.07473
	(0.16575)	(0.12179)	(0.16219)	(0.11531)	(0.16867)	(0.12157)	(0.16654)	(0.12198)
Pcoh_2	0.13893	-0.17462	0.04937	-0.15285	0.15637	-0.17377	0.13511	-0.16967
	(0.16326)	(0.11883)	(0.15929)	(0.11215)	(0.16649)	(0.11863)	(0.16405)	(0.11893)
Pcoh_3	0.21266	-0.17583	0.11832	-0.15216	0.21844	-0.17563	0.20979	-0.17016
	(0.16487)	(0.11980)	(0.16115)	(0.11316)	(0.16774)	(0.11963)	(0.16570)	(0.11996)
Pcoh_4	0.16089	-0.13643	0.08364	-0.11949	0.16515	-0.13515	0.15433	-0.12946
	(0.16503)	(0.12091)	(0.16126)	(0.11446)	(0.16788)	(0.12070)	(0.16578)	(0.12097)
Pcoh_5	0.05779	-0.01457	-0.02249	0.00222	0.05891	-0.01477	0.05368	-0.00899
	(0.16189)	(0.11909)	(0.15822)	(0.11208)	(0.16453)	(0.11879)	(0.16245)	(0.11911)
Pcoh_6	0.06666	-0.01665	-0.02802	0.00550	0.06738	-0.01721	0.05855	-0.01117
	(0.16486)	(0.12086)	(0.16048)	(0.11418)	(0.16744)	(0.12058)	(0.16552)	(0.12089)
Pcoh_7	0.12544	-0.01374	0.03072	0.00544	0.12447	-0.01423	0.12007	-0.00761
	(0.16736)	(0.12254)	(0.16222)	(0.11601)	(0.16979)	(0.12226)	(0.16819)	(0.12262)
Pcoh_8	0.12414	0.02867	0.03407	0.04575	0.12408	0.02834	0.11747	0.03511
	(0.16522)	(0.12140)	(0.16086)	(0.11469)	(0.16785)	(0.12116)	(0.16600)	(0.12150)
Hrs_both	0.00045***	-0.00001	0.00036**	0.00005	0.00040**	-0.00001	0.00045***	-0.00000
	(0.00016)	(0.00012)	(0.00015)	(0.00012)	(0.00016)	(0.00013)	(0.00016)	(0.00012)
D(Hrs_both)	-0.05364	-0.02529	-0.04026	-0.02790	-0.05111	-0.02579	-0.05467	-0.02405
	(0.03529)	(0.02984)	(0.03460)	(0.02969)	(0.03448)	(0.02989)	(0.03568)	(0.03003)
D(Lima)	0.11098***	-0.00590	0.12899***	-0.02065	0.10719***	-0.00509	0.10694***	-0.00698
	(0.01100)	(0.00831)	(0.01971)	(0.01444)	(0.01106)	(0.00841)	(0.01133)	(0.00848)
L(pop_urb_dist)			0.21296***	-0.09550**				
			(0.04555)	(0.04074)				
L(pop_urb_dist)-sq			-0.00878***	0.00413**				
			(0.00215)	(0.00190)				
pop_den_distx1000			-0.00303***	0.00085				
			(0.00108)	(0.00083)				
D(not_slum)			-0.00251	0.02350***				
			(0.00814)	(0.00663)				
D(mid_city)			0.00119	-0.01143				
			(0.00932)	(0.00731)				
D(border)			0.05652***	-0.00754				
			(0.01152)	(0.00973)				

D(tongue)					0.04328***	0.00403		
					(0.01019)	(0.00804)		
D(migrant)					0.00719	-0.00409		
					(0.00880)	(0.00678)		
D(social)					-0.00628	0.00252		
					(0.00816)	(0.00622)		
Station_distx1000							-0.06915***	0.01156
							(0.02327)	(0.01646)
Muni_persx1000							-0.00123	0.00157**
							(0.00113)	(0.00065)
L(budget_dist_pc)							0.00807	-0.00795
							(0.00710)	(0.00514)
D(sdp_dist)							-0.01690	-0.03425**
							(0.01977)	(0.01366)
Constant	0.62877***	-1.06167***	-0.38896	-0.56531**	0.54388**	-1.06416***	0.58465**	-1.02072***
	(0.23758)	(0.16877)	(0.31313)	(0.25750)	(0.24004)	(0.17124)	(0.24199)	(0.16887)
Observations	2,160	2,160	2,160	2,160	2,160	2,160	2,160	2,160
LogL	614.453	1206.746	723.181	1242.653	634.453	1206.201	609.902	1214.526

g) Second-stage regression for IV models with labour supply (GMM version)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.23632***	-0.07939	0.19682***	-0.05441	0.23618***	-0.07959	0.24415***	-0.07797
	(0.06928)	(0.05293)	(0.06542)	(0.05186)	(0.07052)	(0.05404)	(0.06935)	(0.05251)
L(expenditure)	-0.04810***	0.15334***	-0.06265***	0.15393***	-0.03819***	0.15404***	-0.04641***	0.15225***
	(0.01181)	(0.00859)	(0.01164)	(0.00884)	(0.01236)	(0.00901)	(0.01192)	(0.00855)
L(fam_size)	-0.01005	0.01858	-0.00265	0.01571	-0.02166	0.01695	-0.01062	0.01655
	(0.02825)	(0.02088)	(0.02667)	(0.02042)	(0.02858)	(0.02111)	(0.02836)	(0.02070)
D(mem_son)	-0.00696	-0.00191	-0.01258	0.00079	-0.00495	-0.00159	-0.00704	-0.00166
	(0.02070)	(0.01637)	(0.01973)	(0.01603)	(0.02075)	(0.01641)	(0.02082)	(0.01634)
D(mem_oth)	0.02298	-0.02043	0.02590	-0.02175	0.02632	-0.02000	0.02238	-0.02025
	(0.02238)	(0.01438)	(0.02193)	(0.01461)	(0.02229)	(0.01443)	(0.02246)	(0.01433)
P(men)	-0.02428	-0.03168*	-0.01640	-0.03376*	-0.02150	-0.03160*	-0.02277	-0.03284*
	(0.02389)	(0.01751)	(0.02280)	(0.01724)	(0.02383)	(0.01756)	(0.02393)	(0.01747)
D(gender_head)	-0.05599**	0.03392**	-0.05256**	0.02906*	-0.05439**	0.03372*	-0.05755**	0.03130*
	(0.02608)	(0.01725)	(0.02426)	(0.01713)	(0.02655)	(0.01743)	(0.02608)	(0.01734)
Pcoh_1	0.02184	-0.08498	-0.05882	-0.06607	0.04779	-0.08223	0.01308	-0.07594
	(0.16376)	(0.12149)	(0.16107)	(0.11506)	(0.16774)	(0.12128)	(0.16440)	(0.12167)
Pcoh_2	0.12194	-0.17581	0.04894	-0.15432	0.14793	-0.17411	0.11751	-0.16975
	(0.16128)	(0.11855)	(0.15819)	(0.11191)	(0.16556)	(0.11834)	(0.16191)	(0.11863)
Pcoh_3	0.19203	-0.17837	0.11500	-0.15483	0.20756	-0.17734	0.18863	-0.17169
	(0.16289)	(0.11949)	(0.16000)	(0.11289)	(0.16678)	(0.11933)	(0.16356)	(0.11964)
Pcoh_4	0.14532	-0.13685	0.08261	-0.12002	0.15639	-0.13514	0.13812	-0.12831
	(0.16312)	(0.12064)	(0.16024)	(0.11423)	(0.16696)	(0.12041)	(0.16373)	(0.12068)
Pcoh_5	0.04162	-0.01790	-0.02200	-0.00084	0.05030	-0.01730	0.03724	-0.01171

	(0.15998)	(0.11881)	(0.15718)	(0.11186)	(0.16363)	(0.11852)	(0.16040)	(0.11882)
Pcoh_6	0.05276	-0.02001	-0.02568	0.00228	0.06061	-0.01980	0.04468	-0.01394
	(0.16299)	(0.12055)	(0.15943)	(0.11395)	(0.16657)	(0.12029)	(0.16351)	(0.12058)
Pcoh_7	0.11145	-0.01742	0.03347	0.00185	0.11815	-0.01677	0.10622	-0.01095
	(0.16538)	(0.12223)	(0.16107)	(0.11582)	(0.16888)	(0.12197)	(0.16605)	(0.12230)
Pcoh_8	0.10852	0.02633	0.03574	0.04311	0.11682	0.02698	0.10182	0.03352
	(0.16327)	(0.12106)	(0.15975)	(0.11443)	(0.16694)	(0.12085)	(0.16390)	(0.12116)
Hrs_both	0.00046***	-0.00001	0.00037**	0.00004	0.00042***	-0.00001	0.00046***	-0.00001
	(0.00016)	(0.00012)	(0.00015)	(0.00012)	(0.00016)	(0.00012)	(0.00016)	(0.00012)
D(Hrs_both)	-0.05373	-0.02533	-0.04202	-0.02763	-0.05191	-0.02567	-0.05507	-0.02359
	(0.03472)	(0.02970)	(0.03420)	(0.02960)	(0.03415)	(0.02974)	(0.03507)	(0.02989)
D(Lima)	0.11066***	-0.00645	0.12918***	-0.02157	0.10695***	-0.00578	0.10675***	-0.00762
	(0.01087)	(0.00830)	(0.01951)	(0.01440)	(0.01100)	(0.00840)	(0.01119)	(0.00846)
L(pop_urb_dist)			0.21330***	-0.09729**				
			(0.04531)	(0.04061)				
L(pop_urb_dist)-sq			-0.00878***	0.00421**				
			(0.00214)	(0.00190)				
pop_den_distx1000			-0.00315***	0.00087				
			(0.00107)	(0.00083)				
D(not_slum)			-0.00301	0.02326***				
			(0.00808)	(0.00663)				
D(mid_city)			0.00107	-0.01154				
			(0.00923)	(0.00730)				
D(border)			0.05620***	-0.00797				
			(0.01144)	(0.00972)				
D(tongue)					0.04351***	0.00436		
					(0.01011)	(0.00801)		
D(migrant)					0.00662	-0.00379		
					(0.00874)	(0.00677)		
D(social)					-0.00599	0.00242		
					(0.00811)	(0.00620)		
Station_distx1000							-0.06696***	0.01206
							(0.02301)	(0.01642)
Muni_persx1000							-0.00121	0.00157**
							(0.00113)	(0.00065)
L(budget_dist_pc)							0.00840	-0.00845*
							(0.00704)	(0.00511)
D(sdp_dist)							-0.01747	-0.03456**
							(0.01959)	(0.01364)
Constant	0.65441***	-1.06713***	-0.39099	-0.55626**	0.54845**	-1.07175***	0.60836***	-1.02302***
	(0.23084)	(0.16757)	(0.30925)	(0.25680)	(0.23673)	(0.16978)	(0.23487)	(0.16753)
Observations	2,160	2,160	2,160	2,160	2,160	2,160	2,160	2,160
LogL	621.252	1206.879	725.228	1241.792	632.501	1205.832	616.202	1214.369

h) IV models without labour supply (OLS version)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
θ	-0.00061	-0.00925	0.00482	-0.01018	0.00040	-0.00907	-0.00010	-0.00936
	(0.01342)	(0.01051)	(0.01300)	(0.01047)	(0.01346)	(0.01050)	(0.01340)	(0.01047)
L(expenditure)	-0.05365***	0.11065***	-0.06604***	0.10618***	-0.04978***	0.11004***	-0.05326***	0.10958***
	(0.00630)	(0.00558)	(0.00621)	(0.00562)	(0.00641)	(0.00567)	(0.00632)	(0.00558)
L(fam_size)	-0.01188	0.02827	-0.00376	0.02491	-0.01907	0.02892	-0.01166	0.02662
	(0.02626)	(0.01995)	(0.02521)	(0.01964)	(0.02645)	(0.02010)	(0.02623)	(0.01983)
D(mem_son)	-0.01108	0.01627	-0.01618	0.01840	-0.00831	0.01612	-0.01147	0.01650
	(0.01899)	(0.01555)	(0.01856)	(0.01530)	(0.01901)	(0.01560)	(0.01901)	(0.01553)
D(mem_oth)	0.01848	-0.01986	0.02131	-0.02240	0.02070	-0.02019	0.01771	-0.01971
	(0.02153)	(0.01354)	(0.02165)	(0.01403)	(0.02138)	(0.01354)	(0.02157)	(0.01353)
P(men)	-0.02298	-0.02953*	-0.01502	-0.03070*	-0.01915	-0.03063*	-0.02111	-0.03038*
	(0.02265)	(0.01720)	(0.02194)	(0.01697)	(0.02258)	(0.01721)	(0.02263)	(0.01717)
D(gender_head)	-0.02584	0.02015	-0.02744	0.01872	-0.02423	0.01946	-0.02626	0.01770
	(0.02316)	(0.01544)	(0.02220)	(0.01564)	(0.02367)	(0.01555)	(0.02313)	(0.01558)
Pcoh_1	0.08058	-0.16045	-0.02417	-0.15607	0.09251	-0.16489	0.07069	-0.15458
	(0.15186)	(0.10938)	(0.15788)	(0.10898)	(0.15537)	(0.11032)	(0.15144)	(0.10966)
Pcoh_2	0.13832	-0.23893**	0.05082	-0.23457**	0.15284	-0.24409**	0.13250	-0.23648**
	(0.14922)	(0.10655)	(0.15549)	(0.10576)	(0.15291)	(0.10755)	(0.14874)	(0.10670)
Pcoh_3	0.22165	-0.24426**	0.12729	-0.23650**	0.22436	-0.24753**	0.21618	-0.24035**
	(0.15141)	(0.10788)	(0.15738)	(0.10718)	(0.15477)	(0.10889)	(0.15098)	(0.10808)
Pcoh_4	0.18130	-0.19025*	0.10047	-0.18615*	0.18106	-0.19204*	0.17431	-0.18613*
	(0.15232)	(0.10936)	(0.15809)	(0.10876)	(0.15568)	(0.11034)	(0.15198)	(0.10953)
Pcoh_5	0.07146	-0.06076	-0.01014	-0.06086	0.07136	-0.06185	0.06629	-0.05706
	(0.14827)	(0.10668)	(0.15457)	(0.10599)	(0.15165)	(0.10763)	(0.14768)	(0.10681)
Pcoh_6	0.06496	-0.05424	-0.02672	-0.05374	0.06617	-0.05612	0.05594	-0.05137
	(0.15108)	(0.10861)	(0.15683)	(0.10807)	(0.15436)	(0.10953)	(0.15056)	(0.10875)
Pcoh_7	0.09645	-0.02475	0.00906	-0.03045	0.09852	-0.02657	0.08921	-0.02195
	(0.15226)	(0.11026)	(0.15759)	(0.10985)	(0.15550)	(0.11117)	(0.15181)	(0.11046)
Pcoh_8	0.09832	0.00781	0.01550	-0.00092	0.10055	0.00540	0.08964	0.01069
	(0.15062)	(0.10861)	(0.15685)	(0.10800)	(0.15399)	(0.10959)	(0.15011)	(0.10877)
D(Lima)	0.11024***	0.00620	0.12341***	-0.00682	0.10616***	0.00738	0.10668***	0.00572
	(0.01007)	(0.00837)	(0.01832)	(0.01416)	(0.01022)	(0.00849)	(0.01031)	(0.00852)
L(pop_urb_dist)			0.20489***	-0.05205				
			(0.04413)	(0.03965)				
L(pop_urb_dist)-sq			-0.00845***	0.00226				
			(0.00208)	(0.00186)				
pop_den_distx1000			-0.00282***	0.00097				
			(0.00100)	(0.00083)				
D(not_slum)			-0.00491	0.03790***				
			(0.00746)	(0.00608)				
D(mid_city)			-0.00554	-0.00596				
			(0.00864)	(0.00679)				
D(border)			0.05846***	-0.00266				
			(0.01089)	(0.00934)				
D(tongue)					0.03638***	-0.00526		
					(0.00895)	(0.00727)		
D(migrant)					0.01237	-0.00131		
					(0.00806)	(0.00642)		
D(social)					-0.01241*	0.00581		
					(0.00753)	(0.00592)		

Station_distx1000							-0.07139***	0.01314
							(0.02247)	(0.01535)
Muni_persx1000							-0.00079	0.00208***
							(0.00121)	(0.00066)
L(budget_dist_pc)							0.00605	-0.00605
							(0.00673)	(0.00503)
D(sdp_dist)							-0.00816	-0.03789***
							(0.01859)	(0.01299)
Constant	0.83666***	-0.72322***	-0.18545	-0.39408*	0.78751***	-0.71492***	0.81708***	-0.68720***
	(0.16789)	(0.12584)	(0.28722)	(0.23099)	(0.17160)	(0.12676)	(0.16929)	(0.12755)
Observations	2,160	2,160	2,160	2,160	2,160	2,160	2,160	2,160
LogL	758.034	1272.613	821.929	1298.135	769.942	1273.392	761.685	1279.041

i) First-stage regression without labour supply

	L(expenditure)				θ			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
L(fam_size)	0.3330221***	0.3281614***	0.3506159***	0.3454798***	-0.0149537	-0.0158153	-0.0055746	-0.01258
	(0.06908)	(0.06636)	(0.06874)	(0.06864)	(0.04099)	(0.04103)	(0.04127)	(0.04110)
D(mem_son)	0.1046536**	0.1020674**	0.1000008**	0.0985517**	-0.0141545	-0.0124572	-0.0172818	-0.0155529
	(0.04901)	(0.04854)	(0.04894)	(0.04862)	(0.03082)	(0.03073)	(0.03099)	(0.03093)
D(mem_oth)	-0.0261911	-0.0285105	-0.0303384	-0.0268998	-0.0204745	-0.0220951	-0.0240487	-0.0202913
	(0.04794)	(0.04597)	(0.04834)	(0.04776)	(0.03324)	(0.03328)	(0.03332)	(0.03337)
P(men)	0.0682079	0.0804893	0.0719048	0.0770331	0.0033165	0.0033536	0.0028866	0.003846
	(0.05776)	(0.05669)	(0.05735)	(0.05765)	(0.03486)	(0.03500)	(0.03493)	(0.03499)
D(gender_head)	-0.1099522**	-0.0969398**	-0.1080963**	-0.1084435**	0.1276912***	0.1261593***	0.1278621***	0.1280251***
	(0.04528)	(0.04402)	(0.04444)	(0.04515)	(0.03668)	(0.03659)	(0.03700)	(0.03668)
Pcoh_1	-1.049436***	-1.202527***	-1.058106***	-1.114161***	0.3200921	0.3282947	0.3088071	0.3141628
	(0.31621)	(0.32275)	(0.32052)	(0.31633)	(0.31035)	(0.30452)	(0.31470)	(0.31114)
Pcoh_2	-0.9186865***	-1.06156***	-0.9177212***	-0.9693976***	0.150558	0.1606509	0.140574	0.1451287
	(0.30916)	(0.31612)	(0.31339)	(0.30986)	(0.31046)	(0.30420)	(0.31493)	(0.31123)
Pcoh_3	-0.8224865**	-0.9720408***	-0.8159825**	-0.8709521***	0.1155721	0.1270132	0.11502	0.109844
	(0.31485)	(0.32113)	(0.31854)	(0.31508)	(0.30863)	(0.30257)	(0.31279)	(0.30935)
Pcoh_4	-0.4922896	-0.6451437**	-0.505024	-0.5493352*	0.127781	0.1355357	0.1203174	0.1232849
	(0.31824)	(0.32352)	(0.32232)	(0.31829)	(0.31162)	(0.30529)	(0.31592)	(0.31225)
Pcoh_5	-0.4394578	-0.5952918*	-0.4330609	-0.4798477	0.1509307	0.160218	0.1501653	0.1467775
	(0.31407)	(0.32035)	(0.31809)	(0.31384)	(0.30753)	(0.30137)	(0.31157)	(0.30824)
Pcoh_6	-0.5143102	-0.6822801**	-0.502565	-0.5601176*	0.1065411	0.1206193	0.1072454	0.1037114
	(0.32309)	(0.32980)	(0.32719)	(0.32337)	(0.31063)	(0.30472)	(0.31462)	(0.31146)
Pcoh_7	-0.4554262	-0.6212727*	-0.4466062	-0.5094374	0.0385438	0.0467428	0.0379829	0.0332241
	(0.32642)	(0.33296)	(0.33034)	(0.32666)	(0.31455)	(0.30830)	(0.31858)	(0.31540)
Pcoh_8	-0.3873419	-0.569908*	-0.3812901	-0.4469433	0.0420605	0.0508546	0.0388804	0.0363892
	(0.32001)	(0.32743)	(0.32398)	(0.32051)	(0.31089)	(0.30476)	(0.31508)	(0.31177)
D(Lima)	0.2464175***	0.1726248***	0.2314304***	0.2546095***	-0.0208688	-0.0438865	-0.0243137	-0.0182753
	(0.02580)	(0.04644)	(0.02612)	(0.02563)	(0.01557)	(0.03021)	(0.01569)	(0.01583)
theta_educ_prima	0.1663211***	0.1453726***	0.1673245***	0.1668861***	0.0424185***	0.0449184***	0.0437612***	0.0426734***
	(0.02182)	(0.02107)	(0.02203)	(0.02169)	(0.01258)	(0.01275)	(0.01261)	(0.01267)
year_educ_female	0.0432046***	0.0336106**	0.0407131***	0.044965***	0.0559441***	0.0569651***	0.0534841***	0.0561378***

	(0.01491)	(0.01451)	(0.01497)	(0.01483)	(0.00884)	(0.00886)	(0.00887)	(0.00888)
year_educ_female-sq	0.0010482	0.0011952*	0.0011745*	0.0009409	-0.0028631***	-0.0028942***	-0.0027744***	-0.0028747***
	(0.00064)	(0.00063)	(0.00065)	(0.00064)	(0.00038)	(0.00038)	(0.00038)	(0.00038)
D(risk7)	0.1695586***	0.1688514***	0.1644345***	0.1755826***	0.0328389	0.0313388	0.0318679	0.0330192
	(0.05029)	(0.04964)	(0.04926)	(0.04985)	(0.02649)	(0.02648)	(0.02638)	(0.02658)
L(asset)	0.1902762***	0.1847394***	0.188285***	0.1902971***	-0.0110543**	-0.0104043**	-0.012256***	-0.0113107**
	(0.00792)	(0.00762)	(0.00793)	(0.00787)	(0.00454)	(0.00457)	(0.00453)	(0.00455)
D(house1)	-0.0199746	-0.0246864	-0.0206102	-0.014854	-0.0113487	-0.0088106	-0.0112565	-0.0110716
	(0.03275)	(0.03166)	(0.03244)	(0.03282)	(0.02201)	(0.02218)	(0.02190)	(0.02212)
L(pop_urb_dist)		0.4170538***				-0.0694094		
		(0.11731)				(0.06470)		
L(pop_urb_dist)-sq		-0.0172822***				0.0029923		
		(0.00553)				(0.00311)		
pop_den_distx1000		0.0067841***				0.0008785		
		(0.00253)				(0.00168)		
D(not_slum)		0.0806693***				0.0087232		
		(0.02045)				(0.01253)		
D(mid_city)		0.0316513				-0.0266313**		
		(0.02377)				(0.01385)		
D(border)		0.0880407***				-0.0035699		
		(0.03050)				(0.01891)		
D(tongue)			-0.0427794*				-0.0431669***	
			(0.02367)				(0.01475)	
D(migrant)			0.0661511***				0.0151435	
			(0.02133)				(0.01257)	
D(social)			-0.0442737**				-0.0113628	
			(0.01996)				(0.01184)	
Station_distx1000				-0.0854781*				0.0154037
				(0.05113)				(0.05624)
Muni_persx1000				0.0080682***				0.0022449*
				(0.00270)				(0.00136)
L(budget_dist_pc)				0.028343*				-0.001824
				(0.01634)				(0.01040)
D(sdp_dist)				0.0505594				0.0138813
				(0.05347)				(0.02989)
Constant	7.25789***	5.03025***	7.241753***	7.117369***	0.3746718	0.7581394*	0.3967264	0.3776377
	(0.35739)	(0.67833)	(0.36066)	(0.36464)	(0.32228)	(0.43507)	(0.32649)	(0.32912)
Observations	2,160	2,160	2,160	2,160	2,160	2,160	2,160	2,160
R-squared	0.509	0.5339	0.5129	0.5139	0.0849	0.0891	0.0897	0.0859

j) Second-stage regression for IV models without labour supply (2SLS version)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.23606***	-0.06858	0.18782***	-0.04190	0.23761***	-0.06932	0.24084***	-0.06518
	(0.07738)	(0.05767)	(0.07235)	(0.05661)	(0.07875)	(0.05878)	(0.07719)	(0.05718)
L(expenditure)	-0.04523***	0.15363***	-0.06123***	0.15486***	-0.03541***	0.15416***	-0.04355***	0.15277***

	(0.01267)	(0.00897)	(0.01233)	(0.00921)	(0.01334)	(0.00948)	(0.01278)	(0.00894)
L(fam_size)	-0.01121	0.02106	-0.00325	0.01759	-0.02225	0.01940	-0.01202	0.01905
	(0.02837)	(0.02085)	(0.02662)	(0.02038)	(0.02877)	(0.02111)	(0.02846)	(0.02068)
D(mem_son)	-0.00519	-0.00283	-0.01152	0.00046	-0.00351	-0.00251	-0.00536	-0.00232
	(0.02082)	(0.01638)	(0.01975)	(0.01605)	(0.02090)	(0.01643)	(0.02093)	(0.01635)
D(mem_oth)	0.02291	-0.01894	0.02494	-0.02033	0.02604	-0.01857	0.02224	-0.01873
	(0.02235)	(0.01428)	(0.02191)	(0.01457)	(0.02227)	(0.01433)	(0.02243)	(0.01425)
P(men)	-0.02494	-0.03232*	-0.01667	-0.03435**	-0.02178	-0.03246*	-0.02332	-0.03352*
	(0.02407)	(0.01750)	(0.02288)	(0.01726)	(0.02401)	(0.01755)	(0.02409)	(0.01747)
D(gender_head)	-0.05583**	0.03309*	-0.05062**	0.02809	-0.05383**	0.03320*	-0.05690**	0.03036*
	(0.02607)	(0.01727)	(0.02417)	(0.01717)	(0.02655)	(0.01747)	(0.02604)	(0.01735)
Pcoh_1	0.01223	-0.07809	-0.07923	-0.06340	0.03258	-0.07617	0.00448	-0.07123
	(0.16649)	(0.11951)	(0.16334)	(0.11383)	(0.17090)	(0.11932)	(0.16669)	(0.11950)
Pcoh_2	0.10930	-0.16900	0.02591	-0.15184	0.13094	-0.16792	0.10579	-0.16491
	(0.16375)	(0.11642)	(0.16007)	(0.11043)	(0.16847)	(0.11625)	(0.16399)	(0.11634)
Pcoh_3	0.19223	-0.17020	0.10217	-0.14997	0.20078	-0.16987	0.18953	-0.16524
	(0.16555)	(0.11746)	(0.16203)	(0.11155)	(0.16989)	(0.11736)	(0.16583)	(0.11746)
Pcoh_4	0.14587	-0.13140	0.07163	-0.11726	0.15181	-0.12987	0.14004	-0.12505
	(0.16616)	(0.11876)	(0.16242)	(0.11302)	(0.17043)	(0.11859)	(0.16644)	(0.11866)
Pcoh_5	0.03284	-0.00800	-0.04227	0.00444	0.03661	-0.00791	0.02914	-0.00348
	(0.16238)	(0.11686)	(0.15903)	(0.11055)	(0.16653)	(0.11664)	(0.16238)	(0.11669)
Pcoh_6	0.04308	-0.00763	-0.04705	0.01005	0.04713	-0.00798	0.03525	-0.00323
	(0.16501)	(0.11854)	(0.16098)	(0.11256)	(0.16908)	(0.11835)	(0.16512)	(0.11839)
Pcoh_7	0.10176	-0.00414	0.01095	0.01067	0.10483	-0.00452	0.09644	0.00118
	(0.16721)	(0.11986)	(0.16232)	(0.11399)	(0.17122)	(0.11965)	(0.16747)	(0.11976)
Pcoh_8	0.09962	0.03997	0.01387	0.05232	0.10390	0.03979	0.09305	0.04544
	(0.16492)	(0.11860)	(0.16095)	(0.11260)	(0.16909)	(0.11844)	(0.16513)	(0.11851)
D(Lima)	0.11081***	-0.00587	0.13027***	-0.02032	0.10677***	-0.00500	0.10647***	-0.00707
	(0.01097)	(0.00830)	(0.01959)	(0.01448)	(0.01110)	(0.00841)	(0.01130)	(0.00847)
L(pop_urb_dist)			0.21538***	-0.09466**				
			(0.04576)	(0.04081)				
L(pop_urb_dist)-sq			-0.00891***	0.00409**				
			(0.00216)	(0.00191)				
pop_den_distx1000			-0.00302***	0.00086				
			(0.00108)	(0.00083)				
D(not_slum)			-0.00311	0.02356***				
			(0.00809)	(0.00666)				
D(mid_city)			0.00097	-0.01106				
			(0.00931)	(0.00734)				
D(border)			0.05920***	-0.00725				
			(0.01143)	(0.00965)				
D(tongue)					0.04502***	0.00443		
					(0.01046)	(0.00816)		
D(migrant)					0.00757	-0.00423		
					(0.00885)	(0.00677)		
D(social)					-0.00664	0.00277		
					(0.00819)	(0.00620)		
Station_distx1000							-0.07331***	0.01148
							(0.02372)	(0.01627)
Muni_persx1000							-0.00126	0.00154**
							(0.00111)	(0.00066)
L(budget_dist_pc)							0.00747	-0.00817
							(0.00705)	(0.00510)
D(sdp_dist)							-0.01573	-0.03484**

							(0.01991)	(0.01359)
Constant	0.62134***	-1.11582***	-0.39631	-0.61965**	0.51626**	-1.11828***	0.58142**	-1.07284***
	(0.23404)	(0.16509)	(0.31398)	(0.25898)	(0.24177)	(0.16818)	(0.23763)	(0.16509)
Observations	2,160	2,160	2,160	2,160	2,160	2,160	2,160	2,160
LogL	613.243	1212.338	728.290	1244.112	623.415	1211.342	611.654	1220.488

k) Second-stage regression for IV models without labour supply (LIML version)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.26042***	-0.07107	0.20890**	-0.04248	0.24991***	-0.07200	0.26640***	-0.06775
	(0.08667)	(0.06081)	(0.08163)	(0.05860)	(0.08345)	(0.06223)	(0.08680)	(0.06064)
L(expenditure)	-0.04291***	0.15355***	-0.05919***	0.15493***	-0.03413**	0.15407***	-0.04110***	0.15271***
	(0.01345)	(0.00918)	(0.01308)	(0.00935)	(0.01376)	(0.00974)	(0.01359)	(0.00918)
L(fam_size)	-0.01137	0.02105	-0.00341	0.01757	-0.02253	0.01942	-0.01229	0.01905
	(0.02877)	(0.02087)	(0.02692)	(0.02039)	(0.02899)	(0.02115)	(0.02890)	(0.02070)
D(mem_son)	-0.00515	-0.00290	-0.01149	0.00041	-0.00346	-0.00259	-0.00526	-0.00240
	(0.02114)	(0.01640)	(0.01998)	(0.01606)	(0.02106)	(0.01644)	(0.02127)	(0.01636)
D(mem_oth)	0.02343	-0.01899	0.02544	-0.02033	0.02635	-0.01862	0.02278	-0.01878
	(0.02259)	(0.01429)	(0.02206)	(0.01458)	(0.02240)	(0.01435)	(0.02269)	(0.01426)
P(men)	-0.02525	-0.03230*	-0.01697	-0.03435**	-0.02195	-0.03243*	-0.02367	-0.03350*
	(0.02439)	(0.01751)	(0.02311)	(0.01726)	(0.02417)	(0.01756)	(0.02442)	(0.01747)
D(gender_head)	-0.05874**	0.03341*	-0.05313**	0.02817	-0.05529**	0.03355*	-0.05998**	0.03069*
	(0.02675)	(0.01743)	(0.02474)	(0.01726)	(0.02688)	(0.01765)	(0.02675)	(0.01753)
Pcoh_1	0.00729	-0.07735	-0.08308	-0.06307	0.03029	-0.07539	-0.00041	-0.07047
	(0.16967)	(0.11990)	(0.16512)	(0.11391)	(0.17253)	(0.11974)	(0.17009)	(0.11990)
Pcoh_2	0.10833	-0.16868	0.02539	-0.15162	0.13059	-0.16758	0.10501	-0.16457
	(0.16686)	(0.11676)	(0.16172)	(0.11050)	(0.17008)	(0.11663)	(0.16734)	(0.11669)
Pcoh_3	0.19135	-0.16987	0.10173	-0.14974	0.20036	-0.16950	0.18887	-0.16490
	(0.16859)	(0.11780)	(0.16366)	(0.11161)	(0.17146)	(0.11773)	(0.16910)	(0.11781)
Pcoh_4	0.14383	-0.13101	0.07023	-0.11706	0.15090	-0.12946	0.13807	-0.12464
	(0.16922)	(0.11909)	(0.16405)	(0.11308)	(0.17202)	(0.11896)	(0.16972)	(0.11900)
Pcoh_5	0.03025	-0.00758	-0.04418	0.00465	0.03532	-0.00746	0.02661	-0.00305
	(0.16545)	(0.11722)	(0.16065)	(0.11062)	(0.16811)	(0.11704)	(0.16567)	(0.11706)
Pcoh_6	0.04216	-0.00739	-0.04759	0.01022	0.04662	-0.00771	0.03441	-0.00297
	(0.16804)	(0.11888)	(0.16257)	(0.11262)	(0.17064)	(0.11872)	(0.16838)	(0.11875)
Pcoh_7	0.10301	-0.00419	0.01241	0.01074	0.10542	-0.00456	0.09799	0.00112
	(0.17038)	(0.12021)	(0.16403)	(0.11406)	(0.17285)	(0.12002)	(0.17089)	(0.12011)
Pcoh_8	0.10080	0.03997	0.01526	0.05242	0.10447	0.03980	0.09452	0.04544
	(0.16799)	(0.11894)	(0.16256)	(0.11267)	(0.17068)	(0.11881)	(0.16846)	(0.11887)
D(Lima)	0.11048***	-0.00588	0.13070***	-0.02036	0.10667***	-0.00502	0.10604***	-0.00708
	(0.01115)	(0.00831)	(0.01985)	(0.01449)	(0.01118)	(0.00841)	(0.01151)	(0.00848)
L(pop_urb_dist)			0.21536***	-0.09477**				
			(0.04613)	(0.04081)				
L(pop_urb_dist)-sq			-0.00891***	0.00409**				
			(0.00218)	(0.00191)				
pop_den_distx1000			-0.00305***	0.00086				
			(0.00110)	(0.00083)				
D(not_slum)			-0.00332	0.02353***				
			(0.00819)	(0.00666)				
D(mid_city)			0.00160	-0.01109				
			(0.00947)	(0.00736)				
D(border)			0.05915***	-0.00726				
			(0.01156)	(0.00965)				

D(tongue)					0.04561***	0.00435		
					(0.01062)	(0.00822)		
D(migrant)					0.00728	-0.00418		
					(0.00893)	(0.00679)		
D(social)					-0.00635	0.00270		
					(0.00827)	(0.00622)		
Station_distx1000							-0.07358***	0.01150
							(0.02427)	(0.01629)
Muni_persx1000							-0.00133	0.00154**
							(0.00111)	(0.00066)
L(budget_dist_pc)							0.00756	-0.00819
							(0.00714)	(0.00510)
D(sdp_dist)							-0.01649	-0.03476**
							(0.02022)	(0.01361)
Constant	0.58481**	-1.11368***	-0.42845	-0.61944**	0.49692**	-1.11585***	0.54262**	-1.07070***
	(0.24494)	(0.16798)	(0.32084)	(0.26018)	(0.24748)	(0.17163)	(0.24932)	(0.16825)
Observations	2,160	2,160	2,160	2,160	2,160	2,160	2,160	2,160
LogL	584.279	1211.043	706.645	1243.777	608.869	1209.948	580.740	1219.187

l) Second-stage regression for IV models without labour supply (GMM version)

	(1)		(2)		(3)		(4)	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
θ	0.25454***	-0.06999	0.20587***	-0.04363	0.25472***	-0.07094	0.26086***	-0.06806
	(0.07656)	(0.05755)	(0.07178)	(0.05648)	(0.07789)	(0.05866)	(0.07641)	(0.05702)
L(expenditure)	-0.04392***	0.15470***	-0.05934***	0.15572***	-0.03362**	0.15546***	-0.04215***	0.15364***
	(0.01262)	(0.00894)	(0.01230)	(0.00919)	(0.01328)	(0.00944)	(0.01273)	(0.00891)
L(fam_size)	-0.01032	0.02001	-0.00313	0.01681	-0.02257	0.01813	-0.01083	0.01792
	(0.02834)	(0.02079)	(0.02660)	(0.02034)	(0.02875)	(0.02107)	(0.02844)	(0.02062)
D(mem_son)	-0.00560	-0.00196	-0.01138	0.00099	-0.00352	-0.00161	-0.00574	-0.00173
	(0.02081)	(0.01633)	(0.01974)	(0.01600)	(0.02089)	(0.01638)	(0.02091)	(0.01629)
D(mem_oth)	0.02297	-0.01897	0.02620	-0.02012	0.02653	-0.01849	0.02234	-0.01887
	(0.02234)	(0.01427)	(0.02190)	(0.01455)	(0.02227)	(0.01432)	(0.02241)	(0.01424)
P(men)	-0.02528	-0.03277*	-0.01790	-0.03461**	-0.02234	-0.03264*	-0.02376	-0.03397*
	(0.02406)	(0.01748)	(0.02284)	(0.01725)	(0.02401)	(0.01753)	(0.02407)	(0.01744)
D(gender_head)	-0.05563**	0.03232*	-0.05110**	0.02760	-0.05385**	0.03218*	-0.05680**	0.02959*
	(0.02600)	(0.01725)	(0.02410)	(0.01716)	(0.02649)	(0.01744)	(0.02597)	(0.01733)
Pcoh_1	-0.01041	-0.08412	-0.08397	-0.06752	0.02001	-0.08084	-0.01893	-0.07606
	(0.16619)	(0.11931)	(0.16305)	(0.11352)	(0.17074)	(0.11915)	(0.16637)	(0.11932)
Pcoh_2	0.09033	-0.17100	0.02277	-0.15172	0.12098	-0.16881	0.08606	-0.16548
	(0.16349)	(0.11630)	(0.15977)	(0.11016)	(0.16834)	(0.11614)	(0.16370)	(0.11623)
Pcoh_3	0.17027	-0.17312	0.09778	-0.15070	0.18905	-0.17184	0.16697	-0.16703
	(0.16525)	(0.11734)	(0.16170)	(0.11127)	(0.16972)	(0.11726)	(0.16551)	(0.11735)
Pcoh_4	0.12815	-0.13220	0.06812	-0.11651	0.14163	-0.13011	0.12153	-0.12404
	(0.16592)	(0.11864)	(0.16217)	(0.11276)	(0.17030)	(0.11847)	(0.16617)	(0.11856)
Pcoh_5	0.01390	-0.01216	-0.04399	0.00278	0.02595	-0.01121	0.00989	-0.00679
	(0.16216)	(0.11677)	(0.15878)	(0.11033)	(0.16641)	(0.11658)	(0.16214)	(0.11661)
Pcoh_6	0.02791	-0.00987	-0.04611	0.01030	0.03889	-0.00940	0.02026	-0.00460
	(0.16485)	(0.11842)	(0.16074)	(0.11231)	(0.16900)	(0.11826)	(0.16495)	(0.11829)
Pcoh_7	0.08832	-0.00870	0.01380	0.00830	0.09823	-0.00794	0.08338	-0.00289
	(0.16706)	(0.11977)	(0.16209)	(0.11384)	(0.17115)	(0.11960)	(0.16731)	(0.11967)
Pcoh_8	0.08474	0.03685	0.01565	0.05133	0.09658	0.03770	0.07836	0.04312
	(0.16474)	(0.11847)	(0.16069)	(0.11235)	(0.16902)	(0.11835)	(0.16495)	(0.11840)
D(Lima)	0.11039***	-0.00676	0.13097***	-0.02164	0.10646***	-0.00604	0.10612***	-0.00797

	(0.01095)	(0.00829)	(0.01954)	(0.01444)	(0.01107)	(0.00839)	(0.01127)	(0.00846)
L(pop_urb_dist)			0.21570***	-0.10051**				
			(0.04572)	(0.04056)				
L(pop_urb_dist)-sq			-0.00891***	0.00435**				
			(0.00216)	(0.00190)				
pop_den_distx1000			-0.00320***	0.00088				
			(0.00108)	(0.00083)				
D(not_slum)			-0.00378	0.02358***				
			(0.00809)	(0.00664)				
D(mid_city)			0.00111	-0.01116				
			(0.00928)	(0.00733)				
D(border)			0.05899***	-0.00665				
			(0.01143)	(0.00960)				
D(tongue)					0.04584***	0.00506		
					(0.01045)	(0.00814)		
D(migrant)					0.00655	-0.00394		
					(0.00883)	(0.00676)		
D(social)					-0.00640	0.00275		
					(0.00818)	(0.00619)		
Station_distx1000							-0.07182***	0.01230
							(0.02369)	(0.01625)
Muni_persx1000							-0.00130	0.00156**
							(0.00111)	(0.00066)
L(budget_dist_pc)							0.00800	-0.00853*
							(0.00703)	(0.00507)
D(sdp_dist)							-0.01671	-0.03444**
							(0.01989)	(0.01358)
Constant	0.61149***	-1.12015***	-0.42943	-0.59267**	0.49665**	-1.12558***	0.56670**	-1.07306***
	(0.23342)	(0.16498)	(0.31346)	(0.25826)	(0.24115)	(0.16806)	(0.23696)	(0.16497)
Observations	2,160	2,160	2,160	2,160	2,160	2,160	2,160	2,160
LogL	591.452	1209.391	709.843	1241.785	603.014	1207.808	587.616	1217.148

m) Reduced form models

	Original		Reduced Form	
	Informal	Formal	Informal	Formal
θ	-0.25790	-0.05320		
	(0.19113)	(0.16360)		
$\theta \ln(\theta M)$	-0.03052***	0.07280***		
	(0.00928)	(0.00821)		
$(1-\theta)\ln[(1-\theta)M]$	-0.06037***	0.06799***		
	(0.01596)	(0.01389)		
θ_educ			0.11724	0.09298
			(0.19766)	(0.14304)
$\theta\_educ \ln(\theta * asset)$			-0.01505	0.03006***
			(0.01276)	(0.00958)
$(1-\theta\_educ)\ln[(1-\theta\_educ)*asset]$			-0.01318	0.04715***
			(0.01388)	(0.01048)
L(fam_size)	-0.03547	0.02197	-0.03796	0.02471
	(0.03208)	(0.02377)	(0.03164)	(0.02464)
D(mem_son)	-0.00906	0.00388	-0.02084	0.02355
	(0.01880)	(0.01545)	(0.01915)	(0.01629)
D(mem_oth)	0.01445	-0.01512	0.01890	-0.02587*

	(0.02143)	(0.01328)	(0.02100)	(0.01382)
P(men)	-0.01703	-0.03814**	-0.02727	-0.01691
	(0.02243)	(0.01671)	(0.02269)	(0.01793)
D(gender_head)	-0.02452	0.01044	-0.01625	0.00208
	(0.02379)	(0.01522)	(0.02358)	(0.01625)
age_male	-0.00039	0.00125*	-0.00016	0.00098
	(0.00087)	(0.00067)	(0.00088)	(0.00070)
age_female	0.00084	0.00121	0.00070	0.00112
	(0.00100)	(0.00081)	(0.00100)	(0.00083)
year_educ_male	-0.00095	0.00590***		
	(0.00139)	(0.00098)		
year_educ_female	-0.00391***	0.00517***		
	(0.00121)	(0.00090)		
D(chronic_male)	0.00444	0.00375	0.00602	0.00250
	(0.00917)	(0.00698)	(0.00930)	(0.00745)
D(chronic_female)	-0.00163	0.01281*	-0.00190	0.01216*
	(0.00858)	(0.00676)	(0.00862)	(0.00716)
Hrs_male	0.00029	-0.00014	0.00025	-0.00004
	(0.00022)	(0.00017)	(0.00022)	(0.00018)
Hrs_female	0.00009	0.00070***	0.00020	0.00030*
	(0.00020)	(0.00016)	(0.00020)	(0.00016)
D(Hrs_male)	-0.07175***	0.02816	-0.05568***	0.00642
	(0.02139)	(0.01928)	(0.02052)	(0.01805)
D(Hrs_female)	-0.01695	0.00284	-0.01406	0.00166
	(0.01240)	(0.00917)	(0.01201)	(0.00931)
D(Lima)	0.10300***	0.01958**	0.09619***	0.03333***
	(0.01017)	(0.00808)	(0.01055)	(0.00862)
Pcoh_1	0.04035	-0.05426	0.15746	-0.29088**
	(0.15642)	(0.11737)	(0.15241)	(0.11935)
Pcoh_2	0.09319	-0.12998	0.20939	-0.36191***
	(0.15365)	(0.11426)	(0.14971)	(0.11629)
Pcoh_3	0.16517	-0.12882	0.29164*	-0.38364***
	(0.15537)	(0.11484)	(0.15121)	(0.11719)
Pcoh_4	0.11420	-0.06951	0.23500	-0.30847***
	(0.15581)	(0.11597)	(0.15184)	(0.11876)
Pcoh_5	0.00298	0.04044	0.10504	-0.16662
	(0.15256)	(0.11378)	(0.14883)	(0.11641)
Pcoh_6	-0.01460	0.02466	0.09254	-0.19198
	(0.15563)	(0.11568)	(0.15189)	(0.11853)
Pcoh_7	0.02017	0.01523	0.11095	-0.17851
	(0.15885)	(0.11874)	(0.15546)	(0.12169)
Pcoh_8	-0.00247	0.02921	0.10208	-0.18318
	(0.16237)	(0.12135)	(0.15899)	(0.12418)
Constant	1.06608***	-0.61731***	0.38962**	0.05411
	(0.22498)	(0.17985)	(0.19813)	(0.13933)
Observations	2,160		2,160	
LogL	2295.909		2173.723	

n) Consumption groups (PROBIT)

	Foff	HEA
L(expenditure)	0.74934***	0.58865***
	(0.07951)	(0.07813)
L(fam_size)	0.28597	-0.09082

	(0.31127)	(0.30310)
D(mem_son)	-0.21954	0.03068
	(0.17948)	(0.17633)
D(mem_oth)	-0.14787	-0.17205
	(0.18512)	(0.19121)
P(men)	0.11969	-0.53890**
	(0.21971)	(0.22451)
D(gender_head)	0.19623	-0.34899
	(0.22146)	(0.23280)
age_average	-0.03019***	-0.00593
	(0.01113)	(0.01146)
year_educ_average	-0.01060	0.00601
	(0.01226)	(0.01235)
D(chronic_any)	0.03770	0.12056
	(0.07654)	(0.07936)
Hrs_both	0.00334***	0.00027
	(0.00111)	(0.00102)
D(Hrs_both)	0.36389	-0.48697
	(0.27991)	(0.34364)
D(Lima)	-0.08794	0.46922***
	(0.10701)	(0.10940)
Pcoh_1	0.88165	-1.22414
	(1.67838)	(1.64303)
Pcoh_2	0.76471	-1.35659
	(1.65281)	(1.61689)
Pcoh_3	0.44343	-1.07198
	(1.64651)	(1.62752)
Pcoh_4	1.87377	-1.90700
	(1.68402)	(1.62440)
Pcoh_5	1.78563	-1.21527
	(1.65598)	(1.64083)
Pcoh_6	1.58656	-0.87628
	(1.66760)	(1.67239)
Pcoh_7	1.92713	-1.44916
	(1.69055)	(1.68837)
Pcoh_8	2.05720	-0.71175
	(1.69160)	(1.72557)
Hotel_firmPC	0.09845***	
	(0.03425)	
D(rest_dist)	0.14875*	
	(0.08841)	
D(tour_dist)	-0.05534	
	(0.09485)	
Prop(workrest_dist)	1.68972	
	(3.24268)	
D(ca_health)_1		-0.19607**
		(0.09872)
D(ca_health)_2		0.15379
		(0.09743)
n_eventual		0.43309***
		(0.03632)
Health1_distx1000		-0.53325
		(0.54362)
Health2_distx1000		0.14703

		(0.93444)
Health3_distx1000		0.58769
		(0.59596)
Health4_distx1000		0.53803
		(0.55642)
Health_firmPC		-0.18976*
		(0.11419)
Prop(workhealth_dist)		-4.74650
		(3.34460)
Constant	-7.59479***	-2.71804
	(1.86249)	(1.88485)
Observations	2,105	2,105
LogL	-742.355	-793.976
Wald (ident. vars.)	17.86***	146.80***

o) Consumption groups (SUR)

	Fon	Foff	CC	HEA	TC	ED
θ	0.34594**	-0.03839	-0.08762	-0.03271	-0.17543**	0.00913
	(0.16393)	(0.13432)	(0.07560)	(0.11351)	(0.08073)	(0.07233)
$\theta\ln(\theta M)$	-0.06746***	-0.03438***	0.00788**	0.00232	0.03183***	0.01645***
	(0.00773)	(0.00788)	(0.00328)	(0.00492)	(0.00461)	(0.00351)
$(1-\theta)\ln[(1-\theta)M]$	-0.02930**	-0.04369***	-0.00127	0.00001	0.01221*	0.01945***
	(0.01388)	(0.01143)	(0.00663)	(0.01015)	(0.00738)	(0.00699)
L(fam_size)	0.00882	-0.04480*	-0.00102	0.02793*	0.00088	0.04881***
	(0.02368)	(0.02390)	(0.01122)	(0.01538)	(0.01370)	(0.01276)
D(mem_son)	0.02423	-0.00524	-0.01036	-0.00402	0.00600	0.02321***
	(0.01534)	(0.01603)	(0.00714)	(0.01204)	(0.00954)	(0.00753)
D(mem_oth)	0.01083	-0.00657	-0.01365***	-0.00262	0.01597**	-0.01249
	(0.01373)	(0.01281)	(0.00518)	(0.00852)	(0.00800)	(0.00773)
P(men)	0.02249	0.00427	0.00705	0.00181	-0.00987	-0.01313
	(0.01679)	(0.01394)	(0.00798)	(0.01011)	(0.01020)	(0.00916)
D(gender_head)	0.01814	-0.04933**	0.01098*	-0.00459	0.00542	0.01500*
	(0.01777)	(0.02037)	(0.00577)	(0.01423)	(0.00852)	(0.00812)
age_male	-0.00062	0.00055	-0.00028	-0.00010	0.00039	0.00039
	(0.00069)	(0.00057)	(0.00029)	(0.00046)	(0.00040)	(0.00031)
age_female	-0.00013	0.00069	-0.00041	-0.00074*	0.00067	0.00148***
	(0.00074)	(0.00069)	(0.00033)	(0.00043)	(0.00047)	(0.00037)
year_educ_male	-0.00425***	0.00001	-0.00016	0.00061	0.00184***	0.00216***
	(0.00104)	(0.00085)	(0.00044)	(0.00066)	(0.00061)	(0.00049)
year_educ_female	-0.00544***	-0.00126*	0.00111***	-0.00065	0.00212***	0.00297***
	(0.00087)	(0.00073)	(0.00037)	(0.00059)	(0.00055)	(0.00043)
D(chronic_male)	-0.00653	-0.00935*	-0.00194	0.01376***	-0.00037	-0.00395
	(0.00682)	(0.00530)	(0.00318)	(0.00513)	(0.00422)	(0.00356)
D(chronic_female)	-0.02126***	-0.00774	-0.00011	0.01445***	0.00268	0.00399
	(0.00658)	(0.00515)	(0.00314)	(0.00475)	(0.00401)	(0.00345)
Hrs_male	0.00004	0.00022	0.00015**	-0.00004	-0.00023*	-0.00018**
	(0.00017)	(0.00014)	(0.00008)	(0.00011)	(0.00012)	(0.00009)
Hrs_female	-0.00171***	0.00029**	0.00021***	0.00028**	0.00049***	-0.00002
	(0.00017)	(0.00014)	(0.00007)	(0.00011)	(0.00010)	(0.00008)
D(Hrs_male)	-0.02747	-0.02873*	-0.00000	-0.00235	0.01887*	0.01895**
	(0.01762)	(0.01674)	(0.00741)	(0.01698)	(0.01019)	(0.00896)
D(Hrs_female)	0.00531	0.00273	-0.00090	-0.00958	-0.00552	0.00952**
	(0.00968)	(0.00792)	(0.00425)	(0.00613)	(0.00552)	(0.00472)

D(Lima)	0.01928**	0.00396	-0.03001***	0.00161	0.01381***	-0.00350
	(0.00758)	(0.00540)	(0.00268)	(0.00517)	(0.00482)	(0.00414)
Pcoh_1	0.05115	-0.00409	0.05440	-0.11506	0.12436*	-0.14221*
	(0.11810)	(0.09682)	(0.04191)	(0.08434)	(0.06511)	(0.08145)
Pcoh_2	0.05635	0.01084	0.07042*	-0.14209*	0.09630	-0.09555
	(0.11532)	(0.09451)	(0.04007)	(0.08366)	(0.06285)	(0.08064)
Pcoh_3	0.06017	0.02707	0.06405	-0.15475*	0.12700**	-0.11369
	(0.11610)	(0.09628)	(0.04023)	(0.08319)	(0.06366)	(0.08116)
Pcoh_4	-0.02681	0.03620	0.05752	-0.19289**	0.17374***	-0.05888
	(0.11729)	(0.09665)	(0.04022)	(0.08454)	(0.06372)	(0.08192)
Pcoh_5	-0.13440	0.09472	0.05689	-0.14214*	0.18371***	-0.01300
	(0.11444)	(0.09700)	(0.04029)	(0.08295)	(0.06413)	(0.07974)
Pcoh_6	-0.09802	0.05400	0.04713	-0.08344	0.19450***	-0.06402
	(0.11657)	(0.10073)	(0.04298)	(0.08572)	(0.06574)	(0.08132)
Pcoh_7	-0.07994	-0.00459	0.05834	-0.06608	0.18618***	-0.07640
	(0.11915)	(0.10407)	(0.04443)	(0.08666)	(0.06674)	(0.08151)
Pcoh_8	-0.04342	-0.03646	0.04242	-0.04439	0.18684***	-0.09910
	(0.12160)	(0.10668)	(0.04584)	(0.08928)	(0.06787)	(0.08170)
PHI		-0.08764***		-0.02098**		
		(0.02438)		(0.00890)		
Constant	0.84793***	0.56201***	0.05721	0.16930	-0.26488***	-0.24451**
	(0.18015)	(0.16095)	(0.07576)	(0.13208)	(0.09990)	(0.10207)
Observations	2,105					
LogL	15183.850					

p) Markets within groups (PROBIT)

	Fon	Foff		CC		HEA		TC		ED
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
L(expend_Fon)	0.47319***									
	(0.06266)									
L(expend_Foff)		0.02135	0.15324***							
		(0.02753)	(0.03036)							
L(expend_CC)				0.37662***	0.37004***					
				(0.04499)	(0.04018)					
L(expend_HEA)						0.17310***	0.54941***			
						(0.02700)	(0.04000)			
L(expend_TC)								0.33172***	0.74107***	
								(0.04503)	(0.05382)	
L(expend_ED)										0.12263***
										(0.03469)
L(fam_size)	-1.11737***	0.51921**	0.18821	0.00790	-0.48001*	-0.57811*	0.23641	0.23935	0.08706	0.60539**
	(0.25474)	(0.26121)	(0.30732)	(0.29969)	(0.27153)	(0.31825)	(0.38583)	(0.34977)	(0.33599)	(0.27399)
D(mem_son)	0.03697	-0.16172	0.27030	-0.03679	-0.05952	-0.13416	0.06763	-0.17150	0.27465	-0.73172***
	(0.15875)	(0.16216)	(0.18101)	(0.17824)	(0.17050)	(0.19240)	(0.24105)	(0.20529)	(0.20156)	(0.19427)
D(mem_oth)	0.04820	-0.02510	0.00752	0.08113	0.01731	0.30721*	-0.11471	-0.17065	-0.00501	-0.16827
	(0.15242)	(0.16251)	(0.18661)	(0.19089)	(0.16572)	(0.17865)	(0.22769)	(0.21151)	(0.20519)	(0.15833)
P(men)	-0.07918	0.05268	0.11416	-0.43730*	0.01932	-0.10232	-0.73980**	-0.00360	-0.50954**	-0.16718
	(0.18482)	(0.18729)	(0.21115)	(0.23716)	(0.18986)	(0.22451)	(0.29396)	(0.23867)	(0.25190)	(0.19748)
D(gender_head)	0.21652	-0.08631	-0.21933	0.13650	-0.03547	0.09617	0.35501	0.16160	0.25050	0.04158
	(0.17451)	(0.19539)	(0.22296)	(0.21517)	(0.18273)	(0.21107)	(0.23991)	(0.22431)	(0.20388)	(0.18769)
age_average	0.01941**	-0.01787*	-0.00088	-0.00874	0.01005	0.00530	-0.01121	-0.02071	0.02211*	-0.00083
	(0.00980)	(0.01012)	(0.01150)	(0.01151)	(0.01062)	(0.01152)	(0.01439)	(0.01363)	(0.01307)	(0.01085)
year_educ_average	0.03997***	-0.03746***	0.08608***	-0.05388***	0.06139***	0.00045	0.00932	-0.02841*	0.05853***	-0.00817
	(0.00968)	(0.00997)	(0.01104)	(0.01241)	(0.01075)	(0.01204)	(0.01703)	(0.01467)	(0.01391)	(0.01166)

D(chronic_any)	0.01726	0.13696**	-0.01112	-0.00516	-0.02361	-0.06658	0.08842	0.07803	-0.11486	0.14601**
	(0.06391)	(0.06562)	(0.07258)	(0.07938)	(0.06638)	(0.08012)	(0.10324)	(0.08912)	(0.08488)	(0.06904)
Hrs_both	0.00125	0.00252***	-0.00010	-0.00001	-0.00109	-0.00032	0.00070	-0.00127	-0.00055	0.00127
	(0.00087)	(0.00089)	(0.00097)	(0.00107)	(0.00089)	(0.00103)	(0.00152)	(0.00112)	(0.00115)	(0.00092)
D(Hrs_both)	0.02785	0.32097	-0.58826*	-0.06189	0.10429	-0.26353	0.19347	0.21185	-0.26735	-0.04186
	(0.24407)	(0.29900)	(0.35403)	(0.25072)	(0.25997)	(0.26213)	(0.29326)	(0.36708)	(0.35128)	(0.29167)
D(Lima)	0.48586***	0.03122	0.15951	-0.01447	0.54732***	-0.22969	-0.13985	-0.38273*	0.13940	-0.02569
	(0.12314)	(0.12969)	(0.14159)	(0.15237)	(0.13179)	(0.15067)	(0.19553)	(0.22001)	(0.20919)	(0.13538)
Pcoh_1	0.16767	0.60024	-1.77313	2.24661	-0.89272	-0.30231	-1.32111	-3.34266*	1.30628	1.34337
	(1.38771)	(1.43388)	(1.72805)	(1.41690)	(1.41280)	(1.48304)	(1.88890)	(1.90642)	(1.64684)	(1.31723)
Pcoh_2	-0.39275	0.67841	-2.02699	2.38722*	-0.98672	-0.53109	-1.36425	-3.02145	0.78868	2.05090
	(1.36733)	(1.41204)	(1.70759)	(1.37922)	(1.39433)	(1.45025)	(1.82795)	(1.90422)	(1.61034)	(1.29128)
Pcoh_3	-0.05995	0.94618	-1.87491	3.11065**	-1.64454	-0.21626	-0.79774	-3.51309*	1.38612	2.26681*
	(1.37085)	(1.41766)	(1.71230)	(1.40024)	(1.39992)	(1.45104)	(1.84244)	(1.88387)	(1.61560)	(1.28966)
Pcoh_4	-0.76857	1.23500	-0.63003	2.39225*	-1.59635	-0.11378	-1.83594	-1.93801	-0.20920	1.53509
	(1.37531)	(1.42235)	(1.72077)	(1.41482)	(1.40396)	(1.46162)	(1.87005)	(1.92968)	(1.61174)	(1.30237)
Pcoh_5	-1.22710	1.25657	-0.51952	2.41923*	-2.11350	-0.35744	-0.49243	-2.80053	1.04687	2.08805
	(1.35950)	(1.40575)	(1.71259)	(1.36291)	(1.39281)	(1.43996)	(1.83343)	(1.90855)	(1.60456)	(1.27249)
Pcoh_6	-1.68198	1.38933	-0.84915	1.71808	-2.07177	-1.10846	-0.75785	-2.68581	1.16008	2.33333*
	(1.38460)	(1.41890)	(1.73529)	(1.38353)	(1.42205)	(1.45464)	(1.86126)	(1.96030)	(1.64422)	(1.30126)
Pcoh_7	-1.64892	1.70981	-1.17795	1.52896	-1.78209	-0.83787	-0.93241	-1.93817	0.90419	2.68145**
	(1.40521)	(1.43267)	(1.74991)	(1.40321)	(1.44245)	(1.48091)	(1.91274)	(1.98334)	(1.67639)	(1.32053)
Pcoh_8	-1.89515	1.56717	-0.86889	1.60537	-2.08377	-1.05194	-0.87322	-2.25026	0.76566	2.28057*
	(1.43716)	(1.44868)	(1.77035)	(1.44083)	(1.47501)	(1.50611)	(1.94234)	(2.03138)	(1.72706)	(1.34938)
P(pop_slums_dist)	0.08410	0.02918	0.03584	0.17458	-0.09508	-0.03135	0.06199	-0.17662	0.02247	-0.07162
	(0.10644)	(0.11043)	(0.12148)	(0.12710)	(0.10915)	(0.12864)	(0.16608)	(0.15450)	(0.15044)	(0.11064)
P(migrant_dist)	-0.20707	-0.46231	1.34815**	-1.01823*	0.88629*	0.64604	0.95039	-0.78081	1.12935*	-1.17080**
	(0.47249)	(0.49838)	(0.53267)	(0.53072)	(0.48587)	(0.56514)	(0.71354)	(0.58438)	(0.62663)	(0.48856)
P(race_dist)	-0.09123	2.11913***	-1.00065*	0.70335	-0.19086	0.65493	-1.32648**	1.21977**	0.06618	1.10456**
	(0.47648)	(0.50092)	(0.53406)	(0.51518)	(0.47074)	(0.53297)	(0.59483)	(0.51865)	(0.64120)	(0.45301)

informal_index	-0.02772	0.09049	-0.41554***	-0.23020	-0.01284	-0.00789	-0.05188	0.30318*	-0.04395	-0.25994**
	(0.12881)	(0.14076)	(0.15656)	(0.14940)	(0.13584)	(0.15646)	(0.20832)	(0.17894)	(0.20301)	(0.12773)
D(conf_gov_1)	0.02340	-0.05379	0.09349	-0.08713	-0.00407	0.08473	0.26058**	0.06797	0.02208	0.03001
	(0.07586)	(0.07814)	(0.08933)	(0.09405)	(0.07655)	(0.09362)	(0.12690)	(0.10204)	(0.10195)	(0.08333)
D(conf_gov_2)	-0.24880**	0.16629	-0.07407	0.10866	0.03174	0.19204	0.06562	0.01379	-0.30161**	-0.04027
	(0.10935)	(0.10926)	(0.11549)	(0.12474)	(0.10823)	(0.12836)	(0.16225)	(0.14335)	(0.13113)	(0.10743)
D(conf_gov_3)	0.11826	-0.05069	-0.14425	-0.24818	-0.14062	0.01509	-0.19591	-0.27564	0.26315	0.01354
	(0.16227)	(0.16563)	(0.17626)	(0.18167)	(0.16270)	(0.19118)	(0.24515)	(0.20782)	(0.19574)	(0.16144)
D(overall_satis)	-0.22963***	0.12725*	-0.02810	0.02721	-0.07680	0.00110	0.21108*	0.01302	-0.18963*	-0.13455*
	(0.06907)	(0.07032)	(0.07868)	(0.08669)	(0.07098)	(0.08823)	(0.12096)	(0.09120)	(0.10021)	(0.07574)
D(income_satis)	-0.05951	0.05147	-0.15627**	0.03216	-0.00355	0.03770	0.04817	0.24443**	-0.10311	-0.11868
	(0.07192)	(0.07309)	(0.07854)	(0.08716)	(0.07426)	(0.08513)	(0.11011)	(0.09659)	(0.09083)	(0.07480)
D(income_stab)	-0.12401	0.25192***	-0.23432**	0.19881**	0.02913	0.09851	0.03682	-0.30754***	0.18699	-0.06523
	(0.07909)	(0.08085)	(0.09549)	(0.09633)	(0.08150)	(0.09736)	(0.13947)	(0.11782)	(0.12155)	(0.09023)
D(sub_pov)	0.05556	0.03838	-0.07789	0.08667	-0.11918*	0.09067	0.06636	-0.13355	-0.06889	-0.01311
	(0.06499)	(0.06686)	(0.07234)	(0.07954)	(0.06679)	(0.07845)	(0.10281)	(0.08532)	(0.08643)	(0.06945)
D(road_1)	-0.07355	-0.17425	0.08078	0.03389	-0.09750	-0.06824	-0.00676	-0.31523**	0.23018	0.21890*
	(0.11756)	(0.12205)	(0.13809)	(0.12626)	(0.12679)	(0.14465)	(0.18054)	(0.14917)	(0.16722)	(0.11949)
D(road_2)	0.09109	0.00125	-0.11004	-0.20206*	0.02486	-0.07653	0.05643	0.20685	0.03282	0.09936
	(0.10854)	(0.11337)	(0.12620)	(0.12014)	(0.11495)	(0.13550)	(0.17196)	(0.14467)	(0.15701)	(0.11661)
D(road_3)	0.04098	-0.22231	-0.08743	-0.36686	-0.23659	-0.10003	0.44215	-0.17472	-0.20844	-0.23896
	(0.25765)	(0.26638)	(0.29746)	(0.25860)	(0.33877)	(0.33312)	(0.33481)	(0.30575)	(0.31917)	(0.23173)
distance	0.00027	0.00016	-0.00044	-0.00041	-0.00077**	-0.00070*	-0.00036	0.00022	-0.00019	-0.00006
	(0.00028)	(0.00028)	(0.00030)	(0.00030)	(0.00032)	(0.00039)	(0.00039)	(0.00047)	(0.00047)	(0.00027)
P(nowater_dist)	0.10280	0.38563	-0.17886	0.47859*	-0.53036*	-0.12515	-0.68851*	-0.43088	0.13172	0.04072
	(0.28877)	(0.27897)	(0.31362)	(0.27389)	(0.29677)	(0.34305)	(0.37479)	(0.32390)	(0.34215)	(0.28134)
P(nodrain_dist)	-0.39068	-0.49088	0.25944	-0.40196	0.36200	0.40817	0.18976	0.66611	-0.66535	-0.05987
	(0.37893)	(0.34661)	(0.38586)	(0.36729)	(0.36481)	(0.43044)	(0.48417)	(0.41178)	(0.42386)	(0.35257)
P(noelec_dist)	-1.70534**	0.50829	0.79651	1.72161*	0.39822	-0.71585	1.11079	-0.26177	-0.01256	0.52542
	(0.81608)	(0.76969)	(0.82477)	(0.88485)	(0.85102)	(0.96444)	(1.04196)	(0.85346)	(0.87831)	(0.75707)

P(nofloor_dist)	0.33004	0.04602	0.03699	1.68533***	-0.63907	0.08024	-0.05855	0.14105	0.56782	0.30034
	(0.43181)	(0.45962)	(0.49707)	(0.45915)	(0.45487)	(0.55947)	(0.62478)	(0.55444)	(0.56793)	(0.44363)
P(nowall_dist)	0.50196*	-0.00457	-0.05534	-0.93838***	0.20131	-0.18078	0.09074	-0.29869	-0.66269*	0.12254
	(0.25953)	(0.27800)	(0.30758)	(0.29169)	(0.26980)	(0.32036)	(0.40597)	(0.33279)	(0.34736)	(0.27491)
P(noinsu_dist)	-0.18428	-0.45471	-0.68662	0.11860	-0.42293	0.92206	2.77763***	-0.25262	-1.12363*	0.81366*
	(0.50273)	(0.47120)	(0.51561)	(0.57110)	(0.53843)	(0.59860)	(0.68722)	(0.57227)	(0.60208)	(0.47775)
P(illiterate_dist)	3.40970	-1.69790	-0.35974	-6.18368**	8.09241***	-2.13967	6.27624**	-0.91336	5.79958*	-4.02653*
	(2.72306)	(2.38799)	(2.54890)	(2.42074)	(2.21651)	(2.47492)	(3.05419)	(2.67059)	(2.97077)	(2.17182)
P(selfwork_dist)	0.10452	0.94610	0.64813	0.72449	0.04474	1.16998	0.75725	0.82496	0.39639	-0.01742
	(0.81375)	(0.77152)	(0.82550)	(0.89612)	(0.84506)	(0.93939)	(0.97928)	(0.90283)	(0.94660)	(0.73619)
P(unreg_dist)	-5.13255	-5.05780	2.94592	6.50263	3.54274	5.46768	13.30646**	2.74515	0.71516	8.09505*
	(3.92929)	(3.97422)	(4.31546)	(4.45610)	(4.01562)	(4.83413)	(5.58059)	(5.17727)	(5.64743)	(4.14589)
icf_index	0.03283	-1.47878***	0.82123	-0.01526	-0.94842*	-0.03244	0.33142	-0.12017	0.27173	-0.29335
	(0.47256)	(0.46027)	(0.52783)	(0.51492)	(0.50687)	(0.51050)	(0.65475)	(0.61311)	(0.67403)	(0.46674)
P(undernut_dist)	1.39671**	0.23059	1.07311*	-1.05221*	0.55196	1.20481*	0.70194	2.20362***	-0.33151	0.41799
	(0.55910)	(0.55267)	(0.61462)	(0.61214)	(0.54944)	(0.61517)	(0.73346)	(0.74301)	(0.82849)	(0.56282)
Mort_distx1000	0.00857	-0.02424	0.01270	0.03078	-0.01013	0.07920***	-0.02618	-0.06350**	0.01722	-0.05087*
	(0.02339)	(0.02225)	(0.02711)	(0.02558)	(0.02254)	(0.02712)	(0.04440)	(0.02871)	(0.02210)	(0.02740)
P(poor_dist)	-1.45046***	0.70115	-0.19480	-0.07363	0.28241	-0.36929	-2.03156***	-1.08696*	-0.10494	-1.02380**
	(0.51902)	(0.53294)	(0.59866)	(0.57265)	(0.51566)	(0.62316)	(0.77505)	(0.64527)	(0.67876)	(0.52231)
lifeexp_dist	0.03029	-0.01379	-0.02033	0.08589	-0.07080	0.20534**	-0.02949	-0.13347	-0.02591	-0.19821**
	(0.07702)	(0.07438)	(0.08979)	(0.08484)	(0.07614)	(0.09430)	(0.15091)	(0.09770)	(0.07896)	(0.08983)
D(ca_food)_1	-0.01488									
	(0.09689)									
D(ca_food)_2	0.08744									
	(0.09856)									
D(ca_food)_3	0.05360									
	(0.09515)									
D(ca_food)_4	-0.03281									
	(0.09097)									
Agropec_firmPC	-0.00401									

	(0.03611)								
D(agro_dist)	0.03093								
	(0.08457)								
D(pec_dist)	-0.15893*								
	(0.09503)								
D(fish_dist)	-0.08889								
	(0.11659)								
Prop(workagro_dist)	-0.95435**								
	(0.47716)								
Prop(workfish_dist)	1.57452								
	(1.79328)								
Hotel_firmPC		-0.02861	0.19280***						
		(0.04018)	(0.04498)						
D(rest_dist)		-0.02697	0.16790*						
		(0.08583)	(0.09401)						
D(tour_dist)		0.02910	0.00555						
		(0.08861)	(0.09733)						
Prop(workrest_dist)		1.66370	-5.69826						
		(3.85588)	(4.16451)						
Com_firmPC				0.01303	-0.00226				
				(0.00887)	(0.00774)				
Prop(workcom_dist)				0.29707	2.92296**				
				(1.32091)	(1.18588)				
D(ca_cloth)_1				-0.13766	0.06932				
				(0.16268)	(0.15386)				
D(ca_cloth)_2				-0.03983	0.10576				
				(0.16713)	(0.15576)				
D(ca_health)_1						0.06234	0.32558**		
						(0.09903)	(0.13331)		
D(ca_health)_2						0.04734	-0.15003		
						(0.09613)	(0.12252)		

n_eventual						0.17462***	0.10571***			
						(0.03073)	(0.04066)			
Health1_distx1000						-0.06369	2.26392**			
						(0.69866)	(1.11030)			
Health2_distx1000						0.28185	-3.86212**			
						(1.25017)	(1.59632)			
Health3_distx1000						-0.13656	-2.17195*			
						(0.78424)	(1.19732)			
Health4_distx1000						-0.03163	-2.22156**			
						(0.71558)	(1.12998)			
Health_firmPC						0.05078	-0.76529***			
						(0.17314)	(0.22785)			
Prop(workhealth_dist)						9.20844**	11.16693*			
						(4.63229)	(6.18071)			
D(ca_transport)								-0.01996	0.26231***	
								(0.09450)	(0.09274)	
comm1_distx1000								0.14392	0.17575	
								(0.13669)	(0.15737)	
comm2_distx1000								1.48066**	-0.84796*	
								(0.58080)	(0.48832)	
comm3_distx1000								-1.40595	-0.46152	
								(0.95157)	(1.02545)	
comm4_distx1000								-0.61898**	-0.63355**	
								(0.29360)	(0.28467)	
comm5_distx1000								0.01134	0.05467	
								(0.04378)	(0.04384)	
comm6_distx1000								1.96459***	-0.00320	
								(0.75639)	(0.63404)	
D(comm_dist)								0.13426	0.01170	
								(0.10552)	(0.10516)	

D(transp_dist)								-0.02967	0.26980	
								(0.17615)	(0.18814)	
transp_distx1000								-0.12705	0.13362	
								(0.11861)	(0.13272)	
D(good_roads_dist)								0.08431	-0.11040	
								(0.18252)	(0.18473)	
Transport_firmPC								0.33054	-0.33736	
								(0.26504)	(0.26187)	
D(TC_dist)								0.13651	-0.05817	
								(0.15102)	(0.13933)	
Prop(worktransp_dist)								4.97339**	0.66752	
								(2.15580)	(2.03411)	
Commun_firmPC								0.28769**	0.07435	
								(0.12359)	(0.11928)	
D(ca_educ)										0.01480
										(0.07609)
D(ca_entert)										-0.07431
										(0.09984)
school_distx1000										-0.05004**
										(0.02507)
Educa_firmPC										0.38569*
										(0.20404)
Entert_firmPC										-1.76933***
										(0.50187)
Prop(workeduca_dist)										-0.15966
										(1.29089)
Constant	-0.00648	4.69367	-1.94904	-15.68530*	-0.71720	-24.07521***	-13.61778	9.33189	-5.41106	6.01634
	(7.94450)	(7.61202)	(8.92143)	(8.65220)	(7.76230)	(9.19451)	(12.67570)	(9.90060)	(8.94634)	(8.58600)
Observations	2,090	1,794	1,794	2,097	2,097	1,732	1,732	2,011	2,011	1,997
LogL	-1201.192	-1151.687	-903.171	-729.423	-1105.531	-757.782	-423.804	-584.848	-604.473	-1005.104

Wald (ident. vars.)	106.73***	77.07***	76.75***	153.07***	85.05***	114.97***	65.55***	172.38***	83.22***	112.79***

q) Markets within groups (SUR)

	Fon		Foff		CC		HEA		TC		ED	
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
θ	-0.81048***	0.46787	0.20671	0.45732*	0.34401	-0.00971	0.24175	-0.13556	0.04227	-0.08267	0.05886	0.10023
	(0.27644)	(0.54355)	(0.27672)	(0.23479)	(0.21437)	(0.27672)	(0.54383)	(0.17860)	(0.23010)	(0.24588)	(0.14459)	(0.10630)
$\theta\ln(\theta M)$	0.01511	0.00988	-0.15215***	-0.07700***	-0.07074***	-0.06351***	-0.07523***	0.01480	-0.07286***	0.12045***	-0.10708***	0.11529***
	(0.01402)	(0.02599)	(0.01686)	(0.01520)	(0.01292)	(0.01909)	(0.02808)	(0.01120)	(0.01553)	(0.01583)	(0.00997)	(0.00801)
$(1-\theta)\ln\left[(1-\theta)M\right]$	-0.09206***	0.06727	-0.12480***	-0.01675	-0.03765	-0.07855**	-0.03100	-0.01219	-0.06602***	0.10945***	-0.10556***	0.13507***
	(0.02689)	(0.05189)	(0.03224)	(0.02698)	(0.02520)	(0.03243)	(0.06548)	(0.02212)	(0.02514)	(0.02573)	(0.01634)	(0.01245)
L(fam_size)	0.00229	-0.08742	0.15528	-0.09606	-0.09076	0.08061	-0.06696	0.01268	-0.00396	-0.13672**	-0.00657	0.02995
	(0.05385)	(0.05776)	(0.12882)	(0.10001)	(0.06746)	(0.10842)	(0.18674)	(0.07104)	(0.06725)	(0.06572)	(0.06378)	(0.05258)
D(mem_son)	0.01675	-0.05722	-0.13201*	0.00363	-0.06716	-0.00728	-0.06554	0.02716	-0.02670	0.02981	-0.24369***	0.14574***
	(0.03516)	(0.04275)	(0.07913)	(0.06098)	(0.04487)	(0.05990)	(0.11402)	(0.04669)	(0.03894)	(0.03855)	(0.05146)	(0.03608)
D(mem_oth)	0.00479	0.00426	0.03046	-0.03210	-0.01738	0.03213	0.06055	-0.07264	0.03878	-0.02131	-0.04123	0.06372**
	(0.03287)	(0.03757)	(0.07983)	(0.06780)	(0.04121)	(0.06560)	(0.10061)	(0.04518)	(0.04552)	(0.04686)	(0.04136)	(0.03129)
P(men)	-0.07926**	0.04133	0.08208	-0.02352	-0.06675	0.03214	0.30153**	-0.03246	0.05389	-0.09501**	0.03780	-0.03635
	(0.03804)	(0.03751)	(0.09118)	(0.07532)	(0.04627)	(0.04554)	(0.13038)	(0.04989)	(0.04708)	(0.04645)	(0.04194)	(0.03177)
D(gender_head)	-0.01042	-0.01184	-0.04732	0.08169	-0.11213**	-0.03366	-0.00411	-0.04467	0.05797	-0.04514	-0.04898	0.07781**
	(0.03907)	(0.03681)	(0.08595)	(0.07056)	(0.04520)	(0.04999)	(0.12963)	(0.04059)	(0.05046)	(0.05159)	(0.04369)	(0.03105)
age_male	-0.00024	0.00122	-0.00525	0.00670**	0.00124	0.00204	0.00468	-0.00143	-0.00276	0.00339*	0.00137	0.00017
	(0.00144)	(0.00136)	(0.00353)	(0.00299)	(0.00177)	(0.00211)	(0.00569)	(0.00186)	(0.00195)	(0.00188)	(0.00177)	(0.00134)
age_female	0.00369**	-0.00092	0.00618	-0.00173	-0.00137	-0.00215	-0.00274	0.00238	-0.00605***	0.00620***	-0.00274	0.00032
	(0.00167)	(0.00172)	(0.00394)	(0.00328)	(0.00208)	(0.00262)	(0.00619)	(0.00223)	(0.00220)	(0.00215)	(0.00211)	(0.00163)
year_educ_male	0.00600***	0.00690***	-0.01323**	0.01466***	-0.00858***	0.01111***	0.00542	-0.00020	-0.00561*	0.00390	-0.00277	-0.00169
	(0.00208)	(0.00201)	(0.00523)	(0.00466)	(0.00262)	(0.00345)	(0.00947)	(0.00301)	(0.00293)	(0.00291)	(0.00272)	(0.00191)
year_educ_female	-0.00111	0.00145	-0.00647	0.00639	0.00115	0.00216	-0.00695	-0.00448*	-0.01057***	0.00727***	-0.00494**	0.00305*
	(0.00182)	(0.00196)	(0.00472)	(0.00424)	(0.00240)	(0.00321)	(0.00680)	(0.00267)	(0.00260)	(0.00260)	(0.00237)	(0.00170)

D(chronic_male)	0.00813	-0.00880	-0.00018	-0.01467	-0.00386	-0.00053	0.03803	0.03359*	0.00512	-0.00663	0.01493	-0.03064**
	(0.01493)	(0.01459)	(0.03612)	(0.02919)	(0.01885)	(0.02078)	(0.05632)	(0.01897)	(0.01783)	(0.01818)	(0.01720)	(0.01314)
D(chronic_female)	0.01356	0.00401	-0.06241*	0.01315	-0.00964	0.01494	-0.02088	-0.01165	0.03471**	-0.03278*	0.00333	-0.00840
	(0.01429)	(0.01471)	(0.03395)	(0.02782)	(0.01813)	(0.02013)	(0.05235)	(0.01870)	(0.01705)	(0.01702)	(0.01602)	(0.01263)
Hrs_male	0.00015	-0.00007	0.00010	0.00128*	-0.00098**	0.00068	0.00063	-0.00059	-0.00049	0.00034	0.00054	-0.00018
	(0.00037)	(0.00034)	(0.00087)	(0.00072)	(0.00044)	(0.00052)	(0.00135)	(0.00047)	(0.00045)	(0.00044)	(0.00040)	(0.00032)
Hrs_female	0.00147***	0.00030	0.00186**	-0.00058	0.00086**	0.00003	0.00184	0.00003	0.00014	0.00002	-0.00051	-0.00002
	(0.00034)	(0.00036)	(0.00077)	(0.00066)	(0.00042)	(0.00056)	(0.00127)	(0.00041)	(0.00041)	(0.00040)	(0.00039)	(0.00031)
D(Hrs_male)	-0.09307***	0.05079	-0.02858	-0.13221*	-0.04897	-0.16227**	-0.15161	0.08622*	-0.00165	0.02421	-0.08405**	0.01131
	(0.03447)	(0.04027)	(0.10902)	(0.07043)	(0.04479)	(0.06406)	(0.13364)	(0.05161)	(0.04154)	(0.04194)	(0.04256)	(0.03282)
D(Hrs_female)	-0.03116	0.00557	-0.16987***	-0.03571	0.01226	-0.00421	-0.04819	0.04518*	-0.05328**	0.06369**	0.02412	0.04519**
	(0.02064)	(0.02161)	(0.05140)	(0.04258)	(0.02611)	(0.02921)	(0.07858)	(0.02731)	(0.02640)	(0.02597)	(0.02603)	(0.01858)
D(Lima)	0.13893***	0.04908**	0.07692*	-0.00101	-0.02533	0.05714**	-0.08003	0.03080	0.12234***	-0.11483***	0.07854***	-0.01716
	(0.01718)	(0.01944)	(0.03976)	(0.03257)	(0.02092)	(0.02720)	(0.06824)	(0.02150)	(0.01746)	(0.01789)	(0.02151)	(0.01412)
Pcoh_1	-0.03720	-0.00331	0.36589	-0.25284	0.23039	-0.06011	0.49412	-0.25809	-0.34541	0.46658	-0.40514	0.13809
	(0.25464)	(0.21924)	(0.70049)	(0.50148)	(0.35449)	(0.38526)	(0.57985)	(0.30508)	(0.31190)	(0.31585)	(0.42353)	(0.24377)
Pcoh_2	0.01911	0.01637	0.37344	-0.28152	0.23938	-0.14136	0.42462	-0.29566	-0.20447	0.30542	-0.52266	0.00328
	(0.25033)	(0.21453)	(0.69005)	(0.49419)	(0.35063)	(0.37401)	(0.55422)	(0.29807)	(0.30579)	(0.31048)	(0.42041)	(0.23997)
Pcoh_3	0.10876	-0.01456	0.30895	-0.27458	0.25718	-0.11622	0.50903	-0.38453	-0.49194	0.46873	-0.59502	-0.01014
	(0.25180)	(0.21792)	(0.69376)	(0.49125)	(0.35256)	(0.38233)	(0.59234)	(0.30245)	(0.30989)	(0.31442)	(0.42099)	(0.23959)
Pcoh_4	0.11934	0.08970	0.09529	-0.41070	0.26975	-0.10777	0.51154	-0.39312	-0.27306	0.10634	-0.64573	0.20676
	(0.25475)	(0.22205)	(0.69385)	(0.48677)	(0.35460)	(0.37706)	(0.55559)	(0.30422)	(0.31016)	(0.31479)	(0.42174)	(0.24128)
Pcoh_5	0.08533	-0.04006	0.25286	-0.37684	0.22794	0.22133	0.36655	-0.38158	-0.28689	0.13184	-0.69532*	0.28512
	(0.25241)	(0.21844)	(0.68435)	(0.47288)	(0.34954)	(0.38989)	(0.58549)	(0.30217)	(0.30455)	(0.30551)	(0.41783)	(0.23661)
Pcoh_6	0.08199	-0.03164	0.42466	-0.36249	0.20270	0.20256	0.31765	-0.24150	-0.27354	0.17396	-0.47063	0.27721
	(0.26044)	(0.23785)	(0.68979)	(0.47241)	(0.35689)	(0.43068)	(0.59923)	(0.30445)	(0.31022)	(0.30931)	(0.42811)	(0.24134)
Pcoh_7	0.11386	0.02768	0.51240	-0.36076	0.20990	0.31557	0.49037	-0.32675	-0.10072	-0.00310	-0.43220	0.26783
	(0.26528)	(0.24570)	(0.70042)	(0.47470)	(0.36301)	(0.43900)	(0.64405)	(0.31360)	(0.31082)	(0.31058)	(0.43243)	(0.24455)
Pcoh_8	0.09970	-0.04126	0.40100	-0.39175	0.18124	0.22914	0.36297	-0.33602	-0.10264	0.00759	-0.38503	0.24811
	(0.27046)	(0.25695)	(0.70394)	(0.47707)	(0.36910)	(0.44747)	(0.72317)	(0.32435)	(0.31947)	(0.31841)	(0.43696)	(0.24941)

PHI		-0.03192	0.05354	0.05364	-0.36065***	-0.00037	0.21562***	0.20992**	0.50701***	0.50485***	-0.00341	
		(0.03285)	(0.08386)	(0.08351)	(0.06138)	(0.05201)	(0.06948)	(0.08770)	(0.03797)	(0.03435)	(0.05390)	
Constant	1.09156***	-0.42390	1.10421	0.82192	1.02430**	0.41278	-0.08064	1.15450***	1.63057***	-0.80908**	1.88313***	-0.80745***
	(0.35996)	(0.51144)	(0.74151)	(0.53991)	(0.40960)	(0.50652)	(0.80462)	(0.36323)	(0.38115)	(0.38465)	(0.45697)	(0.26117)
Observations	2,090		1,794		2,097		1,732		2,011		1,997	
LogL	1251.457		-1059.761		921.617		21.343		803.544		151.769	

Note: M corresponds to each consumption group’s expenditure.

r) IV model for informal Fon consumption (OLS version)

	With labour supply				Without labour supply			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
θ	0.00786	0.01928	0.00724	0.00941	-0.00658	0.00640	-0.00667	-0.00472
	(0.02245)	(0.02183)	(0.02245)	(0.02227)	(0.02192)	(0.02138)	(0.02190)	(0.02177)
L(expend_Fon)	0.00900	-0.01245	0.01151	0.00991	0.00628	-0.01500	0.00892	0.00730
	(0.01141)	(0.01138)	(0.01148)	(0.01128)	(0.01145)	(0.01141)	(0.01151)	(0.01132)
L(fam_size)	0.04894	0.05757	0.04072	0.04319	0.05086	0.05871	0.04259	0.04506
	(0.04295)	(0.04145)	(0.04324)	(0.04257)	(0.04282)	(0.04129)	(0.04313)	(0.04245)
D(mem_son)	0.03070	0.02169	0.03617	0.03237	0.03338	0.02403	0.03879	0.03488
	(0.03190)	(0.03113)	(0.03204)	(0.03169)	(0.03191)	(0.03115)	(0.03205)	(0.03169)
D(mem_oth)	0.00782	0.00920	0.00983	0.00645	0.00704	0.00809	0.00910	0.00557
	(0.03235)	(0.03108)	(0.03219)	(0.03202)	(0.03232)	(0.03115)	(0.03214)	(0.03200)
P(men)	-0.07477**	-0.06379*	-0.07129*	-0.07428**	-0.07442**	-0.06325*	-0.07079*	-0.07385**
	(0.03669)	(0.03551)	(0.03671)	(0.03655)	(0.03671)	(0.03552)	(0.03673)	(0.03656)
D(gender_head)	-0.00598	-0.00664	-0.00593	-0.00935	-0.00218	-0.00301	-0.00222	-0.00560
	(0.03664)	(0.03476)	(0.03720)	(0.03593)	(0.03621)	(0.03439)	(0.03678)	(0.03552)
Pcoh_1	-0.21108	-0.37588	-0.21619	-0.20852	-0.24141	-0.40798*	-0.24493	-0.23932
	(0.22800)	(0.22934)	(0.22871)	(0.22578)	(0.22982)	(0.23109)	(0.23106)	(0.22760)
Pcoh_2	-0.11914	-0.25252	-0.12120	-0.11390	-0.15363	-0.28708	-0.15375	-0.14833
	(0.22415)	(0.22630)	(0.22494)	(0.22193)	(0.22597)	(0.22800)	(0.22728)	(0.22383)
Pcoh_3	-0.06078	-0.19740	-0.07794	-0.05456	-0.08511	-0.22309	-0.10106	-0.07924
	(0.22690)	(0.22855)	(0.22735)	(0.22475)	(0.22896)	(0.23035)	(0.22992)	(0.22687)
Pcoh_4	-0.02342	-0.15060	-0.04086	-0.01685	-0.04036	-0.17025	-0.05722	-0.03393
	(0.22818)	(0.22999)	(0.22863)	(0.22613)	(0.23041)	(0.23196)	(0.23138)	(0.22843)
Pcoh_5	-0.02907	-0.17209	-0.03712	-0.02455	-0.05622	-0.20179	-0.06302	-0.05202
	(0.22646)	(0.22893)	(0.22701)	(0.22421)	(0.22809)	(0.23038)	(0.22918)	(0.22586)
Pcoh_6	-0.01039	-0.16908	-0.01906	-0.01303	-0.03583	-0.19814	-0.04320	-0.03903
	(0.23059)	(0.23290)	(0.23106)	(0.22839)	(0.23196)	(0.23416)	(0.23295)	(0.22976)
Pcoh_7	0.07381	-0.09461	0.06875	0.07920	0.04904	-0.12289	0.04528	0.05394
	(0.23284)	(0.23478)	(0.23334)	(0.23067)	(0.23427)	(0.23609)	(0.23529)	(0.23209)
Pcoh_8	0.11533	-0.04406	0.10534	0.11896	0.08980	-0.07381	0.08112	0.09272
	(0.22942)	(0.23254)	(0.22992)	(0.22732)	(0.23053)	(0.23363)	(0.23156)	(0.22840)
Hrs_both	0.00052***	0.00047***	0.00050***	0.00051***				
	(0.00018)	(0.00017)	(0.00017)	(0.00018)				
D(Hrs_both)	-0.07638*	-0.04967	-0.07317*	-0.07187*				
	(0.04386)	(0.04381)	(0.04350)	(0.04368)				
D(Lima)	0.13300***	0.12390***	0.12761***	0.12056***	0.13413***	0.12637***	0.12840***	0.12138***
	(0.01579)	(0.03202)	(0.01606)	(0.01617)	(0.01574)	(0.03204)	(0.01601)	(0.01613)
L(pop_urb_dist)		0.33998***				0.34489***		
		(0.06902)				(0.06936)		
L(pop_urb_dist)-sq		-0.01408***				-0.01432***		
		(0.00328)				(0.00329)		
pop_den_distx1000		-0.00247				-0.00244		
		(0.00150)				(0.00150)		
D(not_slum)		0.05277***				0.05302***		
		(0.01196)				(0.01194)		
D(mid_city)		-0.02525*				-0.02517*		
		(0.01397)				(0.01402)		
D(border)		0.07924***				0.08282***		
		(0.01859)				(0.01855)		
D(tongue)			0.04316***				0.04370***	
			(0.01415)				(0.01416)	

D(migrant)			0.03359**				0.03465***	
			(0.01315)				(0.01316)	
D(social)			-0.00594				-0.00695	
			(0.01221)				(0.01225)	
Station_distx1000				-0.13356***				-0.13772***
				(0.03399)				(0.03359)
Muni_persx1000				-0.00328*				-0.00328*
				(0.00173)				(0.00170)
L(budget_dist_pc)				-0.00112				-0.00160
				(0.01026)				(0.01023)
D(sdp_dist)				-0.06286*				-0.06172*
				(0.03424)				(0.03463)
Constant	0.49072*	-1.24153***	0.45175*	0.51473*	0.51116**	-1.22307***	0.47039*	0.54070**
	(0.26398)	(0.44272)	(0.26463)	(0.26345)	(0.25993)	(0.44246)	(0.26112)	(0.25929)
Observations	2,142	2,142	2,142	2,142	2,142	2,142	2,142	2,142
LogL	-273.311	-210.055	-263.974	-264.031	-278.260	-214.005	-268.538	-268.742

s) First-stage regression for IV models for Fon shares (with labour supply)

	L(expenditure_Fon)				θ			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
L(fam_size)	0.2863361***	0.2806483***	0.3052114***	0.2875527***	-0.0216221	-0.0210947	-0.0137437	-0.0187512
	(0.08762)	(0.08560)	(0.08805)	(0.08669)	(0.04027)	(0.04028)	(0.04052)	(0.04040)
D(mem_son)	0.2143842***	0.2142722***	0.2069655***	0.2167356***	-0.0014088	-0.0006027	-0.0036641	-0.0030949
	(0.06715)	(0.06736)	(0.06744)	(0.06712)	(0.03106)	(0.03092)	(0.03120)	(0.03115)
D(mem_oth)	0.0102579	0.0062688	0.0065656	0.0087224	-0.0251461	-0.0265724	-0.0277601	-0.025161
	(0.05730)	(0.05657)	(0.05747)	(0.05747)	(0.03296)	(0.03300)	(0.03304)	(0.03310)
P(men)	0.1067553	0.1175679	0.1130172	0.113451	0.0047234	0.0047219	0.0043471	0.0057609
	(0.07647)	(0.07542)	(0.07638)	(0.07665)	(0.03388)	(0.03405)	(0.03394)	(0.03403)
D(gender_head)	0.0364591	0.0497447	0.0417154	0.0371746	0.1338942***	0.132626***	0.1339685***	0.1344142***
	(0.08408)	(0.08430)	(0.08411)	(0.08277)	(0.03633)	(0.03633)	(0.03664)	(0.03630)
Pcoh_1	-0.5735691	-0.7242102*	-0.5519869	-0.6130142	0.1581487	0.1659335	0.1512578	0.1491591
	(0.40406)	(0.40646)	(0.40202)	(0.40886)	(0.30197)	(0.29588)	(0.30558)	(0.30271)
Pcoh_2	-0.4127199	-0.5570166	-0.3836162	-0.4455544	-0.00399	0.0062154	-0.0096124	-0.0112539
	(0.38902)	(0.39206)	(0.38685)	(0.39481)	(0.30139)	(0.29499)	(0.30523)	(0.30208)
Pcoh_3	-0.2942532	-0.4431698	-0.2557384	-0.3223287	-0.0124594	-0.0015165	-0.011152	-0.0203379
	(0.39858)	(0.40103)	(0.39563)	(0.40404)	(0.30041)	(0.29413)	(0.30388)	(0.30106)
Pcoh_4	-0.325117	-0.4731626	-0.3109738	-0.377706	0.0190996	0.025208	0.0136942	0.0127614
	(0.40405)	(0.40607)	(0.40138)	(0.40969)	(0.30146)	(0.29503)	(0.30511)	(0.30196)
Pcoh_5	-0.5338057	-0.6813312*	-0.5147844	-0.5627934	0.013044	0.0202947	0.0143571	0.0070552
	(0.38987)	(0.39166)	(0.38808)	(0.39512)	(0.29900)	(0.29262)	(0.30236)	(0.29962)
Pcoh_6	-0.5927034	-0.7494015*	-0.564542	-0.6322816	-0.0298701	-0.0177435	-0.0267101	-0.0351096
	(0.40396)	(0.40663)	(0.40209)	(0.40993)	(0.30199)	(0.29576)	(0.30530)	(0.30272)
Pcoh_7	-0.5495615	-0.7039098*	-0.5265385	-0.5883376	-0.0806093	-0.0742519	-0.0785701	-0.0883246
	(0.40758)	(0.41141)	(0.40562)	(0.41336)	(0.30576)	(0.29914)	(0.30906)	(0.30649)
Pcoh_8	-0.3759724	-0.5502392	-0.3484812	-0.4185302	-0.1001237	-0.0912745	-0.0997709	-0.1084257
	(0.39820)	(0.40240)	(0.39611)	(0.40433)	(0.30205)	(0.29560)	(0.30557)	(0.30282)
Hrs_both	-0.0022365***	-0.0022316***	-0.002222***	-0.0022416***	-0.0016928***	-0.0016936***	-0.0016791***	-0.001691***

	(0.00042)	(0.00042)	(0.00042)	(0.00042)	(0.00018)	(0.00018)	(0.00018)	(0.00018)
D(Hrs_both)	0.0979292	0.1451162	0.1023745	0.0854048	0.1089409	0.1070885*	0.1101389*	0.1077836*
	(0.09816)	(0.09953)	(0.09806)	(0.09809)	(0.05867)	(0.05920)	(0.05858)	(0.05864)
D(Lima)	0.300507***	0.1836383***	0.2860999***	0.3051301***	-0.0183473	-0.0300198	-0.0221084	-0.0162712
	(0.03426)	(0.06714)	(0.03433)	(0.03442)	(0.01498)	(0.02985)	(0.01509)	(0.01521)
theta_educ_prima	0.1272289***	0.1082058***	0.1296094***	0.1279237***	0.0389426***	0.0410557***	0.039985***	0.0390872***
	(0.02536)	(0.02551)	(0.02544)	(0.02539)	(0.01247)	(0.01261)	(0.01250)	(0.01256)
year_educ_female	0.0282876	0.0188017	0.0242718	0.0300442*	0.0568976***	0.0579221***	0.0550449***	0.05702***
	(0.01744)	(0.01747)	(0.01751)	(0.01744)	(0.00881)	(0.00884)	(0.00885)	(0.00887)
year_educ_female-sq	0.0003256	0.0004863	0.0005232	0.0002231	-0.002991***	-0.003022***	-0.002921***	-0.0029993***
	(0.00075)	(0.00075)	(0.00075)	(0.00075)	(0.00038)	(0.00038)	(0.00038)	(0.00038)
D(risk7)	0.0805201	0.0785648	0.076467	0.0862917	0.0455305*	0.044595*	0.0443043*	0.0459134*
	(0.06271)	(0.06323)	(0.06228)	(0.06265)	(0.02669)	(0.02663)	(0.02668)	(0.02677)
L(asset)	0.1308416***	0.1255911***	0.1289422***	0.1315638***	-0.0034024	-0.0029717	-0.0044881	-0.0035434
	(0.01059)	(0.01042)	(0.01053)	(0.01057)	(0.00444)	(0.00447)	(0.00442)	(0.00445)
L(pop_urb_dist)		0.4248284**				-0.0487762		
		(0.16651)				(0.06427)		
L(pop_urb_dist)-sq		-0.0174487**				0.0019439		
		(0.00772)				(0.00308)		
pop_den_distx1000		0.0085288**				0.0007593		
		(0.00336)				(0.00162)		
D(not_slum)		0.0774306***				0.0063109		
		(0.02711)				(0.01244)		
D(mid_city)		0.0108429				-0.0224367*		
		(0.03036)				(0.01354)		
D(border)		0.0310749				0.0068517		
		(0.03672)				(0.01850)		
D(tongue)			-0.0499408				-0.0335845**	
			(0.03184)				(0.01423)	
D(migrant)			0.0341891				0.0171236	
			(0.02726)				(0.01241)	
D(social)			-0.0678107***				-0.0108992	
			(0.02574)				(0.01155)	
Station_distx1000				-0.0621391				0.0012379
				(0.07509)				(0.04849)
Muni_persx1000				-0.0015389				0.0020043
				(0.00695)				(0.00123)
L(budget_dist_pc)				0.053817**				-0.0009492
				(0.02392)				(0.00999)
D(sdp_dist)				-0.0168214				0.018512
				(0.07537)				(0.02865)
Constant	6.723974***	4.390253***	6.714677***	6.490674***	0.4915182	0.7765834*	0.5009891	0.4937143
	(0.46067)	(0.95682)	(0.46149)	(0.47567)	(0.31990)	(0.43021)	(0.32327)	(0.32737)
Observations	2,142	2,142	2,142	2,142	2,142	2,142	2,142	2,142
R-squared	0.2445	0.2658	0.2484	0.2467	0.132	0.1354	0.1354	0.133

t) First-stage regression for IV models for Fon shares (without labour supply)

	L(expenditure_Fon)				θ			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
L(fam_size)	0.2874371***	0.2792781***	0.3072909***	0.2878319***	-0.016749	-0.0175234	-0.0081469	-0.0144433
	(0.08821)	(0.08634)	(0.08864)	(0.08721)	(0.04107)	(0.04114)	(0.04133)	(0.04119)
D(mem_son)	0.2069824***	0.2073368***	0.199207***	0.2100302***	-0.0093965	-0.0079137	-0.0118841	-0.0106826
	(0.06841)	(0.06855)	(0.06870)	(0.06836)	(0.03114)	(0.03108)	(0.03129)	(0.03125)
D(mem_oth)	0.0240108	0.0167546	0.0193084	0.0236689	-0.0198288	-0.0218215	-0.0230852	-0.0195853
	(0.05825)	(0.05738)	(0.05833)	(0.05837)	(0.03321)	(0.03325)	(0.03328)	(0.03334)
P(men)	0.1037625	0.1150435	0.1094787	0.1096722	0.0025942	0.0025213	0.0018089	0.0030401
	(0.07712)	(0.07607)	(0.07705)	(0.07731)	(0.03486)	(0.03500)	(0.03493)	(0.03499)
D(gender_head)	0.0291611	0.0425008	0.0342658	0.0295973	0.1280272***	0.1266019***	0.1279973***	0.1283822***
	(0.08315)	(0.08329)	(0.08323)	(0.08165)	(0.03669)	(0.03655)	(0.03701)	(0.03669)
Pcoh_1	-0.3564595	-0.5246614	-0.3408949	-0.387828	0.3142017	0.3193505	0.3031131	0.309086
	(0.39962)	(0.40292)	(0.39845)	(0.40583)	(0.30986)	(0.30488)	(0.31406)	(0.31061)
Pcoh_2	-0.1917646	-0.3533744	-0.1691339	-0.2179071	0.1483084	0.1555765	0.1381793	0.1433655
	(0.38414)	(0.38815)	(0.38284)	(0.39146)	(0.30989)	(0.30450)	(0.31423)	(0.31062)
Pcoh_3	-0.1050958	-0.2732469	-0.0707616	-0.1253084	0.1091505	0.1171593	0.1078111	0.1039092
	(0.39627)	(0.39940)	(0.39425)	(0.40327)	(0.30803)	(0.30281)	(0.31206)	(0.30871)
Pcoh_4	-0.1627651	-0.3280236	-0.1528452	-0.2079144	0.1197314	0.124637	0.1116995	0.1156993
	(0.40196)	(0.40471)	(0.40016)	(0.40887)	(0.31105)	(0.30557)	(0.31523)	(0.31165)
Pcoh_5	-0.3180297	-0.4866043	-0.3033571	-0.3386244	0.151626	0.1570678	0.1502299	0.1479308
	(0.38713)	(0.38985)	(0.38610)	(0.39407)	(0.30696)	(0.30165)	(0.31089)	(0.30761)
Pcoh_6	-0.3731382	-0.5550232	-0.3496894	-0.4028654	0.1105036	0.1197547	0.1107894	0.1083078
	(0.40105)	(0.40463)	(0.39999)	(0.40843)	(0.31012)	(0.30507)	(0.31397)	(0.31089)
Pcoh_7	-0.340934	-0.5189267	-0.3225524	-0.3701955	0.0519087	0.0560508	0.0510979	0.0471719
	(0.40348)	(0.40825)	(0.40245)	(0.41071)	(0.31397)	(0.30857)	(0.31786)	(0.31475)
Pcoh_8	-0.1496671	-0.3524831	-0.1278878	-0.1815558	0.0434485	0.0486655	0.0402798	0.0384268
	(0.39490)	(0.39996)	(0.39365)	(0.40247)	(0.31037)	(0.30511)	(0.31441)	(0.31120)
D(Lima)	0.3009861***	0.172806**	0.2868914***	0.3065272***	-0.0176292	-0.0369379	-0.021148	-0.0148052
	(0.03514)	(0.06956)	(0.03516)	(0.03528)	(0.01547)	(0.03006)	(0.01559)	(0.01574)
theta_educ_prima	0.1310116***	0.1120577***	0.1336611***	0.1318579***	0.0419933***	0.0439289***	0.0432206***	0.0422976***
	(0.02534)	(0.02541)	(0.02543)	(0.02536)	(0.01266)	(0.01282)	(0.01268)	(0.01274)
year_educ_female	0.0270207	0.0170677	0.0225344	0.0288752*	0.0559976***	0.0567456***	0.0537988***	0.0561802***
	(0.01745)	(0.01745)	(0.01755)	(0.01745)	(0.00889)	(0.00891)	(0.00893)	(0.00893)
year_educ_female-sq	0.0004855	0.0006638	0.0006955	0.0003798	-0.0028809***	-0.0029016***	-0.0028018***	-0.0028918***
	(0.00075)	(0.00075)	(0.00075)	(0.00075)	(0.00038)	(0.00038)	(0.00038)	(0.00038)
D(risk7)	0.0621036	0.0594658	0.0583262	0.0674177	0.0326945	0.0311767	0.0316915	0.0328063
	(0.06237)	(0.06287)	(0.06179)	(0.06242)	(0.02651)	(0.02649)	(0.02640)	(0.02659)
L(asset)	0.1230444***	0.1176553***	0.1209895***	0.1236983***	-0.0103031**	-0.0098302**	-0.0114781**	-0.0105715
	(0.01090)	(0.01070)	(0.01082)	(0.01087)	(0.00456)	(0.00459)	(0.00454)	(0.00456)
D(house1)	0.0295482	0.0241008	0.0287137	0.034095	-0.0121048	-0.0098793	-0.0122316	-0.0118686
	(0.03965)	(0.03866)	(0.03944)	(0.03983)	(0.02208)	(0.02225)	(0.02197)	(0.02219)
L(pop_urb_dist)		0.4186855**				-0.0531487		

		(0.16762)				(0.06606)		
L(pop_urb_dist)-sq		-0.0170306**				0.0022392		
		(0.00779)				(0.00317)		
pop_den_distx1000		0.0084922**				0.0007554		
		(0.00348)				(0.00167)		
D(not_slum)		0.0781417***				0.0070608		
		(0.02727)				(0.01264)		
D(mid_city)		0.0062827				-0.0245477*		
		(0.03053)				(0.01387)		
D(border)		0.0140438				-0.0070176		
		(0.03609)				(0.01901)		
D(tongue)			-0.0587655*				-0.0401179***	
			(0.03219)				(0.01471)	
D(migrant)			0.0335221				0.0165262	
			(0.02739)				(0.01265)	
D(social)			-0.0663367**				-0.0098416	
			(0.02601)				(0.01185)	
Station_distx1000				-0.0393366				0.0202458
				(0.07480)				(0.05611)
Muni_persx1000				-0.0012476				0.0022099
				(0.00706)				(0.00136)
L(budget_dist_pc)				0.0527569**				-0.0018856
				(0.02412)				(0.01045)
D(sdp_dist)				-0.0226302				0.0140404
				(0.07774)				(0.02985)
Constant	6.476497***	4.241068***	6.483705***	6.225002***	0.3689078	0.673102	0.3880934	0.3714068
	(0.44716)	(0.95374)	(0.44798)	(0.46748)	(0.32190)	(0.44123)	(0.32597)	(0.32877)
Observations	2,142	2,142	2,142	2,142	2,142	2,142	2,142	2,142
R-squared	0.2305	0.2521	0.2346	0.2326	0.0854	0.0889	0.0896	0.0865

u) Second-stage regression for IV model for informal Fon shares (2SLS version)

	With labour supply				Without labour supply			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
θ	0.26964**	0.26080**	0.25840**	0.26260**	0.32146***	0.29603**	0.31028**	0.31076**
	(0.10839)	(0.10449)	(0.11021)	(0.10729)	(0.12323)	(0.11722)	(0.12534)	(0.12163)
L(expend_Fon)	0.11017***	0.06604**	0.12884***	0.11542***	0.12870***	0.08324**	0.14936***	0.13386***
	(0.02952)	(0.03096)	(0.03084)	(0.02938)	(0.03305)	(0.03443)	(0.03488)	(0.03292)
L(fam_size)	0.03146	0.04397	0.01300	0.02390	0.02897	0.04139	0.00831	0.02126
	(0.04378)	(0.04221)	(0.04475)	(0.04369)	(0.04456)	(0.04263)	(0.04586)	(0.04449)
D(mem_son)	-0.00062	0.00003	0.00123	-0.00049	-0.00069	-0.00061	0.00124	-0.00085
	(0.03518)	(0.03386)	(0.03546)	(0.03496)	(0.03589)	(0.03435)	(0.03629)	(0.03564)
D(mem_oth)	0.01335	0.01513	0.01656	0.01180	0.01116	0.01321	0.01485	0.00935
	(0.03289)	(0.03151)	(0.03277)	(0.03254)	(0.03339)	(0.03188)	(0.03329)	(0.03305)
P(men)	-0.08636**	-0.07382**	-0.08406**	-0.08691**	-0.08800**	-0.07524**	-0.08545**	-0.08838**
	(0.03878)	(0.03717)	(0.03903)	(0.03866)	(0.03994)	(0.03804)	(0.04026)	(0.03977)
D(gender_head)	-0.04520	-0.04313	-0.04431	-0.04767	-0.04797	-0.04416	-0.04713	-0.04997
	(0.04010)	(0.03762)	(0.04076)	(0.03958)	(0.04005)	(0.03733)	(0.04077)	(0.03957)
Pcoh_1	-0.19117	-0.35414	-0.18200	-0.17968	-0.28381	-0.43224*	-0.27127	-0.27150

	(0.21498)	(0.21660)	(0.21389)	(0.21271)	(0.22158)	(0.22075)	(0.22206)	(0.21927)
Pcoh_2	-0.06785	-0.20258	-0.06078	-0.05605	-0.15610	-0.27788	-0.14590	-0.14311
	(0.21072)	(0.21244)	(0.20953)	(0.20842)	(0.21620)	(0.21547)	(0.21657)	(0.21401)
Pcoh_3	-0.01949	-0.15682	-0.03278	-0.00700	-0.08627	-0.21329	-0.09884	-0.07339
	(0.21320)	(0.21470)	(0.21163)	(0.21096)	(0.21897)	(0.21801)	(0.21906)	(0.21683)
Pcoh_4	0.01081	-0.11640	0.00315	0.02582	-0.04043	-0.15996	-0.04675	-0.02434
	(0.21380)	(0.21553)	(0.21226)	(0.21170)	(0.22016)	(0.21932)	(0.22016)	(0.21822)
Pcoh_5	0.02581	-0.11966	0.02609	0.03713	-0.04935	-0.18413	-0.04757	-0.03746
	(0.21209)	(0.21455)	(0.21049)	(0.20960)	(0.21688)	(0.21688)	(0.21687)	(0.21444)
Pcoh_6	0.06091	-0.10061	0.05801	0.06497	-0.00860	-0.16074	-0.01026	-0.00468
	(0.21703)	(0.21904)	(0.21560)	(0.21465)	(0.22080)	(0.22058)	(0.22096)	(0.21851)
Pcoh_7	0.14770	-0.01686	0.14451	0.15868	0.08478	-0.07215	0.08281	0.09557
	(0.22005)	(0.22148)	(0.21855)	(0.21775)	(0.22344)	(0.22266)	(0.22355)	(0.22119)
Pcoh_8	0.18884	0.03291	0.17982	0.19829	0.11932	-0.02805	0.11135	0.12830
	(0.21616)	(0.21841)	(0.21461)	(0.21397)	(0.21911)	(0.21922)	(0.21916)	(0.21698)
Hrs_both	0.00112***	0.00099***	0.00111***	0.00110***				
	(0.00029)	(0.00027)	(0.00029)	(0.00029)				
D(Hrs_both)	-0.10941**	-0.08322*	-0.10788**	-0.10286**				
	(0.04801)	(0.04700)	(0.04730)	(0.04771)				
D(Lima)	0.10376***	0.11309***	0.09635***	0.08856***	0.09895***	0.11597***	0.09118***	0.08289***
	(0.01883)	(0.03360)	(0.01915)	(0.01931)	(0.01987)	(0.03435)	(0.02022)	(0.02040)
L(pop_urb_dist)		0.30006***				0.29935***		
		(0.07365)				(0.07600)		
L(pop_urb_dist)-sq		-0.01238***				-0.01244***		
		(0.00348)				(0.00358)		
pop_den_distx1000		-0.00317*				-0.00332*		
		(0.00165)				(0.00173)		
D(not_slum)		0.04277***				0.03998***		
		(0.01317)				(0.01354)		
D(mid_city)		-0.01972				-0.01801		
		(0.01474)				(0.01519)		
D(border)		0.07619***				0.08409***		
		(0.01966)				(0.01987)		
D(tongue)			0.06620***				0.07363***	
			(0.01662)				(0.01781)	
D(migrant)			0.02348*				0.02290	
			(0.01412)				(0.01453)	
D(social)			0.00358				0.00410	
			(0.01324)				(0.01377)	
Station_distx1000				-0.13594***				-0.14690***
				(0.03308)				(0.03495)
Muni_persx1000				-0.00384**				-0.00402***
				(0.00150)				(0.00146)
L(budget_dist_pc)				-0.00510				-0.00624
				(0.01053)				(0.01077)
D(sdp_dist)				-0.06507**				-0.06318*
				(0.03241)				(0.03393)
Constant	-0.53409	-1.84047***	-0.69085*	-0.52146	-0.66219*	-1.92725***	-0.83605**	-0.63449*
	(0.36702)	(0.47951)	(0.37564)	(0.36477)	(0.38336)	(0.49497)	(0.39680)	(0.38062)
Observations	2,142	2,142	2,142	2,142	2,142	2,142	2,142	2,142
LogL	-391.546	-302.536	-394.581	-383.647	-461.121	-355.022	-467.245	-450.639

v) Second-stage regression for IV model for informal Fon shares (LIML version)

	With labour supply				Without labour supply			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
θ	0.27536**	0.26812**	0.26097**	0.26770**	0.33365***	0.31477**	0.31653**	0.32224**
	(0.11060)	(0.10747)	(0.11122)	(0.10928)	(0.12776)	(0.12467)	(0.12769)	(0.12594)
L(expend_Fon)	0.11136***	0.06760**	0.12944***	0.11651***	0.13131***	0.08753**	0.15083***	0.13638***
	(0.02986)	(0.03146)	(0.03100)	(0.02969)	(0.03384)	(0.03586)	(0.03532)	(0.03368)
L(fam_size)	0.03128	0.04371	0.01287	0.02372	0.02855	0.04067	0.00796	0.02082
	(0.04384)	(0.04230)	(0.04478)	(0.04375)	(0.04476)	(0.04292)	(0.04598)	(0.04469)
D(mem_son)	-0.00092	-0.00034	0.00110	-0.00076	-0.00123	-0.00150	0.00096	-0.00136
	(0.03526)	(0.03396)	(0.03550)	(0.03503)	(0.03607)	(0.03463)	(0.03639)	(0.03581)
D(mem_oth)	0.01348	0.01531	0.01662	0.01191	0.01133	0.01355	0.01496	0.00951
	(0.03294)	(0.03157)	(0.03280)	(0.03259)	(0.03353)	(0.03207)	(0.03336)	(0.03318)
P(men)	-0.08651**	-0.07403**	-0.08414**	-0.08706**	-0.08831**	-0.07581**	-0.08563**	-0.08870**
	(0.03885)	(0.03726)	(0.03907)	(0.03872)	(0.04013)	(0.03833)	(0.04036)	(0.03995)
D(gender_head)	-0.04602	-0.04420	-0.04468	-0.04841	-0.04963	-0.04676	-0.04799	-0.05155
	(0.04028)	(0.03785)	(0.04084)	(0.03975)	(0.04043)	(0.03792)	(0.04097)	(0.03994)
Pcoh_1	-0.19145	-0.35421	-0.18209	-0.17985	-0.28646	-0.43535**	-0.27254	-0.27386
	(0.21509)	(0.21662)	(0.21392)	(0.21279)	(0.22236)	(0.22147)	(0.22245)	(0.21997)
Pcoh_2	-0.06733	-0.20168	-0.06053	-0.05552	-0.15704	-0.27854	-0.14635	-0.14389
	(0.21082)	(0.21245)	(0.20955)	(0.20850)	(0.21688)	(0.21603)	(0.21691)	(0.21463)
Pcoh_3	-0.01918	-0.15618	-0.03265	-0.00665	-0.08720	-0.21392	-0.09937	-0.07415
	(0.21329)	(0.21469)	(0.21165)	(0.21102)	(0.21962)	(0.21856)	(0.21939)	(0.21741)
Pcoh_4	0.01095	-0.11597	0.00325	0.02602	-0.04140	-0.16064	-0.04720	-0.02512
	(0.21388)	(0.21551)	(0.21228)	(0.21176)	(0.22082)	(0.21988)	(0.22049)	(0.21882)
Pcoh_5	0.02626	-0.11882	0.02630	0.03759	-0.05025	-0.18462	-0.04802	-0.03820
	(0.21216)	(0.21452)	(0.21050)	(0.20965)	(0.21749)	(0.21733)	(0.21717)	(0.21497)
Pcoh_6	0.06173	-0.09930	0.05837	0.06576	-0.00872	-0.16001	-0.01034	-0.00470
	(0.21712)	(0.21905)	(0.21562)	(0.21472)	(0.22141)	(0.22105)	(0.22126)	(0.21906)
Pcoh_7	0.14885	-0.01508	0.14502	0.15978	0.08549	-0.07008	0.08315	0.09635
	(0.22018)	(0.22154)	(0.21859)	(0.21786)	(0.22408)	(0.22321)	(0.22387)	(0.22178)
Pcoh_8	0.18998	0.03467	0.18033	0.19938	0.11982	-0.02629	0.11157	0.12889
	(0.21627)	(0.21844)	(0.21465)	(0.21406)	(0.21972)	(0.21968)	(0.21946)	(0.21753)
Hrs_both	0.00113***	0.00101***	0.00111***	0.00111***				
	(0.00029)	(0.00028)	(0.00030)	(0.00029)				
D(Hrs_both)	-0.11005**	-0.08411*	-0.10818**	-0.10343**				
	(0.04821)	(0.04725)	(0.04739)	(0.04789)				
D(Lima)	0.10344***	0.11294***	0.09621***	0.08825***	0.09825***	0.11574***	0.09083***	0.08217***
	(0.01891)	(0.03368)	(0.01919)	(0.01939)	(0.02008)	(0.03461)	(0.02032)	(0.02060)
L(pop_urb_dist)		0.29940***				0.29774***		
		(0.07388)				(0.07671)		
L(pop_urb_dist)-sq		-0.01235***				-0.01237***		
		(0.00349)				(0.00362)		
pop_den_distx1000		-0.00319*				-0.00336*		
		(0.00166)				(0.00175)		
D(not_slum)		0.04262***				0.03953***		
		(0.01321)				(0.01368)		
D(mid_city)		-0.01952				-0.01748		
		(0.01478)				(0.01534)		
D(border)		0.07612***				0.08420***		
		(0.01971)				(0.02001)		
D(tongue)			0.06634***				0.07400***	

			(0.01665)				(0.01790)	
D(migrant)			0.02340*				0.02273	
			(0.01414)				(0.01458)	
D(social)			0.00367				0.00429	
			(0.01326)				(0.01382)	
Station_distx1000				-0.13595***				-0.14714***
				(0.03312)				(0.03520)
Muni_persx1000				-0.00385**				-0.00404***
				(0.00150)				(0.00146)
L(budget_dist_pc)				-0.00512				-0.00629
				(0.01055)				(0.01081)
D(sdp_dist)				-0.06518**				-0.06338*
				(0.03241)				(0.03401)
Constant	-0.54799	-1.85496***	-0.69753*	-0.53399	-0.69003*	-1.96362***	-0.85130**	-0.66087*
	(0.37043)	(0.48203)	(0.37728)	(0.36785)	(0.39088)	(0.50270)	(0.40095)	(0.38778)
Observations	2,142	2,142	2,142	2,142	2,142	2,142	2,142	2,142
LogL	-395.423	-307.291	-396.401	-387.124	-471.408	-370.356	-472.691	-460.325

w) Second-stage regression for IV model for informal Fon shares (GMM version)

	With labour supply				Without labour supply			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
θ	0.27767**	0.27026***	0.25814**	0.26993**	0.33881***	0.31355***	0.31574**	0.32661***
	(0.10804)	(0.10428)	(0.10971)	(0.10697)	(0.12227)	(0.11672)	(0.12415)	(0.12067)
L(expend_Fon)	0.11088***	0.06687**	0.12875***	0.11604***	0.13052***	0.08537**	0.14964***	0.13544***
	(0.02951)	(0.03095)	(0.03080)	(0.02937)	(0.03294)	(0.03435)	(0.03471)	(0.03280)
L(fam_size)	0.03383	0.04617	0.01365	0.02600	0.03044	0.04268	0.00805	0.02239
	(0.04373)	(0.04218)	(0.04473)	(0.04365)	(0.04448)	(0.04255)	(0.04581)	(0.04441)
D(mem_son)	-0.00231	0.00021	-0.00046	-0.00198	-0.00104	0.00046	0.00159	-0.00089
	(0.03505)	(0.03376)	(0.03532)	(0.03485)	(0.03575)	(0.03423)	(0.03614)	(0.03550)
D(mem_oth)	0.01189	0.01427	0.01612	0.01044	0.01011	0.01266	0.01524	0.00840
	(0.03286)	(0.03150)	(0.03277)	(0.03252)	(0.03336)	(0.03186)	(0.03327)	(0.03302)
P(men)	-0.08683**	-0.07336**	-0.08543**	-0.08748**	-0.08798**	-0.07562**	-0.08633**	-0.08857**
	(0.03872)	(0.03713)	(0.03898)	(0.03861)	(0.03988)	(0.03797)	(0.04022)	(0.03971)
D(gender_head)	-0.04472	-0.04299	-0.04264	-0.04696	-0.04765	-0.04449	-0.04475	-0.04942
	(0.04003)	(0.03759)	(0.04063)	(0.03951)	(0.04000)	(0.03730)	(0.04068)	(0.03952)
Pcoh_1	-0.20181	-0.35569*	-0.17950	-0.18798	-0.30039	-0.43921**	-0.27230	-0.28552
	(0.21467)	(0.21606)	(0.21377)	(0.21244)	(0.22101)	(0.22015)	(0.22176)	(0.21874)
Pcoh_2	-0.07607	-0.20328	-0.05819	-0.06205	-0.16953	-0.28553	-0.14562	-0.15396
	(0.21049)	(0.21192)	(0.20943)	(0.20822)	(0.21572)	(0.21481)	(0.21627)	(0.21354)
Pcoh_3	-0.02828	-0.15860	-0.03144	-0.01372	-0.10019	-0.22001	-0.10031	-0.08498
	(0.21297)	(0.21422)	(0.21154)	(0.21076)	(0.21858)	(0.21749)	(0.21884)	(0.21644)
Pcoh_4	0.00312	-0.11449	0.00335	0.02007	-0.05136	-0.16353	-0.04805	-0.03335
	(0.21368)	(0.21517)	(0.21225)	(0.21160)	(0.21984)	(0.21885)	(0.22000)	(0.21791)
Pcoh_5	0.01653	-0.11737	0.02595	0.02998	-0.06223	-0.18584	-0.04773	-0.04787
	(0.21193)	(0.21425)	(0.21045)	(0.20947)	(0.21651)	(0.21651)	(0.21668)	(0.21410)
Pcoh_6	0.05468	-0.09521	0.05764	0.06025	-0.01753	-0.15893	-0.01067	-0.01178
	(0.21696)	(0.21877)	(0.21559)	(0.21459)	(0.22057)	(0.22022)	(0.22082)	(0.21829)
Pcoh_7	0.14151	-0.01122	0.14501	0.15437	0.07741	-0.06851	0.08484	0.09013
	(0.21996)	(0.22120)	(0.21853)	(0.21768)	(0.22325)	(0.22234)	(0.22346)	(0.22103)
Pcoh_8	0.18256	0.03862	0.18047	0.19400	0.11135	-0.02533	0.11313	0.12261
	(0.21606)	(0.21810)	(0.21459)	(0.21389)	(0.21886)	(0.21882)	(0.21902)	(0.21675)
Hrs_both	0.00114***	0.00101***	0.00110***	0.00112***				
	(0.00029)	(0.00027)	(0.00029)	(0.00029)				

D(Hrs_both)	-0.11041**	-0.08375*	-0.10848**	-0.10362**				
	(0.04799)	(0.04698)	(0.04729)	(0.04769)				
D(Lima)	0.10222***	0.11356***	0.09561***	0.08694***	0.09725***	0.11596***	0.09059***	0.08123***
	(0.01878)	(0.03355)	(0.01912)	(0.01927)	(0.01976)	(0.03428)	(0.02010)	(0.02027)
L(pop_urb_dist)		0.29662***				0.29371***		
		(0.07356)				(0.07589)		
L(pop_urb_dist)-sq		-0.01220***				-0.01216***		
		(0.00347)				(0.00358)		
pop_den_distx1000		-0.00338**				-0.00355**		
		(0.00165)				(0.00172)		
D(not_slum)		0.04226***				0.03956***		
		(0.01315)				(0.01349)		
D(mid_city)		-0.01872				-0.01720		
		(0.01471)				(0.01517)		
D(border)		0.07613***				0.08430***		
		(0.01966)				(0.01986)		
D(tongue)			0.06686***				0.07346***	
			(0.01656)				(0.01771)	
D(migrant)			0.02361*				0.02236	
			(0.01408)				(0.01449)	
D(social)			0.00366				0.00416	
			(0.01322)				(0.01374)	
Station_distx1000				-0.13555***				-0.14631***
				(0.03308)				(0.03494)
Muni_persx1000				-0.00387***				-0.00408***
				(0.00150)				(0.00145)
L(budget_dist_pc)				-0.00535				-0.00617
				(0.01052)				(0.01075)
D(sdp_dist)				-0.06664**				-0.06558*
				(0.03238)				(0.03387)
Constant	-0.54059	-1.84488***	-0.69032*	-0.52713	-0.68160*	-1.93123***	-0.84429**	-0.65306*
	(0.36687)	(0.47943)	(0.37506)	(0.36458)	(0.38275)	(0.49480)	(0.39552)	(0.37994)
Observations	2,142	2,142	2,142	2,142	2,142	2,142	2,142	2,142
LogL	-396.032	-307.731	-394.368	-387.652	-473.564	-367.416	-470.667	-461.713