University of Sussex

The role of market imperfections in shaping rural household livelihoods: evidence from South Africa.

Stefania Lovo

Submitted for the degree of Doctor of Philosophy
Department of Economics
University of Sussex
September 2010

Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

Signature:

Stefania Lovo

UNIVERSITY OF SUSSEX

Stefania Lovo, Doctor of Philosophy

The role of market imperfections in shaping rural household livelihoods: evidence from South Africa.

Summary

This thesis analyses farm household behaviour and livelihood strategies in the presence of market imperfections. The first chapter uses a farm household model to explain the presence of three household groups determined on the basis of the labour regime adopted: small-scale peasants, self-cultivators and hiring-in households. A partial generalised ordered logit model is used to test the main predictions of the model using data from the 1997 Rural Survey. The results show that access to liquidity and market imperfections matter in the choice of the labour strategy and that liquidity constrained households are more likely to sell labour off-farm. The second chapter provides an analysis of household technical efficiency (TE) using using data on the KwaZulu Natal Province. The analysis is conducted at household-level and off-farm activities are considered as additional outputs of production. This approach better captures the jointness between farm and non-farm activities generated by the presence of market imperfections. An important source of liquidity for the household is the receipt of a pension. Its effect on household TE is identified exploiting the age eligibility criteria adopted by the pension program. The results show that access to liquidity and income diversification has positive effects on household TE. Finally the last chapter investigates the relationship between land and household welfare. It uses the year of arrival in the current location as an instrument for land access and size for households in the former homelands. This identification strategy relies on the argument that African households have been forcibly relocated to the homelands since the introduction of the Native Land Act in 1913. Because of increasing population pressure in the homelands, later arrivals were less likely to have access to land and to larger plots of land. Results show that access to land positively affects the welfare of rural household.

Acknowledgements

I would like to thank my supervisor Robert Eastwood for his meticulous reviews of my drafts. I am grateful for his useful comments and for helping me improve on my weaknesses. I would also like to thank Andy Newell, Federico Perali and Emilie Perge who kindly commented on previous versions of this thesis. I would like to express my gratitude to Merle and Michael Lipton for their helpful advises and for sharing with me their enthusiasm and knowledge with great kindness. I also appreciated the constructive observations provided by the participants of the PhD conference and sandwich seminars at the University of Sussex. I would like to thank all friends and colleagues that made my PhD life so much easier; a big thank goes to Gonzalo that, besides being the person with whom I have shared all the ups and downs of the PhD, has also been an always present (and I do mean always!) friend. I would like to thank my parents for their confidence and continuous support during the writing of this thesis. Finally a particular thank goes to Milton for staying by my side with understanding and lots of patience during these years.

Contents

List of Tables

List of Figures

Abbreviations

2SLS	Two Stage Least Squares
AE	Allocative Efficiency
CRS	Constant Returns to Scale
DEA	Data Envelopment Analysis
HSRC	Human Sciences Research Council
KIDS	KwaZulu Natal Income Dynamic Survey
KZN - DIHS	KwaZulu Natal Development Indicator Household Survey
IV	Instrumental variables
LATE	Local Average Treatment Effect
LRAD	Land Redistribution and Agricultural Development
OAP	Old Age Pension programme
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
SD	Standard Deviation
TE	Technical Efficiency
TLU	Tropical Livestock Unit
VRS	Variable Returns to Scale

Introduction

This thesis consists of three self-contained essays devoted to the analysis of rural household livelihoods in the presence of labour and credit market imperfections. Each chapter deals with a particular aspect of rural livelihoods, including household labour and resource allocation and asset accumulation. The focus on South Africa is motivated by the role that agriculture and rural development can play in addressing unemployment and poverty. Statistical records reveal that around 43% of the total population lives in rural areas. The high unemployment rate, around 23% at national level, is even higher in rural areas, particularly in the former homelands. Poverty, as well, is more severe as the majority of the poor (62 % in 1996 and 57% in 2001) are found in rural areas (Leibbrandt et al., 2006). In rural South Africa household livelihoods are characterised by a mix of land-based and off-farm activities. This is common to most developing countries were households engage in various activities within and outside agriculture (Dercon and Krishnan, 1996). These activities, together with government transfers, are very important sources of income for the rural economy and households are faced with the decision of allocating family labour and resources across alternative tasks and activities. Understanding what influences and constrains household decisions is important to design and target policies for rural development. The first two chapters of this thesis analyse household resource allocation and use in the presence of market imperfections, in particular, when households are liquidity constrained. Labour allocation decisions are analysed in the first chapter by dividing the population into three categories on the basis of the labour regime adopted. In the second chapter a measure of resource use intensity is obtained by computing an indicator of technical efficiency at household level. The last chapter focuses on the contribution of land to rural household livelihoods. Land-based activities can contribute to the overall well-being of the rural population by providing a return to uneducated family labour (Carter and May, 1999) and goods and services for home consumption. According to Lipton and Lipton (1993), small scale agriculture has the potential to address unemployment and poverty in South Africa by providing a labour-intensive solution to the country's large endowment of labour.

Each chapter develops a line of argument based on the standard microeconomic theory of the farm household. In this framework households' decisions are the result of an optimising behaviour where market imperfections can play a crucial role. In a perfect markets scenario, household decisions are based on market prices, and production choices are made independently of consumption preferences. In the presence of multiple market imperfections this "separation property" no longer holds and household labour allocation decisions depend on both the production and the consumption side of the household. Although all chapters depart from this consolidated theory of the farm household, each of them focuses on a different aspect of households' livelihoods and employs specific empirical strategies particular to the empirical questions to be addressed. The empirical analyses have been conducted using household level data. The first and third chapter use the 1997 Rural Survey, which collects information on rural households in the former homelands of South Africa. In addition, the third chapter employs the 1996 Kwazulu-Natal Development Indicators Household survey. Finally, the second chapter employs the third wave (2004) of the Kwazulu-Natal Income Dynamic Survey. Each chapter, therefore, has its own data section, theoretical framework and empirical methodology. The contributions of this thesis are presented separately for each chapter, together with a short description of the methodology and the results obtained.

In the first chapter of this thesis I extend the theoretical model proposed by Sadoulet et al. (1998) incorporating a liquidity constraint to describe the influence that a lack of pre-harvest liquidity has on labour allocation decisions. In the farm household model, labour market participation is explained by the relation between the on-farm marginal productivity of labour and the opportunity costs of family labour (also referred as shadow wage). The model is consistent with the argument that a liquidity constrained household attributes a higher value to off-farm opportunities reflecting the fact that extra income helps to ease the liquidity constraint. Households are classified into small-scale peasants (working both on and off farm), self cultivators (autarkic in labour) and hiring-in households. The empirical analysis tests the main predictions of the model and analyses the factors determining the membership in the three household categories. Existing empirical analyses have usually neglected the heterogeneity in which households are affected by market imperfections, and have therefore, assumed homogeneous shadow wages across households. In this chapter I employ a generalised order logit and a Brant test on threshold constancy in order to relax this assumption. This test has recently been introduced in the literature and no applications have been found in this context. A Brant test is a statistical tool, described in more detail in the chapter, that has also an economic interpretation. The test allows the researcher to choose the factors (variables) that affect household-specific exposure to market imperfections, i.e. household specific shadow wages. The results show that liquidity and market imperfections matter in the choice of the labour strategy. Liquidity constrained households are more likely to sell labour off-farm while access to information facilitates the hiring in of workers.

The second chapter analyses farm household technical efficiency adopting a household-level approach, which considers non-farming activities as additional outputs of production, to understand the constraints that prevent the optimal use of household resources. Understanding the reasons underlying poor performance in agriculture is, in fact, important for improving the role that agriculture plays in contributing to the livelihood of rural households. Because rural households engage in a wide range of activities in order to generate a livelihood, the standard analysis of technical efficiency usually conducted at farm level is extended to capture the linkages between farming and non-farming activities. This paper follows the work of Chavas et al. (2005) and expands his theoretical analysis by applying the concept of jointness in production, usually applied to multi-product agricultural production, to farming and non-farming activities to show that in the presence of market imperfections a household-level analysis of technical efficiency is more appropriate than a farm-level analysis. The analysis focus on the impact of liquidity constraints on household behaviour. This is done by considering the effect of the pension transfer provided by the South African Old Age Pension Program (OAP) which is expected to cause a substantial improvement in the households' liquidity position. This chapter contributes to the literature by adding to the current debate on the effects of the South African OAP Program on household behaviour and by providing an identification strategy for the liquidity effect on household technical efficiency. The liquidity effect is identified by exploiting the age eligibility criteria adopted by the South African OAP Program, which reaches all women over age 60 and men over age 65. Pension eligibility is used instead of actual pension receipt and several robustness checks are conducted in order to examine the presence of potential confounding effects between the eligibility indicator and age trends or differences in background. The results show that access to liquidity and income diversification have positive effects on household technical efficiency, suggesting that institutional reforms to improve access to labour and credit markets can allow a more efficient use of farm household resources.

The last chapter focuses on the relationship between land endowments and household welfare in the former homelands. The homelands or bantustans were designated reserves were the African population was segregated during the apartheid era. Given the available data, household welfare is measured using an asset index obtained by applying principal component analysis. Although the economic theory of the farm household provides support for a positive relationship between land and household welfare, little empirical evidence is available mainly due to the difficulties in identifying the causal relationship between land and household welfare, given the non-random allocation of land and the lack of suitable instruments. This chapter contributes to the literature by identifying the effects of land endowments on household welfare exploiting historical data on migration to the former homelands. The identification strategy relies on the fact that, since the introduction of the Native Land Act in 1913, African households have been forcibly relocated to the homelands. Movements within the homelands can also be largely explained by government 'betterment planning' for the reorganisation of the territory in the reserves. The year of arrival in the current location is used as an instrument for land endowments since later incomers were less likely to have access to land and to larger plots of land, given the increasing population pressure in these areas. The year of arrival is expected to be independent of households' unobserved ability to generate welfare since the pace and timing of relocations are probably unrelated to unobserved household-specific characteristics. Results show a positive and large effect of land access on household welfare. Moreover, because the homelands are relatively more disadvantaged and less fertile areas, these results are likely to provide a lower bound for the positive effects of land on household welfare.

Chapter 1

Market imperfections, liquidity and farm household labour allocation

1.1 Introduction

Household endowments and market imperfections shape the organisation of agricultural production and lead to different production regimes within rural farm households. Using data on the rural South Africa population, landed households have been categorised into three distinct groups on the basis of the labour regime adopted: small-scale peasants (sellers of labour), self-cultivators (self-sufficient in labour) and hiring-in households (buyers of labour). The labour regime adopted by the household is the result of an optimising behaviour in an imperfect market scenario and is an important indicator of household response to market imperfections. In rural South Africa household livelihoods are characterised by a mix of land-based and off-farm activities. Only few households have access to formal credit, and the decision of allocating family labour across alternative tasks and activities is likely to be influenced not only by endowments and comparative advantages but also by the lack and the search for liquidity.

The categorisation of households into homogenous groups is a useful strategy to understand the determinants and the consequences of particular household decisions. Dercon and Krishnan (1996), for example, consider five household groups defined on the basis of their income portfolio, to investigate the constraints that prevent households from engaging in particular activities. Empirical analyses conducted using data on rural South Africa also suggest the presence of quite distinct household categories. Carter and May (1999) identify eight classes on the basis of the livelihood strategies undertaken by each group. Such categorisation results from not only differences in tangible (land and labour) and intangible assets (welfare rights, social reciprocity), but also from the ability to effectively exploit such endowments. In Eastwood et al. (2006), the authors define categories of households on the basis of their specialisation by income source and identify three groups within rural households in a former homeland area of the Limpopo Province (South Africa): factor-reliant, migration-dependent and pension-dependent households. In this chapter I also intend to exploit the categorisation of farm households into homogenous groups to obtain useful insights on what influences household labour allocation decisions.

The membership in the three categories above mentioned is determined on the basis of the labor regime adopted and is explained using a theoretical framework built on the classical agricultural household model literature (Singh et al., 1986). The theoretical model adopted in this chapter consider the presence of labour and credit market imperfections. In a perfect-markets scenario, household decisions are based on market prices, and production choices are taken independently from consumption preferences, a result commonly called 'separation property'. In the presence of multiple market imperfections this property no longer holds and household labour allocation decisions depend on both the production and the consumption side of the household[1]. A farm household model is,

[1]The separation property may hold also in the absence of one market (for example the land market).

therefore, necessary to correctly model household behaviour. Eswaran and Kotwal (1986), following the endogenous class formation framework of Roemer (1982) who first formalised household labour regime adoptions, use a farm household model and consider credit and labour market imperfections to analyse membership in different modes of cultivation. In their model the amount of credit received depends on land ownership. This framework, however, cannot be applied to rural South Africa where land does not serve as collateral and the agricultural credit sector is underdeveloped (Fenwick and Lyne, 1999). In a similar fashion, Sadoulet et al. (1998) suggest a farm household model that incorporates transaction costs and fixed land endowments and leads to the identification of the three regimes above mentioned. Departing from this model I incorporate an additional liquidity constraint to show that the lack of pre-harvest liquidity also affects the labour strategy adopted. In the farm household model, labour market participation is explained by the relationship between the on-farm marginal productivity of labour and the opportunity costs of family labour for buyers and sellers. The opportunity costs are given by the off-farm and hiring-in effective wages respectively, which incorporate transaction costs, augmented by an endogenous mark up indicating the liquidity position of the household. The model indicates that a liquidity constrained household attributes a higher value to off-farm opportunities reflecting the fact that extra incomes help to ease the liquidity constraint (Barrett et al., 2001). An opposite effect is expect on the decision of hiring in workers.

A partial generalised ordered logit is used to test the main predictions of the model and to analyse the factors determining the membership in the three household categories. Only few empirical analyses have attempted to model the heterogeneity in which households are affected by market imperfections (Cafiero et al., 2004). In this chapter, I employ a gener-

The presence of multiple market imperfections ensures that the model is non-separable (Bardhan and Udry, 1999).

alised order logit and a Brant test on threshold constancy in order to allow for household heterogeneous exposure to market imperfections. A Brant test is a statistical tool, which function is described in the next sections, that has a useful economic interpretation. The test allows to distinguish the variables that affect household-specific exposure to market imperfections from those determining the on farm marginal productivity of labour. The results of the generalised order logit model show that liquidity and market imperfections matter in the choice of the labor strategy adopted. Liquidity is found to affect the choice of hiring in and out labour, inducing liquidity constrained household to sell labour off-farm and preventing the hiring in of labour.

The chapter proceeds as follows: section 2 introduces the logic behind household category formation while section 3 reports, in more details, the theoretical model in which the three household groups are analysed separately. Section 4 presents the empirical specification and the data used, in section 5 the results are shown and discussed; finally, section 6 concludes.

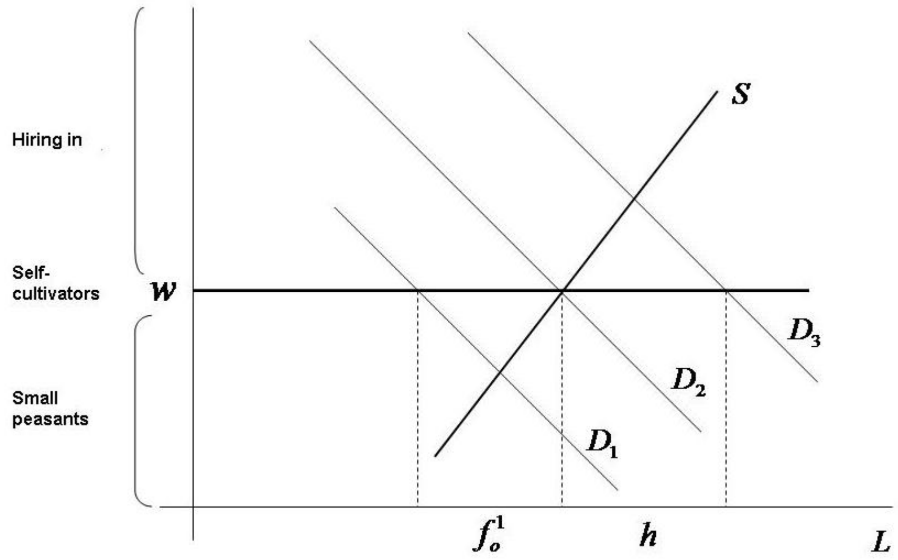
1.2 Explaining household membership in different labour regime categories

This section introduces the rationale behind the existence of distinct household categories based on the labour regime adopted. Households are categorised into small-scale peasants (sellers of labour), self-cultivators (autarkic in labour) and hiring-in households (buyers of labour)[2]. Labour market participation is explained by the relationship between the on-farm marginal productivity of family labour and its opportunity cost off-farm. In the presence of a perfectly competitive labour market, as represented in figure 1.1, household

[2]This analysis does not separately consider crop production and livestock rearing, the latter involving higher entry constraints, as it is instead done in Dercon and Krishnan (1996). Moreover off-farm self-employed activities are also excluded from the analysis given the lack of reliable data.

decisions are based on market wages, w, and farm technology which determines the on-farm demand for labour.

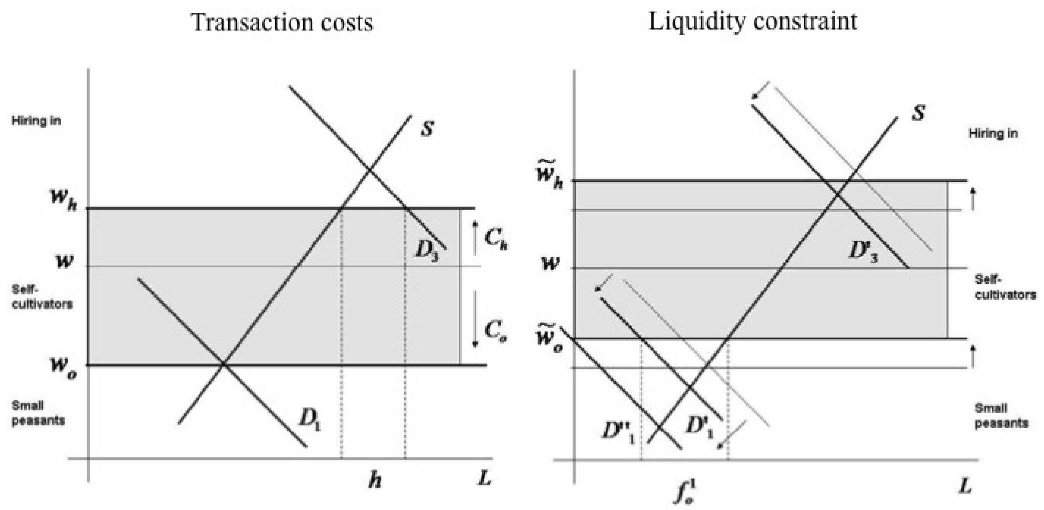
Figure 1.1: Labour market participation with perfect labour market



The graph depicts the labour supply curve (S) and three different patterns of demand (D) for labour. The intersection between the curves determines the on-farm marginal productivity of labour in the case of self-sufficiency in labour (autarky). When the market wage lies above the intersection, as in the case of a low demand for labour D_1, the household has an incentive to sell labour off-farm, and the quantity sold to the market is denoted by f_o^1. The opposite occurs when the off-farm wage exceeds the autarkic on-farm marginal productivity of labour (higher demand for labour, D_3); in this case the household will find it profitable to replace family members with the less costly hired labour, h. When on-farm labour productivity equals the off-farm wage, the household will opt for self-sufficiency. It is worth noting, however, that in the case of a perfect labour market, farm households will likely either sell or hire labour.

The inclusion of labour market transaction costs determines a wage differential between the lower wage the household can gain off-farm and the higher costs of hiring-in labour, and is depicted by the shifts in the wage line (figure 1.2, left panel). Transaction costs, such as search and transport costs (C_o), reduce the off-farm *effective* wage, w_o, while increase the *effective* cost of hiring in labour, w_h, due to supervision and other costs, C_h. Because transactions costs are known to be household specific, different households are expected to face different wage bands. Transaction costs have a negative effect on the quantities of labour hired in and out by the household, increasing the probability of being self-cultivators. For buyers and sellers of labour, the cost of labour equals the effective off-farm and hiring-in wages respectively, while in the case of autarky the cost is endogenously determined and lies within the wage band.

Figure 1.2: Household categories under market imperfections



When households face a liquidity constraint, the effects are observed in both the on-farm marginal productivity of labour and the off-farm opportunity costs (figure 1.2, right panel). A liquidity shortage restricts the purchasing of non-labour inputs; on-farm labour demand falls and consequently lowers the marginal productivity of labour. At the same time, households attribute an additional value to off-farm earnings, reflecting the fact that extra incomes help to ease the liquidity constraint. The higher opportunity cost of working on-farm, $\widetilde{w_o}$, is given by the effective off-farm wage and the liquidity premium. On the other hand, the lack of liquidity prevents households from buying labour and is represented by a higher cost of hired labour, $\widetilde{w_h}$. Figure 1.2 depicts two illustrative situations. The first case refers to a lower demand household (D_1): a cash shortage forces the household out of self-sufficiency and could even lead to the abandonment of agricultural activities ($D_1^{''}$). The second example, in contrast, corresponds to a higher demand household D_3 where the binding liquidity constraint prevents the household from hiring in labour causing a switch to self-sufficiency. In both cases the household shifts down the hierarchy; the overall effect of a cash shortage is, therefore, an increase in the likelihood of being small-scale peasants, for lower demand households, and a decrease in the likelihood of being in the hiring-in category for the higher demand households.

The above discussion has shown how household characteristics influence the demand and supply of family labour and its opportunity cost. The location of the demand and supply curve is determined by household technology and endowments. Transaction costs, which depend both on household-specific characteristics and other factors exogenous to the household, determine the effective market wages. Changes in household endowments, given their effect on the liquidity position of the household, can cause the wage band to shift and affect, at the same time, the marginal productivity of labour. Depending on the initial conditions, changes in household characteristics can cause a switch from small-scale peasants to self-cultivators or from self-cultivators to the hiring-in category, and vice versa, through their effects on the marginal productivity of labour and on the thresholds: $\widetilde{w_o}$ and $\widetilde{w_h}$. These two effects, however, cannot be separated a priori; this issue is addressed during the empirical estimation with the support of a Brant test on threshold constancy.

1.3 A farm household model to explain labour market participation

The model presented in this section provides the theoretical foundations for the emergence of the above-mentioned household categories in a imperfect markets scenario. In particular it considers imperfections in the labour market, allowing for the presence of transaction costs which are directly translated into the *effective cost* of labour, and a pre-harvest liquidity constraint. The model departs from that presented in Sadoulet et al. (1998) and is extended via the inclusion of non-labour inputs, as an additional component of the farm production function, and a pre-harvest cash constraint. The effective cost of hiring (unskilled) labour (h) is given by the market wage plus search and supervision costs and is defined as w_h. The effective off-farm wage includes search and other transaction costs and differs between unskilled (w_o^1) and skilled labour (w_o^2). The imperfections in the labour market are therefore translated into the following relation $w_o^1 < w_h < w_o^2$. Family labour is also allocated to on-farm activities (skilled and unskilled: f_q^1, f_q^2) and to leisure (f_l^1, f_l^2). Unskilled family labour is assumed to be homogenous implying that all members possess the same on-farm productivity. I also employ the standard assumptions of increasing, strictly concave and continuously differentiable utility $u()$ and production $q()$ functions to ensure an unique solution to the model. Production inputs, x, A and L (where $L = f_q^1 + f_q^2 + h$) are assumed to be complementary as well as income and leisure, $u_{1y}, u_{2y} > 0$. As reported by several authors[3], in rural South Africa income from non-agricultural sources, such as wage employment, is important in providing working capital for the purchase of seeds, fertilisers, and other production inputs; to take this into account I include a liquidity constraint on pre-harvest transactions where α is the fraction of hiring

[3]The complete list of study suggesting such conclusions is reported in Van Zyl et al. (1995).

costs and household earnings that occurres before the harvesting ($\alpha > 0$):

$$p_x x + \alpha w_h h + K \leq \alpha(w_o^1 f_o^1 + w_o^2 f_o^2) + S,$$

where x represents non-labour production inputs which price is p_x, K are fixed setup costs and S is the pre-harvest amount of exogenous transfers. The household maximises utility, which is a function of leisure and income, y. The maximisation is subject to the above liquidity constraint together with additional non-negativity constraints on labour demand and supply, and on the amount of inputs and leisure time:

$$\max_{f_o^1, f_o^2, f_q^1, f_q^2, h, x} U(f_l^1, f_l^2, y),$$

where $y = pq(A, h + f_q^1 + f_q^2, x) - w_h h - w_x x - K + w_o^1 f_o^1 + w_o^2 f_o^2 + T$ and $f_l^1 = f^1 - f_q^1 - f_o^1$, $f_l^2 = f^2 - f_l^2 - f_o^2$. The lagrangean function for this problem and the respective first order conditions are reported below.

$$\begin{aligned} L = {} & U(f_l^1, f_l^2, pq(A, h + f_q^1 + f_q^2, x) - w_h h - w_x x - K + w_o^1 f_o^1 + w_o^2 f_o^2 + T) \\ & + \lambda(\alpha(w_o^1 f_o^1 + w_o^2 f_o^2) + S - p_x x - \alpha(w_h h) - K) + \mu_k^n f_k^n + \mu_h h + \mu_x x, \end{aligned}$$

with $k = q, o, l$ and $n = 1, 2$.

$$
\begin{array}{lll}
1a) & h: & u_y(pq_L - w_h) + \mu_h - \lambda\alpha w_h = 0, \quad h \geq 0, \quad \mu_h h = 0, \\
2a) & f_q^1: & u_y pq_L - u_1 + \mu_q^1 - \mu_l^1 = 0, \qquad f_q^1 \geq 0, \quad \mu_i^1 f_q^1 = 0, \\
3a) & f_q^2: & u_y pq_L - u_2 + \mu_q^2 - \mu_l^2 = 0, \qquad f_q^2 \geq 0, \quad \mu_q^2 f_i^2 = 0, \\
4a) & f_o^1: & u_y w_o^1 - u_1 + \mu_o^1 - \mu_l^1 + \lambda\alpha w_o^1 = 0, \quad f_o^1 \geq 0, \quad \mu_o^1 f_o^1 = 0, \\
5a) & f_o^2: & u_y w_o^2 - u_2 + \mu_o^2 - \mu_l^2 + \lambda\alpha w_o^2 = 0, \quad f_o^2 \geq 0, \quad \mu_o^2 f_o^2 = 0, \\
6a) & x: & u_y(pq_x - p_x) + \mu_x - \lambda\alpha p_x = 0, \qquad x \geq 0, \quad \mu_x x = 0, \\
7a) & \lambda: & \alpha(w_o^1 f_o^1 + w_o^2 f_o^2) + S - p_x x - \alpha w_h h - K \geq 0, \quad \lambda \geq 0, \\
 & & \qquad \lambda(\alpha(w_o^1 f_o^1 + w_o^2 f_o^2) + S - p_x x - \alpha w_h h - K) = 0, \\
8a) & \mu_k^n: & f_k^n \geq 0, \quad \mu_k^n \geq 0, \quad \mu_k^n f_k^n = 0, \qquad n = 1, 2,
\end{array}
$$

where λ is the marginal value of liquidity, T are exogenous transfers mainly represented by pensions and remittances and A is the fixed amount of assets owned by the household, which includes land. I use the notation q_L to indicate the marginal productivity of labour while u_1 and u_2 refers to the marginal utility of leisure time for unskilled and skilled members respectively. Since land in South Africa is mainly allocated by local or tribal authorities which provides only use rights, land is assumed to be fixed and cannot serve as collateral for credit (Fenwick and Lyne, 1999).

The inclusion of a liquidity constraint does not affect farm labour allocation of skilled household members. As in Sadoulet et al. (1998), skilled members do not work on-farm. This can be directly derived from the initial assumption $w_h < w_o^2$, which implies that there are no incentives for the household to employ skilled members on-farm since the foregone wage is higher than the cost of hiring labour[4]. The model does not explain the

[4]In the model, the initial assumption $w_h < w_o^2$ implies a positive complementary slack variable associated with on-farm skilled labour f_q^2 which expression can be obtained substituting equations 1a and 5a

presence of households that are both hiring in and out unskilled labour. Given the initial assumption, $w_o^1 < w_h$, the cost of hiring-in labour exceeds the wage unskilled members can gain off-farm. The household, therefore, has an incentive to replace hired workers with family labour[5]. Moreover, given the small percentage (about 4%) of households both selling and buying labour in our sample, this category will not be considered in this study.

Following the same assumptions of Sadoulet et al. (1998) I focus on unskilled family labour only, assuming that skilled members work a fixed number of hours off-farm, $f_o^2 = kf^2$. The model can be simplified and reduced to the following:

$$\max_{h,x,f_o^1,f_q^1,} U(f_l^1, f^2(1-k), y),$$
$$s.t \quad p_x x + \alpha w_h h + K \leq \alpha(w_o^1 f_o^1 + w_o^2 k f^2) + S,$$

where $y = pq(A, h + f_q^1, x) - w_h h - p_x x - K + w_o^1 f_o^1 + w_o^2 k f^2 + T$ and $f_l^1 = f^1 - f_q^1 - f_o^1$. The first order conditions for this problem are reported below:

$$1b) \quad h: \quad u_y(pq_L - w_h) + \mu_h - \lambda \alpha w_h = 0, \quad h \geq 0, \quad \mu_h h = 0,$$

$$2b) \quad f_q^1: \quad u_y pq_L - u_1 + \mu_q^1 - \mu_l^1 = 0, \quad f_i^1 \geq 0, \qquad \mu_q^1 f_i^1 = 0,$$

$$3b) \quad f_o^1: \quad u_y w_o^1 - u_1 + \mu_o^1 - \mu_l^1 + \lambda \alpha w_o^1 = 0, \quad f_o^1 \geq 0, \quad \mu_o^1 f_o^1 = 0,$$

$$4b) \quad x: \quad u_y(pq_x - p_x) + \mu_x - \lambda \alpha p_x = 0, \quad x \geq 0, \quad \mu_x x = 0,$$

$$5b) \quad \lambda: \quad \alpha(w_o^1 f_o^1 + w_o^2 f_o^2) + S - \alpha(p_x x - w_h h) - K \geq 0, \quad \lambda \geq 0,$$

$$\lambda(\alpha(w_o^1 f_o^1 + w_o^2 f_o^2) + S - \alpha(p_x x - w_h h) - K) = 0,$$

$$6b) \quad \mu_k^1: \quad f_k^1 \geq 0, \quad \mu_k^1 \geq 0, \quad \mu_k^1 f_k^1 = 0.$$

into 3a to obtain $\mu_q^2 = u_y(w_o^2 - w_h) + \lambda(w_o^2 - w_h) + \mu_o^2 + \mu_h > 0$. If $\mu_q^2 > 0$, by complementary slackness, $f_q^2 = 0$.

[5] The same intuition can be derived substituting equations 1a and 2a into 4a and considering households with positive hired labour ($\mu_h = 0$) to obtain $\mu_o^1 = u_y(w_h - w_o^1) + \mu_q^1 + \lambda(w_h - w_o^1) \geq 0$. A positive complementary slack variable implies zero off-farm unskilled labour. At the same time, considering households whose members work off-farm, no hired labour is admitted.

Below the three household groups and the landed workers category are analysed separately.

Landed workers. A lower bound on working capital (including land, human capital and agricultural capital) is specified here so that households below this threshold will consider cultivation unprofitable. As reported in Eswaran and Kotwal (1986), a household will engage in farming activities only if its optimal utility in agriculture, U_a^*, exceeds the utility of being a pure worker, U_w^*. A_0 represents a set of assets such that, when $A > A_0$:

$$U_a^*(f^1, f^2, w_o, w_h, p_x, T, K, A,) > U_w^*(f^1, f^2, T, w_o),$$

where:

$$U_w^*(f^1, f^2, T, w_o) = \max_{f_o^1} U(f_l^1, f^2(1-k), w_o^1 f_o^1 + w_2^o k f^2 + T).$$

The presence of fixed set-up costs, K, and the lack of liquidity to cover labour and non-labour inputs costs can prevent households from engaging in agricultural activities. These conditions determine the emergence of a category of landed workers. For households with $A > A_0$, below I analyse the characteristics of each category separately.

Self-cultivators. This category includes households self-sufficient in labour: households that do not sell or hire unskilled labour. The model reduces to the following four equations:

$$1c) \quad pq_L = u_1/u_y = w^*,$$

$$2c) \quad pq_x = p_x(1 + \lambda\alpha/u_y) = \widetilde{p_x},$$

$$3c) \quad \alpha w_o^2 f_o^2 + S - p_x x - K \geq 0, \quad \lambda \geq 0,$$

$$\lambda(\alpha w_o^2 f_o^2 + S - \alpha p_x x - K) = 0,$$

$$4c) \quad y = pq(A, f_q^1, x) - p_x x + w_o^2 k f^2 + T - K.$$

The marginal productivity of labour in the autarkic case, from now w^*, is a function of household characteristics and technology and will be used as a benchmark for the identification of the other categories. When the liquidity constraint is binding, the price of inputs ($\widetilde{p}_x$) is given by the market price and an endogenous markup (λ/u_y) representing the marginal utility of liquidity. The above system of equations defines w^* as a function of all the exogenous variable:

$$w^* = w^*(A, f^1, f^2, T, K, w_o^2, \widetilde{p}_x, \alpha) = w^*(A, f^1, f^2, T, w_o^2, K, p_x, S, \alpha). \tag{1.1}$$

Since $w_h > w_o$, the marginal productivity of labour lies between the two thresholds, $w_o^1 < w^* < w_h$, implying zero incentives to buy and sell labour.

Small-scale peasants. Households belonging to this category allocate labour both on- and off-farm while no hired labour is required. Considering equations 2b and 3b, appropriately adapted to this specific case ($\mu_q^1 = 0$ and $\mu_o^1 = 0$), I obtain the following expression:

$$pq_L = \frac{u_1 + \mu_l^1}{u_y} = w_o^1 \left(1 + \frac{\lambda\alpha}{u_y}\right) = \widetilde{w}_o^1. \tag{1.2}$$

The on-farm marginal productivity of labour equals the off-farm shadow wage ($\widetilde{w}_o^1$), which is given by the effective price of labour augmented by an endogenous markup caused by the presence of a liquidity constraint (liquidity premium). Because the marginal productivity of labour is a decreasing function of labour, the marginal productivity of labour for small-scale peasants (pq_L) is expected to be higher than in the case of autarky (w^*). The condition determining the membership in this category is, therefore, the following:

$$w^* < \widetilde{w}_o^1.$$

The household will sell labour if the opportunity cost given by the effective off-farm wage and the liquidity premium is greater than the on-farm remuneration they would get if all family members worked on the household's farm (w^*). The presence of a binding liquidity constraint increases the opportunity cost of being a self-cultivator, thereby shifting the threshold delimiting this category (figure 1.2, right panel) and inducing the family to sell labour.

Hiring-in households. Households belonging to this category hire in labour in addition to their own family members. Members do not work off-farm and, considering equations 1b and 2b with the opportune adjustments ($\mu_h = 0$ and $\mu_q^1 = 0$), it follows that:

$$pq_L = \frac{u_1 + \mu_l^1}{u_y} = w_h \left(1 + \frac{\lambda\alpha}{u_y}\right) = \widetilde{w}_h. \tag{1.3}$$

Because the marginal productivity of labour is a decreasing function of labour, then $w^* > \widetilde{w}_h$. When the household buys labour, family members can be substituted by hired workers; the opportunity cost of family labour, $\widetilde{w}_h$, is given by the effective wage plus a liquidity markup. If w^* is greater than the hiring-in shadow wage ($\widetilde{w}_h$), the household will have an incentive to hire in labour. The presence of a binding liquidity constraint raises the opportunity cost and discourages the hiring of workers; the upper threshold in figure 1.2 (right panel) shifts upward.

Table 1.1: Results of the comparative statics exercise

Effect	Sign	Effect	Sign
$\frac{\partial w^*}{\partial f^1}$	$-$	$\frac{\partial w^*}{\partial f^2}$	$\pm$
$\frac{\partial w^*}{\partial T}$	$+$	$\frac{\partial w^*}{\partial p_x}$	$-$
$\frac{\partial w^*}{\partial A}$	$+$		

The household behaviour illustrated in figure 1.2 can be also depicted through a comparative statics exercise which, in addition, identifies the sign of the effects and offers a set of conditions that can be empirically tested. Table 1.1 summarises the expected signs derived using the procedure reported in the Appendix 1. Larger endowments of land and higher income transfers raises the marginal productivity of labour thereby lowering the probability of selling labour and increasing the probability of hiring in agricultural workers, *ceteris paribus*. Larger endowments of unskilled family labour lower the marginal productivity of labour causing a switch from the hiring-in category to the self-cultivators and to the small-scale peasant categories (depending on the initial conditions). Finally, the effect of skilled labour is ambiguous[6]. The predictions reported in table 1.1 will be tested using a partial generalised ordered logit as described in the following section.

1.4 Data and empirical specification

Membership of one of the three household categories is determined by the endogenous marginal productivity of labour, w^*, function of the household characteristics and technology, and the shadow wage bands, which are influenced by household specific transaction costs and liquidity position. The three household groups can be ordered in accordance with the latent marginal productivity of labour. The probability of belonging to one of the three categories can be estimated considering the model reported below:

$$\begin{aligned} P(d_j = 1) &= P(i \in \text{small-scale peasants}) = P(w_i^* + v_i < w_{oi}^1(C_{oi}, \lambda_i), \\ P(d_j = 2) &= P(i \in \text{self-sufficients}) = P(w_{oi}^1(C_{oi}, \lambda_i) < w_i^* + v_i < w_{hi}(C_{hi}, \lambda_i), \\ P(d_j = 3) &= P(i \in \text{hiring-in households}) = P(w_i^* + v_i > w_{hi}(C_{hi}, \lambda_i), \end{aligned}$$

where i indicates the i-th household and j are the categories. C_{hi} and C_{oi} represent the determinants of transaction costs associated with hiring labour in and out, λ_i indicates the

[6]For a discussion of this ambiguity see the Appendix 1.

household-specific liquidity status and v_i is the error term. Considering the first condition, for example, the probability of belonging to the first category (small-scale peasants) corresponds to the probability that the endogenous marginal productivity of labour (w^*) is lower that the off-farm shadow wage. The model will be tested using a partial generalised ordered logit which allows for household specific thresholds. This procedure relaxes the assumption underlying the classical ordered logit model in which the relationship between the explanatory variables and the response does not vary across categories (parallel equations assumption). The standard ordered logit model estimates common thresholds, which corresponds to estimate similar effective shadow wages across categories (as it is done in Sadoulet et al. (1998)). It is widely recognised in the literature (de Janvry et al., 1991) that households are not affected by market imperfections with the same intensity; endowments and environmental, social and cultural factors affect their specific exposure to market imperfections. Households with different characteristics are therefore expected to face different transaction costs, i.e. different effective market wages. Moreover, thresholds are also affected by the household-specific liquidity position. In this chapter I use a partial generalised ordered logit, which allows for household specific thresholds and helps relaxing the assumption of constant shadow wages across households.

The marginal productivity of labour in autarky, w^*, namely the latent variable of the model, is assumed to be a linear function[7] of the shadow price of inputs $\widetilde{p}_i$ and of those household characteristics, X_i^w that have been identified as determinants of the latent variable by the Brant test:

$$w_i^* = X_i^w \beta_w + \widetilde{p}_i \theta + \varepsilon_{wi}, \tag{1.4}$$

where β_w is a vector of coefficients and the error ε_w has a standard logistic distribution.

[7]The same linear approximation has been used also in Bedi and Tunali (2005).

The procedure undelying the Brant test is described in more detail below. The shadow price of inputs, such as seeds, fertilisers and chemicals can be also linearly approximated by:

$$\widetilde{p}_i = X_i^p \beta_p + \varepsilon_{pi}, \tag{1.5}$$

where the explanatory variables, X_i^p, are representative of both the demand (for example household technology and endowments) and the supply (for example market prices) of inputs. Given the relationship reported in equation 1.1, equation 1.5 can be substituted into 1.4 to obtain the reduced form expression for the shadow wage:

$$w_i^* = X_i^w \beta_w + (X_i^p \beta_p + \varepsilon_{pi})\theta + \varepsilon_{wi}.$$

Collecting terms and setting the intercept to zero, the expression reduces to the following:

$$w^* = \mathbf{X}\beta + v.$$

The two thresholds are assumed to be a linear function of the household specific determinants of transaction costs and liquidity status:

$$w_n = \delta_0 + \mathbf{Z}\delta_n + \varphi_n,$$

where $n = o, h$, which indicates the off-farm and hiring-in wage respectively.

When a variable affects both the thresholds and the latent marginal productivity of labour, as in the case of liquidity related variables, the two effects cannot generally be separated. This identification problem can be solved including the variable either in the **X** or in the **Z** set of regressors; this does not imply any loss of generality since it does not matter whether the liquidity effect is interpreted as a labour demand or a shadow wage effect on the probability of hiring labour in or out. Departing from a standard ordered logit, a threshold constancy test (Brant test[8]) is used to determine which variables have to be considered as determinants of thresholds. In the case of a statistically significant test statistic, the constancy assumption has to be rejected, and the variables must be included in the threshold equations. Once the variables have been allocated either to the threshold or to the latent regression, a partial generalised ordered logit is estimated to characterise membership in the three household categories. The log likelihood function I estimate is therefore the following:

$$\begin{aligned} l_i(\beta,\delta) &= 1[d_j=1]\log[\Lambda(w_o-\mathbf{X}\beta-\mathbf{Z}\delta_o)]+1[d_j=2]\log[\Lambda(w_h-\mathbf{X}\beta-\mathbf{Z}\delta_h) \\ &\quad -\Lambda(w_o-\mathbf{X}\beta-\mathbf{Z}\delta_o)]+1[d_j=3]\log[1-\Lambda(w_h-\mathbf{X}\beta-\mathbf{Z}\delta_h)]. \end{aligned} \tag{1.6}$$

where β is the vector of coefficients satisfying the parallel regression assumption and δ_n are the vector of coefficient that vary across household categories[9].

1.4.1 The partial generalised ordered logit and the Brant test

In this section the procedure employed to test the main predictions of the theoretical model is presented in details. It is useful to start from a standard order logit specification which assumes constant thresholds and can be summarised as follows:

$$P(d_j=1)=P(w_i^*<w_o),$$

$$P(d_j=2)=P(w_o<w_i^*<w_h),$$

$$P(d_j=3)=P(w_i^*>w_h).$$

[8]An alternative method to test threshold constancy is presented in Pudney and Shield (2000).
[9]The model has been estimated using the stata command gologit2 (Williams, 2006).

where $w_i^* = \mathbf{M}\beta + v$ is the latent regression and $\mathbf{M}$ is a set of households characteristics and other factors exogenous to the household. According to the theoretical model the thresholds correspond to the shadow wages, where $\widetilde{w_o}$ and $\widetilde{w_h}$ are the hiring out and in wages, respectively. In this case they are considered constant across households. The loglikelihood function is the following:

$$\begin{aligned} l_i(\beta) \;=\; & 1[d_j = 1] \log[\Lambda(w_o - \mathbf{M}\beta)] + 1[d_j = 2] \log[\Lambda(w_h - \mathbf{M}\beta) - \\ & \Lambda(w_o - \mathbf{M}\beta)] + 1[d_j = 3] \log[1 - \Lambda(w_h - \mathbf{M}\beta)]. \end{aligned}$$

In this model a unique vector of β coefficients is estimated and, considering for simplicity only one independent variable, m, the cumulative probability outcomes are:

$$P(d \leq 1|m) = F(w_o - \beta m),$$

$$P(d \leq 2|m) = F(w_h - \beta m),$$

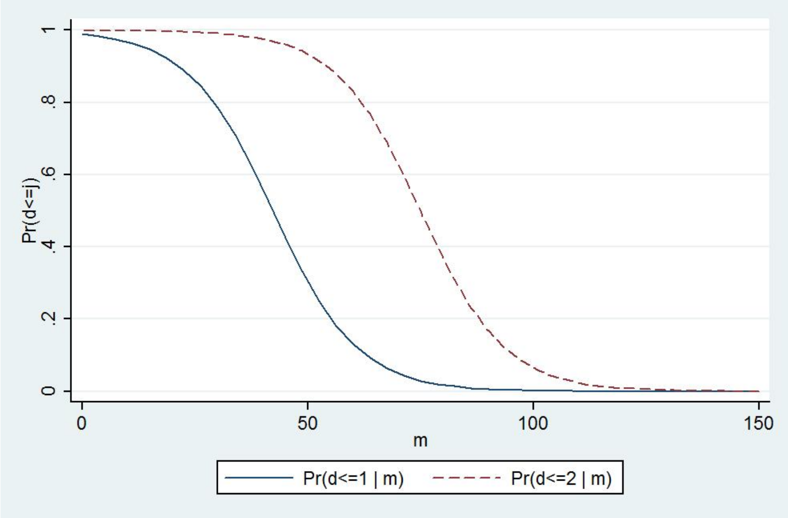
which are represented in figure 1.3[10]. The two curves are parallel as a consequence of equal coefficients across categories implied by the parallel regression assumption (or proportional odds assumption in the case of ordered logit) underlining standard ordinal models.

It is now possible to introduce a new feature in the model. Transactions costs and liquidity position are household specific factors. Therefore, the thresholds are expected to differ across households and across categories. This can be implemented using a generalised ordered logit which allows for heterogenous thresholds. Assuming, for the moment, that the above mentioned set of characteristics, $\mathbf{M}$, also affect the thresholds:

$$w_n = \delta_0 + \mathbf{M}\delta_n + \varphi_n,$$

[10] This graph has been taken from Long and Freese (2003) with opportune modifications.

Figure 1.3: Parallel regression assumption



where $n = o, h$, then the log likelihood functions of the generalised ordered logit model is:

$$\begin{aligned} l_i(\beta, \delta) &= 1[d_j = 1] \log[\Lambda(w_o - \mathbf{M}(\beta - \delta_o))] + 1[d_j = 2] \log[\Lambda(w_h - \mathbf{M}(\beta - \delta_h)) \\ &\quad -\Lambda(w_o - \mathbf{M}(\beta - \delta_o))] + 1[d_j = 3] \log[1 - \Lambda(w_h - \mathbf{M}(\beta - \delta_h))]. \end{aligned}$$

The estimation gives two distinct vectors of coefficients which differ across categories. It is worth noting that the coefficients β, δ_o, and δ_h cannot be identified since only the differences $\beta - \delta_o$ and $\beta - \delta_h$ are estimated. Although coefficients are allowed to vary across categories some (or all) of them may not be statistically different across categories. This suggests the possibility of testing the parallel regression assumption through the pairwise comparison of the compound coefficients. A Brant test, which performs a Wald test on coefficient constancy for each variable, tests the null hypothesis that there is no difference between each pair of coefficients ($H_o : \delta_o = \delta_h = 0$). Variables presenting a statistically significant test statistic violate the parallel regression assumption and are used as regressors in the threshold equations. Defining $\mathbf{Z}$ as the subset of variables failing the test and $\mathbf{X}$ as its complement, the model reduces to:

$$P(d_j = 1) = P(\mathbf{X}\beta + v < w_o),$$

$$P(d_j = 2) = P(w_o < \mathbf{X}\beta + v < w_h),$$

$$P(d_j = 3) = P(\mathbf{X}\beta + v > w_h),$$

where $w_n = \delta_0 + \mathbf{Z}\delta_n + \varphi_n$ for $n = o, h$. The log likelihood function for the partial generalised logit model is the following:

$$\begin{aligned} l_i(\beta, \delta) &= 1[d_j = 1] \log[\Lambda(w_o - \mathbf{X}\beta + \mathbf{Z}\delta_o)] + 1[d_j = 2] \log[\Lambda(w_h - \mathbf{X}\beta + \mathbf{Z}\delta_h) \\ &\quad -\Lambda(w_o - \mathbf{X}\beta + \mathbf{Z}\delta_o)] + 1[d_j = 3] \log[1 - \Lambda(w_h - \mathbf{X}\beta + \mathbf{Z}\delta_h)], \end{aligned}$$

and corresponds to equation 1.6, which is estimated in this chapter. The results are reported in table 1.5.

1.4.2 Data

This chapter uses the Rural Survey for South Africa conducted by Statistics South Africa in 1997. The survey collected information on 6000 rural households located in the former homeland areas. Within the ‘old’ South Africa, 10 homelands (Bantustans) were created, four of which were granted ‘independence’ by South Africa (not recognised by any other country in the world): Transkei (*1976), Bophuthatswana(*1977), Venda (*1979), Ciskei(*1981), KaNgwane, KwaNdebele, KwaZulu, Gazankulu, Qwaqwa and Lebowa[11]. About 13 million people (31% to total population) lives in rural areas in the former homelands. The unemployment rate in South Africa, 23% at national level, is higher in rural areas than in the urban centres. In particular, with the adoption of apartheid policies,

[11] The asterisk precedes the year of independency declaration.

millions of Africans were segregated in the former homelands areas which are, in general, characterised by poor land quality and little job opportunities (BEPA, 2004). About 70% of households has access to land for farming purposes and farms are in general small with an average land size of about 1.4 hectares with one third of them cultivating a field which is smaller than 1 hectare.

This chapter identifies three groups of households on the basis of the allocation of unskilled family labour[12] and on the presence of hired labour. Households with members working both on- and off-farm are defined small-scale peasants while self-cultivators devote all family labour to their own farming activities, constituting the largest category in the sample (54%). Hiring-in households, besides family labour, employ additional hired workers and correspond to 13% of the sample. Households engaged in both hiring labor in and out have been excluded and constitute 3% of the entire sample. The sample has additionally been restricted to those households only involved in maize production (75%), so that a total of 2038 have been used in the estimation. Category characteristics are summarised in table 1.2.

Table 1.2: Descriptive statistics by category

Variables	Small-scale peasants (33.65 %)	Self-cultivators (53.75 %)	Hiring-in hhs (12.6 %)
Land hectares (mean)	1.77	2.14	2.60
Unskilled members (male - mean)	2.76	2.11	2.08
Unskilled members (female - mean)	3.00	2.59	2.47
Skilled members (mean)	1.16	1.21	1.25
Hhs with access to pensions (%)	15.72	29.19	27.11
Age of the household head (mean)	51.93	55.62	56.14

Agricultural assets are represented by the hectares of land for growing crops[13] and by household human capital (proxied by the age of the household head[14]). Hiring-in

[12]Skilled labour includes those household members with a level of education higher than the compulsory *general education and training* which runs from grade 0 to grade 9 (Department of Education, Republic of South Africa). In the South Africa education system there are other two educational bands: *further education and training* (from grade 10 to 12) and the *higher education and training* which includes undergraduate and postgraduate degrees, certificates and diplomas

[13]Grazing land is mainly communal and no information is available on the disposal size.

[14]If the head of the household belongs to the skilled labour forces employed off-farm than the oldest

households cultivate, on average, larger areas and have fewer unskilled members in the household (table 1.2). Exogenous transfers are represented by pensions measured by the number of pension eligible members in the household. In the 1997 the South Africa pension system provided a maximum benefit of 370 rand a month (around half of average household income) to all women over age 60 and men over age 65, which can however be reduced on the basis of individual incomes including income from assets (Case and Deaton, 1998). The use of this variable is further discussed in the next section. Additional variables are included and concern access to information: a dummy variable indicating whether or not the household obtains information on-farming through newspapers, media and other methods and an index of labour market development constructed as the average number of employed and unemployed members in the household by district. Finally I also include a measure of the intensity of the extension officer activities, computed as the average number of household visited by the officers in the district.

1.4.3 Identification

To test the validity of the comparative statics exercise conducted in the theoretical part of this chapter, it is first important to take into account the possible consequences of an endogenous household structure that is here represented by the number of skilled and unskilled household members. According to table 1.1, a higher number of unskilled members is expected to have a negative effect on the endogenous marginal productivity of labour. There are, however, two possible sources of endogeneity: the potential reverse causation between marginal productivity and household structure and the joint correlation with omitted unobservables. In the first case a lower on-farm labour productivity may induce households to reduce their size mainly through migrations or child fostering. This could lead to simultaneity bias problem which would potentially bias our coefficient toward zero.

member working on-farm is considered.

In the second case household structure could be correlated, for example, with unobserved household attitude and ability which also affects labour productivity. Unfortunately there is no valid instrument that could address this potential endogeneity problem. Moreover the generalised ordered logit approach adopted here cannot directly handle the endogeneity problem even in the presence of suitable instruments.

The exogeneity of income transfers is now discussed in order to address the potential identification problem concerning the liquidity effect. In our model income transfers are represented by pensions. Particular features of the Old Age Pension Program (OAP) in South Africa make it suitable for this analysis. The program reaches all women over age 60 and all men over age 65. The amount transferred does not take into account labour history; the means test is based on individual income only (other household members are not taken into account) and, in practise, only affects the white population (Duflo (2003), Edmonds (2006)). This provides no incentives for other household members to stop working and excludes the possibility of any adjustments in household composition in order to obtain the pension. Because pension take-up could be potentially endogenous, in this chapter I use the number of age-eligible members rather than the actual number of pensioners in the household. Pension eligibility is exogenous to household unobserved ability and it also provides a better proxy for the presence of pensioners in the household than that available using the survey. The survey, in fact, only asks about the pension to those members that are not currently working. It is not possible, therefore, to compute the take-up rate. Nevertheless Case and Deaton (1998), using a nationally representative survey conducted in 1993, two years after the expansion of the program, show that about 80% of African women and about 77% of African men age-qualified for the pension actually receive the transfer.

An additional concern on identification regards the possible endogeneity of the distribution of land. Land in South Africa is mainly allocated by the local or central authority

(85% of the sample). Households in general possess only use rights and cannot sell or rent the land (only 2% of interviewed households declare to have purchased or rented the land); this excludes the possibility of land size adjustment through access to the market. A further source of concern, however, regards the fact that land endowments could still be correlated with household unobservable characteristics, for example in the case that more influential households use their bargaining power to obtain larger land plots. The data do not offer the possibility to test this hypothesis and this potential endogeneity problem cannot be fully addressed. Moreover, even in the presence of suitable instruments (presented in chapter 3) the generalised ordered logit approach adopted here cannot easily handle instrumental variable estimations, as mentioned above. Finally, variations in the unobserved land quality could also lead to bias results. This problem, however, is partly ruled out since land in the former homelands, where the survey has been conducted, was generally of poor quality (BEPA, 2004). Moreover, by controlling for province and area fixed effects this problem is further reduced.

1.5 Results

The negative relationship between land size and productivity reported in table 1.3 support, at *prima facie*, the presence of imperfections in the factor markets. In the presence of perfect markets, equal output to land ratio is expected across households since all farmers should adjust demand and supply to market conditions. On the other hand, table 1.3 reveals significant variations in productivity across groups.

The results of the threshold constancy test, reported in table 1.4, determine the allocation of each variable either to the threshold or to the response regressions. Variables presenting a high Chi-square statistic do not satisfy the parallel regression assumption: their coefficients differ significantly across categories and have been classified as regressors in the threshold equations.

Table 1.3: Maize produced per hectare by class of land size

#	Land size in hectares	Obs	mean (tons/hectare)		Differences
1	Less than 0.15	228	21.200		
			(5.031)		
2	Between 0.15 and 0.25	255	0.756	2-1	20.445
			(0.048)		(4.756)
3	Between 0.25 and 0.5	385	0.449	3-2	0.306
			(0.023)		(0.048)
4	Between 0.5 and 1	512	0.355	4-3	0.094
			(0.036)		(0.046)
5	Between 1 and 3.5	349	0.243	5-4	0.112
			(0.041)		(0.056)
6	More than 3.5	145	0.145	6-5	0.099
			(0.018)		(0.064)

Standard errors are in parenthesis. The last column report the differences in mean between each land size category and the previous.

Land has been categorised as determinant of the marginal productivity of labour. Because rural households possess only use rights, land in general cannot be sold, rented or used as collateral and is, therefore, expected to positively affect the endogenous marginal productivity of labour. The fact that land is not considered as a covariate in the threshold regressions reasonably suggests that it does not influence household specific liquidity status and transaction costs. The number of skilled members also affects w^*. In the theoretical model presented in section 1.3, skilled workers are assumed to work a fixed fraction of total time endowment off-farm. In our sample, however, given the underdevelopment of the rural labour market, a considerable fraction of skilled members are found to be unemployed (41%). The presence of skilled members, therefore, is less probable to have an income/liquidity effect for the household; on the other hand, a positive knowledge spillover effect on other unskilled members combined with an opposite negative consumer-worker ratio effect (see Appendix 1) are likely to affect the marginal productivity on-farm, as it is suggested by the test results. Unskilled members (male and female) are also included as regressors in the latent regression (w^*) in line with the theoretical predictions. The Brant test suggests that the number of pension eligible members should be included as regressors in the threshold equations. While this finding does not alter the overall interpretation of their impact on category determination it confirms the presence of a

Table 1.4: Brant test on Parallel Regression Assumption

Variable	**chi2**	**p>chi2**
Determinants of the latent variable: w^*		p>0.05
Land (ha)	3.57	0.059
Unskilled labor - male	0.36	0.549
Unskilled labor - female	0.66	0.417
Skilled labor	0.08	0.781
Age of household head	0.85	0.358
Length of residency	3.00	0.083
Determinants of the thresholds: $\widetilde{w_n}$, $n = o, h$		p<0.05
Gender of household head (male)	16.36	0.000
Education of household head	4.99	0.026
Access to information (dummy)	10.14	0.001
Contact with extension officer (by district)	12.39	0.000
Labor market dev index (by district)	7.62	0.006
Number of pension eligible members	8.81	0.003

binding liquidity constraint which affects household participation in the labour market. Considering, for example, equation 1.2, the threshold delimiting the small-scale peasants category, $\widetilde{w_o^1}$, incorporates a liquidity premium. This implies that, when households face a cash shortage ($\lambda > 0$), off-farm earnings are valued more than their effective price (w_o). This is supported by the above results where pensions, which are expected to relax the liquidity constraint (having an impact on λ), have been found to affect the thresholds. This result suggests that households in rural South Africa face a binding liquidity constraint which affects their labour allocation choices. The intensity of extension officer services and the level of development of the local labour market have been found to affect the thresholds probably through their influence on transaction costs. The magnitude and the sign of such effects are discussed through the interpretation of the partial generalised ordered logit results which are reported in table 1.5.

In the first column (table 1.5), where the determinants of the marginal productivity of labour are reported, variables with a positive coefficient are expected to have a positive effect on w^* increasing the probability of belonging to a higher category where the lowest one is represented by the small-scale peasants category by self-cultivators households and by hiring-in households. Results confirm the predictions of the comparative static exercise. Larger land size increases the marginal productivity of labour and lowers the probability of working off-farm while increasing the probability of hiring in labour. The number of unskilled male and female members negatively affects the marginal productivity of labour implying that larger families tend to sell labour off-farm and are less likely to hire in workers.

Table 1.5: Results of the partial generalised ordered logit

Variable	w^*	**Small peasants** $\widetilde{w}_o$	**Self-cultivators** $\widetilde{w}_h$
Land (ha)	0.040**		
	(0.013)		
Unskilled labor - female	-0.152***		
	(0.031)		
Unskilled labor - male	-0.181***		
	(0.029)		
Skilled labor	0.037		
	(0.037)		
Age of household head	-0.0014		
	(0.004)		
Gender of household head (male)		-0.937***	-0.314**
		(0.111)	(0.137)
Education of household head		0.062***	0.024
		(0.015)	(0.017)
Pensions		0.874***	0.426***
		(0.103)	(0.110)
Access to information		-0.034	0.608***
		(0.123)	(0.188)
Contacts with extension officers (by district)		2.66***	5.991***
		(0.753)	(0.691)
Labor markt dev index (by district)		-0.768***	-0.286
		(0.174)	(0.170)
Sample size: 2037			
Goodness of fit			
Mckelvey and Zavoina R^2: 0.24			
Loglikelihood Full Model: -1727.935			
Outcomes correctly predicted: 60.28%			

Province and area dummies omitted. Numbers reported in parentheses are the standard errors; *, ** and *** indicates a significance at the 10%, 5%, and 1% level respectively.

To interpret the last two columns of coefficients, results can be related to those from two binary logit models in which the first category (small-scale peasants) is related to the other two and, in the last column, the first two classes are compared to the highest (hiring-in households). Coefficients can be interpreted as representing the effect of the selected variables on the probability of being in a higher class through their influence on the threshold delimiting the category and reported in the table header. Exogenous transfers (pensions) lower the liquidity premium and decrease $\widetilde{w}_o$ (equation 1.2). As expected, households receiving pensions are less likely to sell labour off-farm; coefficients indicate a positive effect on the probability to move from small-scale peasants to a higher category. Exogenous transfers have also a positive effect on the probability of hiring in labour thorough their effect on the hiring-in shadow wage ($\widetilde{w}_h$). The impact is, however, smaller than that on the small-scale peasant category. The degree of development of the local labour market is positively related to the effective off-farm wage since a better developed market implies lower transaction costs and favours the participation of the household in the labour market; however no effect is found for the hiring-in category. Being in a district where extension officers' visits are more frequent has a positive impact on transaction costs and increase the probability of being in a higher category, the effect is larger on the probability of hire in labour. Finally, results suggest that a male headed household is more likely to send members to work off-farm and less likely to hire in labour; because in rural South Africa male heads usually tend to migrate to urban centres (Makhura, 2001), their presence in the household seems to signal a liquidity shortage which affects the labour strategy adopted by the household.

1.6 Conclusions

This chapter has explored how, in the presence of labour market imperfections, asset endowments and liquidity constraints affect the choice of the labour regime adopted by the household. Households are categorised on the basis of their labour strategy (small-scale peasants, self-cultivators and hiring-in households) and a standard farm household model is used to explain the membership in the three groups. Labour market imperfections are translated into wage differentials where transactions costs determine the difference between the hiring-in and out wages. The presence of a binding liquidity constraint raises the opportunity cost of family labour, making off-farm opportunities more attractive and the hiring-in of labour harder. A partial generalised ordered logit is used to test the main predictions of the model and a Brant test on threshold constancy is employed to empirically identify the variables affecting the opportunity costs of family labour. The use of a more flexible estimation approach, in contrast with the ordered logit model proposed in Sadoulet et al. (1998), allows the modelling of household heterogeneous exposure to market imperfections. Results support the theoretical prepositions. A lack of liquidity induces household members to work off-farm. This confirms what has been previously found in the literature (Van Zyl et al. (1995) and Fenwick and Lyne (1999)). The results suggest the need for policy reforms in the rural credit sector which, according to the results, could be initially addressed to small-scale peasants since liquidity shortages seem to have a larger impact on their labour allocation decisions. On the other hand, when non-farm activities constitute the solution to cope with cash shortage, policies aimed to the development of the local labour market can reduce transactions costs and improve the participation to the labour market. This can help farmers to gain alternative sources of income and escape poverty.

Chapter 2

Liquidity constraints and farm household technical efficiency.

2.1 Introduction

This chapter provides an analysis of household technical efficiency using a sample of farm households in the KwaZulu Natal Province. About 60% of KwaZulu Natal households are estimated to be involved in agricultural activities. Although a large fraction of rural households has access to land for agricultural purposes, farming activities remain a marginal source of income. Understanding the reasons underlying poor performance in agriculture is an important task to provide insights for the ongoing land reform programs, and to improve the role of agriculture in contributing to the livelihood of rural households. Analysing farm household technical efficiency can help in understanding farm households' behaviour and the constraints that prevent them from optimally employing their resources. A previous study done by Piesse et al. (1996) provides the first analysis of South African farms' technical efficiency, which is however confined to a limited sample of households in the three homelands of KaNgwane, Lebowa and Venda. This chapter provides an analysis of technical efficiency with a particular focus on the role of access to liquidity and income diversification. It is, in fact, recognised in the literature that rural households engage in a wide range of activities in order to generate a livelihood. The standard analysis of technical efficiency is here extended to capture the linkages between farming and non-farming activities that characterise the livelihoods of most rural households.

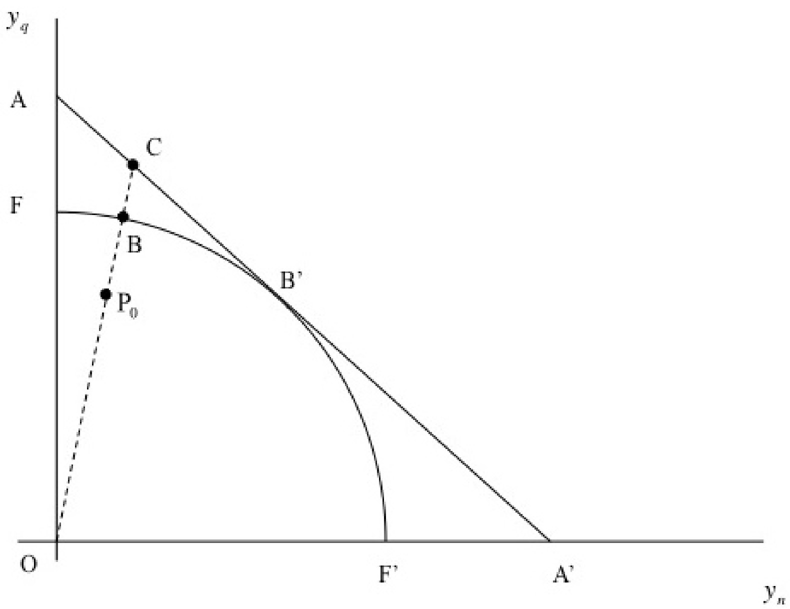
In the empirical specification, the liquidity effect is identified by exploiting the age eligibility criteria adopted by the South Africa Old Age Pension Program (OAP). Pension eligibility is used instead of actual pension receipt and several checks are conducted in order to examine the presence of potential confounding effects between the eligibility indicator and age trends or differences in background. Instrumental variable technique is also used to address the potential endogeneity of the income diversification index. The results show that access to liquidity and income diversification have a positive effect on household technical efficiency suggesting that institutional reforms to improve access to labour and credit markets can allow a more efficient use of farm household resources.

This chapter begins with a discussion of alternative concepts of efficiency, exploring the reasons behind the focus on farm household technical efficiency. Some empirical and theoretical considerations are also introduced as a foundation for the use of a household-level analysis of technical efficiency. Section 2.3 provides a review of the relevant literature while section 2.4 provides theoretical support for the use of a household-level analysis. This is followed by the description of the non-parametric technique adopted for the estimation of technical efficiency. Section 2.5 offers an overview of the data and presents the results of the estimation of the technical efficiency scores. Section 2.6 describes the empirical strategy adopted for the analysis of the determinants of household efficiency and discusses the results. Finally section 2.7 concludes.

2.2 Background on efficiency

Farm household efficiency is a multidimensional concept that has been widely analysed in the empirical literature and consists of two main components: technical and allocative efficiency. This chapter focuses on farm household technical efficiency (TE) and adopts a household-level approach that takes into account the role of non-farming activities. The concept of technical efficiency is based on the identification of a production frontier that represents the maximal combination of outputs attainable given an available set of inputs. Farm households operating on the frontier are considered technically efficient while those located below the frontier are considered inefficient. For a given input vector and considering only two outputs, the production possibility frontier is depicted in figure 2.1. The technical efficiency level of a farm located in P_0 corresponds to $TE = OP_0/OB$. At this point output could be increased to B without requiring extra inputs, and the household is considered inefficient.

Figure 2.1: Technical and allocative efficiency



Some considerations are required to reconcile the theoretical considerations above mentioned with the empirical possibilities. Methodological practicalities and the availability of data necessitate an adaptation of the concept of technical efficiency. The aim of any empirical analysis of technical efficiency is to provide a measure that captures the relationship between the observed production and some ideal, or potential production (the frontier). In principle, if all the possible combinations of inputs and outputs of production are known, a measure of pure technical efficiency could be obtained that would be in line with the above theoretical formulation. However, not all input-output combinations are known, quality may not be observed and data are usually available only for a sample of productive units. Therefore two main issues arise.

First, because it is not possible to observe the ideal or potential productive frontier, this concept needs to be adapted to what is observable and measurable. Departing from the underlying theoretical proposition that no units can exceed the ideal level of production, two main practices are conventionally adopted to obtain an estimate of the productive frontier: a) assuming a specific functional form for the relationship between inputs and outputs, b) considering the best performing units in the sample as forming the frontier.

Second, the interpretation of inefficiency scores need to be adapted to the availability of information on each farm in the sample. The assumption of homogenous inputs and outputs is necessary when their quality is not observed. By neglecting differences in input and output varieties, these unobservable characteristics contribute to the variation in the observed (estimated) efficiency. If, for example, better quality inputs are classified as homogenous with lower quality ones, they are likely to lead to higher observed technical efficiency. Moreover, the use of aggregate product and input values raises some concerns that will be discussed in section 2.4.1.

In general, because the concept of technical efficiency needs to be adapted to accommodate empirical possibilities and the availability of data, caution needs to be used in the interpretation of efficiency scores. Although they may not capture pure technical efficiencies, i.e. pure technical and engineering relationships, they provide, however, a useful representation of the variation in the intensity and effort in the use of observed inputs across farm households (Carter, 1984). In the rest of this chapter, I will refer to this modified interpretation of technical efficiency as 'observed technical efficiency'. Besides these general considerations, additional issues arise in relation to the specific method of estimation adopted and will be discussed with the empirical methodology.

Before introducing the use of a household-level measure of technical efficiency, it is necessary to explain why farm household allocative efficiency (AE) has not been considered in this chapter. Allocative efficiency reflects the ability to choose the optimal proportion of inputs or outputs given their prices and the farm household technology (Coelli et al., 2005). A household is considered to be allocatively efficient if it is minimising costs (or maximising revenues). Considering figure 2.1, the line AA' represents the isorevenue curve. Allocative efficiency is measured as the ratio between the revenues associated with the technical efficient output vector, point B, and those associated with the revenue efficient output levels, point C, that is $AE = OB/OC$. Because the estimation of allocative efficiency requires information on either input or output prices, some concerns arise. Besides the fact that obtaining prices for composite products or inputs may be a challenging and controversial task, the analysis of allocative efficiency in agriculture also raises some conceptual issues. Barret (1997), for example, points out that small farmers make consumption, production and labour supply decisions simultaneously on the basis of household specific shadow prices which differ from market prices. Using market prices in the estimation of allocative efficiency, pure allocative inefficiencies are confounded with misallocations caused by the presence of market imperfections. Although the concept can be adapted to embrace both sources of inefficiencies, the centrality of prices in the estimation of allocative efficiency remains a major concern. This chapter focuses on technical efficiency, whose estimation does no require information on prices although the use of aggregate values is sometimes required. This latter issue will be discussed below.

In this chapter I adopt a household-level measure of technical efficiency which takes into account the role of off-farm activities. Conventional analyses of farm household technical efficiency are, in contrast, conducted at farm level. Section 2.4 shows that when farm households engage in multiple activities, a farm-level analysis of technical efficiency is appropriate only if the following conditions apply: a) there are no technical interdependencies between farming and non-farming activities, i.e. skills acquired off-farm do not affect farm management; b) family and hired labour are perfect substitutes, and c) the farm household is not liquidity-constrained. If these conditions do not hold, farming and non-farming activities can be considered part of a joint production process. In general, while a farm level analysis could be appropriate when all of the above conditions are satisfied, a household-level analysis does not require such assumptions and, at the same time, can capture the reciprocal relationships between farming and non-farming activities that characterise the livelihood of rural households. Moreover, from an operational point of view, it does not require a distinction between labour employed on- and off-farm. It is, in fact, not always possible to separate the amount of hours worked on- and off-farm; in particular, the survey used in this chapter does not provide such information. On the other hand, data on the number of household members of working age are usually available in most surveys and in the one used here.

2.3 Literature review

The efficiency of South African farmers have been studied by Piesse et al. (1996) and Van Zyl et al. (1995). They find that inadequate land size causes efficiency losses for households in the three homelands of KaNgwane, Lebowa and Venda. The analysis of South African farm households' efficiency is extended here using a household-level approach with a particular focus on liquidity aspects and income diversification.

This chapter follows the work of Chavas et al. (2005), who show that in the presence of market imperfections or when farming and non-farming technologies are joint, farm and off-farm decisions are non-separable and a household-level analysis of technical efficiency is more appropriate than a farm-level analysis. This approach has also been adopted by Fletschner (2008), Fletschner and Zepeda (2002), Anriquez and Daidone (2008) and Fernandez-Cornejo (2007). This chapter extends the analysis of the relevance of market imperfections in the estimation of technical efficiency at farm household-level using the concept of jointness usually adopted in multi-product agricultural productions. Chavas et al. (2005) estimate technical efficiency using a non-parametric technique and, in the second part of the paper, several explanatory variables are considered in the analysis of the factors influencing household efficiency. Various measures of access to financial resources are considered, including income from off-farm activities. However, the identification of such effects is not discussed. In this chapter a liquidity effect is identified exploiting the age eligibility criteria adopted by the South African Old Age Pension Program and instrumental variable techniques are adopted to identify the effect of income diversification on household technical efficiency.

As anticipated, the impact of access to liquidity is analysed considering the pension transfer provided by the South Africa OAP to all women over age 60 and men over age 65. Through this analysis, this paper contributes to the current debate on the effects of the OAP on household behaviour. One of the controversies lies on whether the pension induces an income effect that reduces recipient and, possibly, other family members' labour supply. On the other hand, when the household faces a credit constraint, the pension can have a positive effect on labour supply, enabling farm investment and financing job searching, also through migration. On one side Bertrand et al. (2003) argues that the pension transfer has a negative effect on the labour supply of the prime age adults living with a pensioner, the impact differs according to the age and gender of the individuals. Ranchhod (2006) also finds a negative effect of the pension on the labour supply of the beneficiaries. On the other side, Klasen (2008) finds no effect of the pension income on the reservation wage of the unemployed and Jensen (2004) finds no evidence that households reduce labour supply when they receive the pension. Moreover, Posel et al. (2006) and Ardington et al. (2009) questioned the findings in Bertrand et al. (2003), arguing that once migrants are included in the analysis the results change considerably. This chapter attempts to further address this issue by focusing on farm households which have the peculiar characteristic of being a supplier and an employer of labour at the same time. The relationship between pension and labour supply is analysed from a different perspective. As it will be discussed below, in the empirical estimation of technical efficiency, forced by data limitations, I consider the number of adult family members, instead of the number of hours worked, as input in the production of on and off-farm outputs. In this context, a negative labour supply effect will imply a negative impact of pension receipt on technical efficiency since labour inputs are left unproductive. On the other hand, if households are liquidity constrained, access to the transfer is expected to improve household technical efficiency, for example, by enabling the use of more expensive and higher quality inputs and factors or by allowing households to overcome the entry barriers in the labour market. However, if pension receipt partly crowds out private transfers such as remittances, as analysed in Jensen (2004), the potential income and liquidity effects are neutralised.

2.4 Technical efficiency at farm household-level

Conventional analyses of technical efficiency at farm level have generally neglected the linkages between household farm and non-farm activities produced by technical interdependencies and market imperfections. As suggested by Chavas et al. (2005) the use of a household-level analysis of technical efficiency relies on the argument that on and off-farm activities are jointly produced. This section starts with a formal definition of the concept of joint production as applied to multi-product firms. The definition is then applied to farm and non-farm activities at household-level. In the second part of this section, I analyse how access to liquidity and income diversification can impact household observed technical efficiency.

Originally, the definition of joint production applied to multi-product firms referred to multiple outputs that cannot be produced separately, but are joined by the use of common non-allocable or public inputs (c1). Public inputs, once acquired to produce one output, are available costlessly for the production of other outputs (Baumol et al., 1982). Common examples are wool and mutton from sheep or wheat and straw. A second commonly cited cause of joint production is the presence of technical interdependences (c2), for example, when the pesticide used in a field affects the yields of the nearby field. These conditions are still regarded as primary causes of jointness. In addition, several authors (Shumway et al. (1984), Moschini (1989) and Leathers (1991)) consider the presence of multiple outputs competing for an allocable input that is fixed at the productive unit level (c3) as an additional source of jointness in production.

I will here follow Lau (1972) to provide a formal definition of joint production that will be related to the above mentioned causes of jointness. According to Lau (1972) two types of jointness can be identified: jointness in outputs and jointness in inputs:

- A production function $F(y_1, y_2; x, z) = 0$ with two outputs, y_1 and y_2, and two inputs, x and z, is said to be non-joint in inputs if there exist individual production functions $y_1 = f_1(x_1, z_1)$ and $y_2 = f_2(x_2, z_2)$ such that $F(y_1, y_1; x, z) = 0$ if and only if $x_1 + x_2 = x$ and $z_1 + z_2 = z$. That is to say that separate production functions can be obtained for each of the products and no inputs simultaneously contribute to the production of the two goods (Leathers, 1991).

- The same production function is said to be non-joint in outputs if there exist individual input requirement functions $x = g_x(y_{1x}, y_{2x})$ and $z = g_z(y_{1z}, y_{2z})$ such that $F(y_1, y_1; x, z) = 0$ if and only if $y_{1x} + y_{1z} = y_1$ and $y_{2x} + y_{2z} = y_2$. That is to say that separate input requirement functions can be obtained for each of the input. According to Shumway et al. (1984) this condition is rarely descriptive of real world, therefore, I will focus on the first type of jointness.

The definitions above provide a mathematical representation of joint production but are not easily testable, therefore alternative behavioural propositions are provided in Lau (1972) and are reported below. A necessary and sufficient condition for non-jointness in inputs is for the profit function to be additively separable in outputs:

$$\Pi = \sum_i p_i \pi_i(w/p_i, r/p_i), \tag{2.1}$$

where π_i refers to the profit function of output i, w and r are the prices of input x and z respectively. The maximised profit has the following property:

$$\frac{\partial \Pi^2}{\partial p_i \partial p_j} = \frac{\partial y_i^*}{\partial p_j} = 0, \quad i \neq j. \tag{2.2}$$

Where y_i^* is the optimal level of output i. According to equation 2.2, the two outputs are attained from a non-joint production process if the supply of a good is affected by changes in its own price but not in the price of the other product. While it is possible to see that the above conditions are violated in the presence of non-allocable inputs (c1) and technical interdependencies (c2) within a multiple output production process, a further step must be taken in order to explore how the presence of fixed allocable inputs (c3) can lead to jointness in production. Following Shumway et al. (1984), when a constraint on the total amount of inputs available is introduced ($\bar{z} = z_1 + z_2$) the problem becomes:

$$\begin{aligned} \max_{x_i, z_i, \lambda} \Pi &= \sum_i p_i f_i(x_i, z_i) - w x_i \\ & \quad s.t. \quad z_1 + z_2 = \bar{z}, \end{aligned}$$

where the two production functions $f_i(x_i, z_i)$ are considered independent for simplicity. The Lagrangian of this problem is:

$$L = p_1 f_1(x_1, z_1) + p_2 f_2(x_2, z_2) - w(x_1 + x_2) - \lambda(\bar{z} - z_1 - z_2),$$

where λ is the Lagrangian multiplier associated to the input availability constraint. The solutions of this problem are the optimal input allocations, $x_i^*(w, p_1, p_2, \bar{z})$ and $z_i^*(w, p_1, p_2, \bar{z})$, and the endogenous price of the fixed input $\lambda^*(w, p_1, p_2, \bar{z})$. Once the optimal demand for inputs are substituted to obtain the optimal output supplies, $y_i^*(w, p_1, p_2, \bar{z})$ and $y_2^*(w, p_1, p_2, \bar{z})$, it is possible to notice that equation 2.2 is violated since the supply of one output is not independent of changes in the price of the other output.

The above definition can be applied to farm and non-farm activities. In particular, three conditions can possibly lead to jointness between farm and non-farm production: d1) there are technical interdependencies and non-allocatable inputs within farming and non farming technologies, d2) family and hired labour are imperfect substitutes and d3) the household is liquidity constrained.

The first condition usually emerged when skills acquired off-farm improve farm management (Chavas et al., 2005). Ravallion (2003) points out the importance of externalities for rural development given the fact that most rural households engage in multiple activities. By engaging in off-farm activities, for example, farmers can learn about new techniques of production (Feder and Slade (1985), Foster and Rosenzweig (1995)), bookkeeping and finance management. In particular, the presence of local non-farm industries that encourage the acquisition of knowledge and skills can also benefit local farmers at household-level, through knowledge sharing within the household (Basu et al., 2002). Moreover some public inputs can be shared between farm and non-farm activities, for example, the housing infrastructure, food provision and equipments such as vehicles and other tools.

The imperfect substitutability between family and hired labour (point d2) is usually induced by the presence of transaction costs in the labour market. Regarding this latter aspects, the presence of supervision and other transaction costs, give family labour specific features that distinguish it from hired labour. In this context, family labour can be considered as a quasi - fixed allocable input in the short run since no perfect substitutes are available. In general, the presence of multiple outputs competing for a limited amount of inputs implies that the production of one output reduces the availability of resources and has a negative effect on the production of the other output. This argument applies to the allocation of family labour between on and off-farm activities and implies the jointness between farm and non-farming activities. The last aspect (point d3) is related to the presence of a binding liquidity constraint. Farming decisions are constrained by the availability of financial resources. In this case, off-farm earnings can promote farm production by relaxing the on farm liquidity constraint leading again to the jointness between farm and non-farm activities.

In general, while a farm production function can be entirely separated from the non-farm production function when none of the above conditions applies, a joint household-level analysis does not require such assumptions. In practice, this also refers to the ability of quantifying the separate amount of inputs used for farm and non-farm activities. The difficulty of obtaining data on activity specific inputs partly arises from the joint nature of the two production processes as described above. Using standard surveys, for example, externalities between on and non-farm activities cannot be measured. Moreover, inputs are not usually recorded with sufficient detail (distinction between hours worked on and off-farm) and, because their allocation is affected by seasonality, often, only the total quantities available at household-level can be observed. Therefore, both the inherent jointness between farm and non-farm activities and the data limitations suggest the use of a household-level analysis of technical efficiency.

In the remaining of this section I will analyse how access to liquidity and income diversification affect household observed technical efficiency. As anticipated, because not all inputs, outputs and their quality are observed, unobservable factors contribute to the variation in the estimated level of technical efficiency across households. The use of low quality inputs, for example, can result in technical inefficiencies although the timing and the method of production employed are optimal. The following discussion refers to the this definition of observed technical efficiency.

Access to transfers, such as pensions, can produce alternative effects. First, in the presence of a binding liquidity constraint, the transfer can help ease the constraint and allow the purchase of new technological packages that can increase the amount of output produced and therefore the observed technical efficiency. The household, for example, might be able to purchase higher yielding seeds or adopt a more remunerative cropping mix and increase production (Carter, 1989). Moreover, access to liquidity may help farmers to better cope with adverse shocks and afford the costs of entering better quality and more remunerative jobs, such as the cost of equipments, rents and skill acquisition. Dercon and Krishnan (1996), for example, found that entry constraints are important determinants of the choice of income portfolio for rural households in Tanzania and Ethiopia. Even when the transfer is not used for productive purposes, but for food consumption, it can induce a more intensive use of land and family labour if the improved nutritional levels of family members are translated into higher labour productivities. On the other hand, a negative impact can also be observed and is specific to the methodology adopted for the estimation of technical efficiency. Because in the estimation of technical efficiency, that will be described in the next section, I consider the overall number of family members rather than the hours worked as inputs of production, labour supply effects can also be captured. Exogenous transfers can produce an income effect that reduces household labour supply. In this case, a negative effect of the transfer on technical efficiency is expected since labour inputs are left unproductive. Finally, if the pension receipt partly crowds out private transfers such as remittances, as analysed in Jensen (2004), the potential income and liquidity effects are neutralised.

A similar analysis can be conducted for the impact of off-farm earnings on household technical efficiency. Non - farming activities can have a positive effect on technical efficiency mainly because: a) non farm earnings can provide liquidity to the household and produce similar effects to those reported above, b) skills acquired off-farm can generate positive knowledge spillovers improving farmers' managerial ability. On the other hand, a negative effect is expected when off-farm opportunities subtract time for farm management therefore preventing the adoption of management-intensive innovations. Moreover if diversification is the result of a income-smoothing strategy, case studies have shown that lower risk is often obtained at the cost of lower incomes (Dercon, 2002). Which effects prevail is an empirical question[1] and will be discussed in the next sections.

2.4.1 Measuring technical efficiency

There are two main approaches to the estimation of technical efficiency. In this chapter, farm household technical efficiency is estimated using a non-parametric approach known as Data Envelopment Analysis (DEA). Before describing the estimation procedure I will briefly introduce the most common alternative method of estimation, the parametric stochastic frontier analysis.

The stochastic frontier approach has been introduced by Aigner, Lovell and Schmidt (1977) and Meesen and van den Broek (1977). This method econometrically estimates the production frontier by explicitly specifying the functional form of the production function,

[1] Goodwin and Mishra (2004) for example, using a farm level efficiency analysis finds that the involvement in off-farm activities decreases farm efficiency for a sample of US farms. The analysis, however is not extended at household-level. Fletschner and Zepeda (2002) find a positive effect of income diversification on allocative efficiency using data on rural farm households in eastern Paraguay.

$f(\mathbf{x}, \beta)$:

$$y_i = f(\mathbf{x}, \beta) \cdot TE_i \cdot e^{v_i}.$$

The observed level of output, y_i, is therefore given by a deterministic component $f(\mathbf{x}, \beta)$, a standard random error, e^{v_i}, capturing measurement error and other random factors out of farm control and a random component representing technical efficiency, TE_i. This latter factor is bounded to be positive and lower than one and measures the distance from the frontier. The identification of the two error components requires some distributional assumptions since only the aggregate term is estimated. This method has the advantage of distinguishing between pure inefficiency and unobserved random noise and can also be easily adapted to exploit longitudinal data. On the other side, misspecifications of the functional form can be erroneously interpreted as inefficiencies and no methods are available to deal with multiple outputs other than using an output aggregator. Moreover, the presence of zero-valued inputs is troublesome when using the most common functional forms that require logarithmic transformations: Cobb-Douglas or trans-log. Common solutions are arbitrary manipulations of the data, for example by adding small constant values to the input variables, which however could lead to bias estimates. In addition inputs could be considered perfect substitutes and aggregated or alternative functional forms can be used that can accommodate zero-valued inputs. A functional form chosen merely on the basis of data requirements, however, might not appropriately describe the underlying production process and could rise problems of misspecification of the production function.

The method adopted here, DEA, is a deterministic approach that has been first introduced by Charnes et al. (1978). In contrast with the previous method, this approach does not impose restrictions on the underlying farm technology [2]. This methodology is more

[2]The statistical properties of the estimator have been analysed in Banker (1993) where its consistency is proved. However, since the estimates are obtained from a finite sample, they are sensitive to sampling variations. Simar and Wilson (1998) propose a bootstrapping technique to estimate confidence intervals for efficiency scores that reveal their sensitiveness to sample variation

suitable for the analysis conducted here mainly because of its adaptability to multiple inputs and outputs that can be quantified using different units of measurement[3]. Moreover, at the same time, DEA does not require the distinction between the amount of labour employed on and off-farm that is not available in the survey used. However, because it is a deterministic approach, deviations from the frontier are all attributed to inefficiency. Therefore, for example, it fails to take into consideration differences in environmental and weather conditions. However, as far as this chapter is concerned, the use of data on the KwaZulu-Natal province only, restricts the potential variation in such aspects[4].

DEA models can be either input or output oriented. While the two measures are equivalent under constant returns to scale, they differ when variable returns to scale are assumed. The input-oriented approach considers a proportional contraction in the use of inputs given the level of outputs. The output-oriented approach measures the proportional expansion of outputs that could be attained given the available inputs (Coelli et al., 2005). In this chapter I opted for a output oriented analysis since most of the inputs considered, such as land and family labour cannot be easily increased or decreased in the short run according to production requirements.

The farm household technology can be represented by the following technology set $F(y_q, y_n; X, H, L)$ such that family labour (L), hired labour (H) and other inputs (X) can produce the farm and non farm outputs, y_q and y_n. Technical efficiency (TE) is intended as the distance of the household input/output bundle to the multi-input multi-output productive frontier constructed using the information on all the farm households in the sample. Given the presence of multi inputs and outputs, the empirical estimation

[3]An alternative way to handle multiple inputs and outputs is the use of stochastic input or output distance functions (Kumbhakar and Lovell, 2000). These methods, however, require the use of logarithmic transformations that are not suitable for this analysis since most of the input and output variables have zero entries and the use of arbitrary measures, such as replacing zeros with very small numbers, would be necessary and questionable.

[4]Moreover a set of sensitivity exercises are conducted to understand the sensitivity of DEA estimates.

of technical efficiency, discussed in the next section, is based on the concept of output distance function:

$$TE = \min\{\phi : F(y_q/\phi, y_n/\phi; X, H, L) = 0\}.$$

Following the DEA approach the productive frontier is computed as the larger upper bound set of all the possible input - output combinations. The frontier, therefore, is composed by the best performing farm households in the sample. Because it is likely to be sensitive to outliers, I employ the method proposed by Wilson (1993) to eliminate the identified outliers[5]. The output oriented technical efficiency is based on obtaining an optimal set of weights from the maximisation of each household's ratio of all outputs and inputs, given by $\mu_{\mathbf{i}}' \mathbf{y}_i / \nu_{\mathbf{i}}' \mathbf{z}_i$, where $\mathbf{y}_i$ and $\mathbf{z}_i$ are the vectors of M outputs and N inputs of firm i. The associated vector of weights, μ_i (Mx1) and ν_i (Nx1) are obtained by solving the mathematical programming model reported below, where household efficiency is maximised subject to other firms' efficiency measures being lower than 1.

$$\begin{aligned} &\max_{\mu_i, \nu_i} \frac{\mu_{\mathbf{i}}' \mathbf{y_i}}{\nu_{\mathbf{i}}' \mathbf{x_i}} \\ &s.t \quad \frac{\mu_{\mathbf{i}}' \mathbf{y_j}}{\nu_{\mathbf{i}}' \mathbf{x_j}} \leq 1 \quad j = 1, 2 ... I, \\ &\qquad \mu_{\mathbf{i}}, \nu_{\mathbf{i}} \geq 0 \end{aligned}$$

This linear programming model is solved for each of the I households in the sample so that each unit is assigned a set of weights that is most favourable to them. Departing from this base specification a normalisation is imposed, $\nu' \mathbf{x_i} = 1$, to ensure the existence of a unique solution to the model. The problem can be represented in its envelopment

[5] Outliers are defined as observations with very low probability. The elimination of 11 outliers does not affect significantly the average estimated technical efficiency that only changed slightly. Moreover, the elimination of outliers does not alter the general results obtained in the second stage analysis.

(dual) form where ϕ_i is a scalar associated to firm i, and $\lambda_{\mathbf{i}}$ is a (Ix1) vector of constants.

$$\max_{\phi_i,\lambda_i} \phi_i,$$

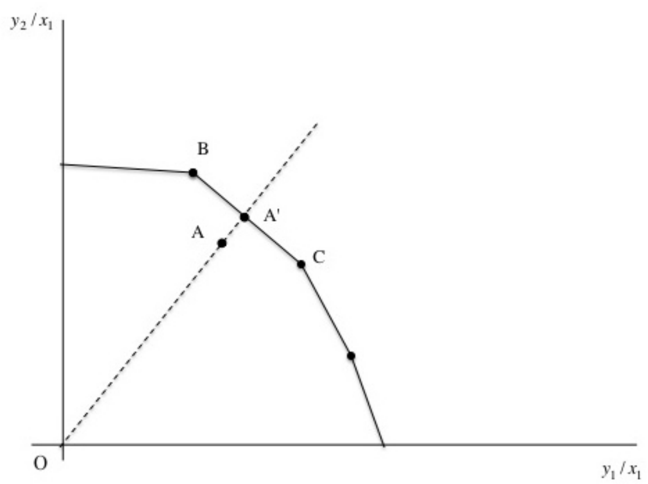
$$s.t \quad -\phi_{\mathbf{i}}\mathbf{y_i} + \mathbf{Y}\lambda_{\mathbf{i}} \geq 0,$$

$$\mathbf{x_i} - \mathbf{X}\lambda_{\mathbf{i}} \geq 0.$$

$$\lambda_{\mathbf{i}} \geq 0,$$

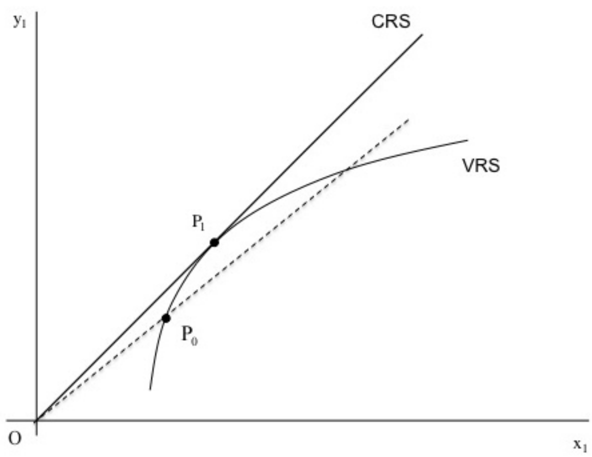
where $\mathbf{X}$ is the (NxI) matrix of inputs and $\mathbf{Y}$ is the (MxI) matrix of outputs of all households in the sample. The measure of technical efficiency is given by $1/\phi_i$. The dual problem is solved for each of the I households in the sample. Intuitively each problem maximises the radial expansion of the output vector of firm i while remaining within the feasible output set. The projected point $(\mathbf{X}\lambda_{\mathbf{i}}, \mathbf{Y}\lambda_{\mathbf{i}})$ is a linear combination of the observed data points. The elements of the vector λ_i are non-zero in correspondence to those households that form part of the relevant part of the frontier. In the one-input two-output case depicted in figure 2.2, the projected point for household A is the point A' and the element of the vector λ_A are non-zero in correspondence to households B and C (that are usually referred to as peers). The scalar ϕ_A captures the distance between A and the projected point A'.

Figure 2.2: Output oriented DEA



Before proceeding to the estimation of technical efficiency, a choice needs to be made about the use of variable or constant returns to scale in the estimation of the production frontier. The two options are depicted in figure 2.3 using a one-input one-output production technology where P_0 and P_1 are two households. The presence of market imperfections (in particular the absence of the land market) and constraints on liquidity are likely to cause households to not operate at optimal scale, that is the household is unable to become more productive by changing the scale of operation. The use of constant returns to scale (CRS) implies a linear production frontier hence, as shown in figure 2.3, only household P_1 would be considered technically efficient. Household P_0 is not operating at optimal scale since it could become more productive by changing its scale of operation. Using a CRS production frontier technical inefficiency is confounded with scale efficiency that measures whether the household is operating at the most productive scale. When using variable returns to scale (VRS), in contrast, P_0 and P_1 are both considered technically efficient although they are not equally productive as indicated by the dashed line. Therefore, when markets are not perfectly competitive, a VRS frontier is preferable since technical efficiency measures are not affected by the use of an inadequate scale of operation.

Figure 2.3: Variable versus constant returns to scale



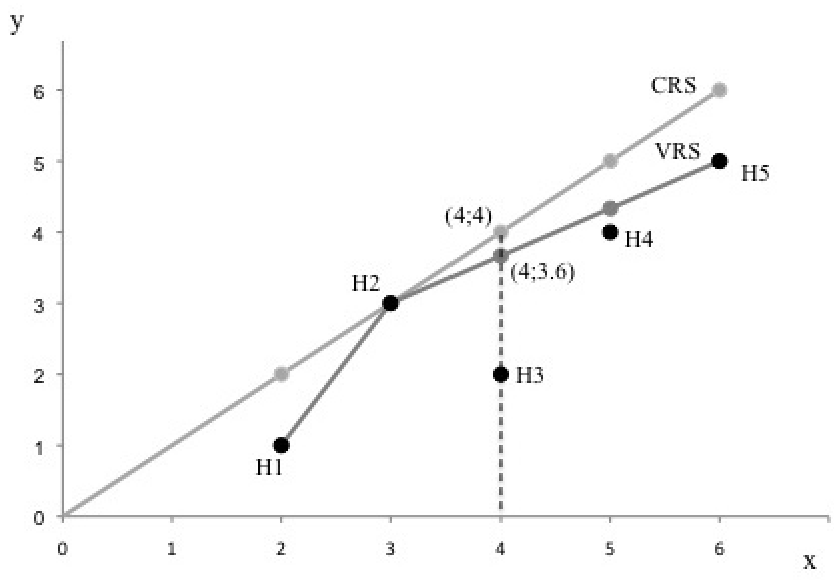
The production possibility frontier is, therefore, estimated using VRS by including the constraint $\mathbf{I1}'\lambda_\mathbf{i} = \mathbf{1}$, where $\mathbf{I1}$ is a vector of ones, that is the projected point for each household is a convex combination of the corresponding peers (in contrast of being a linear combination). This ensures that each household is compared with households of similar scale. The role of the convexity constraint is better explained using a simple example considering 5 households producing one output (y) and employing one input (x).

Table 2.1: Example of DEA under constant and variable returns to scale

			Variable returns to scale				Constant return to scale			
Unit	Y	X	Peers	λ	(λX; λY)	TE	Peers	λ	(λX; λY)	TE
H1	1	2	H1	1	(2;1)	1	H2	0.67	(2;2)	0.50
H2	3	3	H2	1	(3;3)	1	H2	1	(3;3)	1
H3	2	4	H2,H5	0.67, 0.33	(4;3.6)	0.55	H2	1.33	(4;4)	0.50
H4	4	5	H2,H5	0.33, 0.67	(4.3;5)	0.92	H2	1.67	(5;5)	0.80
H5	5	6	H5	1	(6;5)	1	H2	2	(6;6)	0.83

Considering the estimates reported in table 2.1, the presence of the convexity constraint in the VRS method is confirmed by the facts that the estimated λ add up to one (column 5). Considering for example household H3, the projected point in the VRS case is a convex combination of its peers H5 and H2 while, in the CRS case, the projected point is a linear combination of the peer H2, as depicted in figure 2.4. It is possible to notice that, in the CRS case, all households in the sample are compared to household H2 independently of the size of operation. In contrast, under VRS, households are benchmarked to peers that are closer in terms of size and the convexity constraint ensures that a lower weight is given to the more distant households. Considering again household H3 a lower weight is attributed to the peer H5 since the difference in size, in absolute terms, between H3 an H5 is larger than that between H3 an H2.

Figure 2.4: Example: variable versus constant returns to scale



2.5 Household technical efficiency in KwaZulu Natal

The analysis of technical efficiency has been conducted using the third wave (2004) of the KwaZulu-Natal Income Dynamic Survey (KIDS)[6]. The KIDS is a comprehensive household survey that includes information on household characteristics, expenditure, income and farming activities. A sample of 547 farm households has been used for the estimation of technical efficiency[7] .

About 80% of farm households produces maize which is often grown together with other cereals, vegetables and fruits. About 60% of the farms own some livestock and are engaged in animal husbandry. Farms are in general small and the average land size is of about 1.4 hectares. Farm households rely also on off-farm earnings and about 53%

[6]KIDS data have been collected thanks to following collabourating institutions: University of KwaZulu-Natal (UKZN), the University of Wisconsin-Madison and the International Food Policy Research Institute (IFPRI) However, in order to accommodate new areas of interest, the participating institutions have been broadened to include the London School of Hygiene and Tropical Medicine (LSHTM) and the Norwegian Institute of Urban and Regional Studies (NIBR). In addition to the resources provided by each of the collabourating institutions, the study was funded by the UK Department for International Development (DFID) through DSD, the National Research Foundation, the Norwegian Research Council, USAID and the Mellon Foundation.

[7]The initial sample of farm households, including all households conducting agricultural activities, have been reduced following the method proposed by Wilson (1993) in order to eliminate few outliers.

are involved in casual or permanent off-farm activities which constitutes an important component of overall household income. Non-farming earnings, excluding income from pensions, other transfers and remittances, contributes to the 58% of total income. Only 15% of households employes hired labour and about 30% do not use fertilisers, sprays or purchased seeds. The survey do not provide specific information on the credit status of the household, however only 20% of the households have access to formal credit, in particular only 5% has received a loan from a bank or building society. This evidence supports the presence of limited access to credit facilities for the households.

Table 2.2: Descriptive statistics of variables used for efficiency estimation

	Obs	Mean	Std. Dev.	Min	Max
Outputs					
Maize (in kg)	326	17.46	56.72	0.02	625.00
Vegetables (value in RAND)	388	38.34	98.74	0.07	870.83
Fruits (value in RAND)	71	27.08	48.65	0.12	301.25
Others (value in RAND)	121	8.97	23.88	0.25	250.00
Income from animals (in RAND)	294	146.24	299.04	0.13	2710.42
off-farm income (in RAND)	298	2272.72	2593.80	20.00	13267.0
Inputs					
Male members (in adult equivalent)	518	2.24	1.37	0.10	9.40
Female members (in adult equivalent)	542	2.56	1.59	0.30	12.00
Land (hectares)	557	1.38	7.04	0.00	75.00
Hired labour (number of workers)	80	1.629	2.51	0.08	14.83
Livestock (Tropical livestock unit)	340	2.10	3.71	0.01	35.00
Cost of inputs (value in RAND)	386	37.15	71.00	0.25	1016.6

The estimation of technical efficiency employes 6 outputs and 6 inputs which are reported in table 2.2. The number of inputs and outputs used in the estimation of technical efficiency, relatively to the sample size, affects the shape of the efficiency frontier. The higher the number of inputs and outputs included in the estimation of technical efficiency the higher the number of households identified as efficient. At the same time, however, excluding some variables, for example those capturing the quality of inputs and outputs, can increase the bias of the estimates. In the light of this trade off and of the previously mentioned limitations of the observed technical efficiency measure, I opt for the choice of a minimum common set of inputs and outputs for the first stage estimation of technical efficiency, leaving quality related variables for the second stage analysis presented in the next section. The total production of maize has been measured in kilograms while vegetables, fruits and others products have been aggregated using farm level prices when available and median prices at district level otherwise. An additional aggregate output includes the revenues from the sale of animals, meat and animal products such as eggs and milk. Finally, off-farm income includes the earnings from regular and casual employment. Other forms of non-agricultural self employment have not been considered since data were not reliable. The set of inputs includes the number of male and female adults which have been computed using the equivalence scale proposed by Deere and de Janvry (1981)[8]. Land represents the total surface devoted to farming activities while hired labour is measured using the number of permanent and temporary workers employed on the farm. The cost of inputs include the cost of seeds, fertilisers, sprays, ploughing and veterinary expenses. Finally livestocks has been measured in tropical livestock unit (TLU) which is a standard procedure used to aggregate across different species[9]. It is worth noting that the aggregation of inputs and outputs into aggregate categories such as labour, capital and purchased inputs using farm level prices introduces an additional conceptual issue. Technical inefficiency measures can be confounded with allocative errors between individual inputs and outputs within aggregate categories[10] (Ali and Byerlee, 1991).

The summary results of the estimation of technical efficiency are reported in table 2.3[11]. Efficiency estimates are low, the average estimates of technical efficiency are lower then those reported in Piesse et al. (1996). However, DEA estimates largely depend on

[8]This procedure attributes a weight of 0 to members aged below 3, 0.1 to children aged between 3 and 5, 0.3 to members aged between 5 and 8 and over 75, 0.5 to those aged between 8 and 12 and between 65 and 75, 0.8 to those aged between 13 and 17 and between 59 and 65 and 1 to the remaining members aged between 17 and 59.

[9]Cattle correspond to 1 TLU while sheeps and goats correspond to 0.7 TLU.

[10]In particular, Fare et al. (2004) show that when different types of inputs are aggregated using their prices as weights, the technical efficiency scores will be biased downward and when different types of outputs are aggregated using their prices as weights technical efficiency scores will be further biased downward.

[11]These estimates have been obtained using the package FEAR for R (Wilson, 2007).

Table 2.3: Technical efficiency by district municipalities

District	Obs	Technical efficiency	% Efficient
Ugu	69	0.41	21.73
Umgungundlovu	27	0.44	22.22
Uthukela	64	0.41	18.75
Umzinyathi	28	0.31	7.14
Amajuba	39	0.42	17.94
Zululand	71	0.28	8.45
Uthungulu	94	0.40	17.02
iLembe	17	0.39	17.64
Vhembe	47	0.50	21.27
eThekwini	93	0.44	23.65
Total	547	0.40	18.03

the characteristics and size of the sample considered. Therefore, comparisons with other findings are not possible. Considering for example the Ugu district, the average farm household can possibly increase output by 59% without changing the bundle of inputs employed. Because this analysis considers also off-farm activities together with conventional farm outputs, high inefficiencies could also signal the presence of barriers to non farm employment. This will be analysed in the next section where the determinants of technical efficiency are explored. Before proceeding to the analysis of the determinants of technical efficiency, few exercises are conducted to understand the sensitivity of technical efficiency estimates to the choice of alternative input and output bundles and of sample variations. When skilled and unskilled family members are considered as separated input of production average technical efficiency increases to 0.49. This increase is equally distributed across districts. As previously mentioned the use of a higher number of inputs (or outputs), relatively to the sample size, in the estimation of technical efficiency leads to a higher number of households identified as 'efficient'. This is confirmed since 26% of the households are considered efficient when family labor is distinguished by education attainments. On the other hand, the omission of this distinction increases the bias of the estimates. The fact that the average technical efficiency estimates do not change notably, i.e. the average technical efficiency remains low, however suggests that failing to account for difference in labor quality at this stage of the analysis does not significantly alter the overall picture. Therefore labor quality differences will be analysed in the second stage analysis proposed in the next section.

The DEA method does not account for variations in weather or environmental conditions across districts that are therefore capture by the efficiency measures. To investigate whether the presence of districts with particularly favourable (or unfavourable) conditions are significantly affecting the estimates, technical efficiency scores have been re-estimated excluding each district at a time[12]. The average estimates varies between 0.40, excluding ILembe district municipality, and 0.47, excluding eThekwini district municipality, which are the smallest and the largest district in the sample. The variation in the average estimates seems, therefore, to be mainly driven by the reduction in the sample size rather then by the presence of particular district-specific characteristics that could significantly affect the estimates. In addition the analysis of the determinants of technical efficiency will control for common characteristics at district municipality level. Finally, I also computed bias-corrected estimates using the bootstrapping method proposed by Simar and Wilson (1998) with 2000 replications and obtained a very similar average estimate of 0.37. This bootstrapping method repeatedly estimates technical efficiency scores on random samples obtained by simulating the true data generating process. However, because this method seems to be quite sensitive to the choice of the initial random value for the drawing of random samples it is here used only for comparison.

The exercises conducted in this section add some confidence on the estimates of technical efficiency obtained using the DEA method. These estimates are employed in the next section to analyse the factor determining the variation in the observed technical efficiency

[12]Technical efficiency scores have also been re-estimated excluding very large farms (above 50 hectares, 1% of the sample) and the average technical efficiency differs only in the third decimal place. These households have also been excluded from the second stage analysis during robustness checks and their exclusion does not affect the results.

across households.

2.6 Analysis of factors affecting technical efficiency

In the analysis of the determinants of technical efficiency, the efficiency estimates are regressed on a set of contextual factors usually considered in the literature such as human capital and other household and market characteristics. In contrast with the inputs and outputs variables considered in the estimation of technical efficiency, these factors are intended to capture differences in managerial abilities and access to factor markets that affect household decision making. There is an ongoing debate on the use of this two stage procedure that involves the estimation of technical efficiency scores, in the first step, and regressions to relate efficiency scores to contextual factors in the second. On one side, Simar and Wilson (2007) argue that efficiency scores are serially correlated and proposed a seven step double bootstrapping procedure to produce consistent estimates in the second stage. While this approach has been adopted in the literature, it has not received general consensus. McDonald (2009) argues that it is valid only under the proposed data generating process and not robust to reasonable departures from it. Moreover, Banker and Natarajan (2008) provide statistical foundation for the simple two-stage procedure. Their simulation results indicate that a two-stage DEA based approach performs better than a commonly adopted set of one-stage and two-stage parametric procedures. However, hypothesis testing is not discussed. Given the computational complexity of Simar and Wilson (2007) approach, the drawbacks identified by McDonald (2009) and the arguments proposed by Banker and Natarajan (2008), I opted for a simple two stage procedure which has also been extensively adopted in the literature. The variables considered are reported in table 2.4 together with the descriptive statistics. Human capital endowments are represented by the age and education of the household head and by the ratio of skilled members over overall adult family members. The regressions also include a dummy variable indicating whether the household has the title deeds on the land. Finally the employment rate at municipality level is intended to partially capture the presence of transaction costs and the degree of development of the local labour market. The employment rate has been constructed using data from the 2001 South Africa population census on 10% of total population. All regressions include district dummies to control for variations in environmental conditions and soil quality.

Table 2.4: Descriptive statistics of variables used in the efficiency analysis

Variables	Obs	Mean	Std. Dev.	Min	Max
Access to labour market and liquidity					
Share off-farm income	547	30.89	34.27	0	100
Household eligibility (HE)	547	0.39	0.49	0	1
Employment rate (municipality level)	547	42.96	13.32	20.86	73.68
Household characteristics					
Gender of household head (male)	547	0.51	0.59	0	1
Land title	547	0.28	0.45	0	1
Human capital					
Age of household head	547	54.6	14.07	18	96
Education of household head	547	5.48	4.82	0	20
Ratio of skilled adults	547	0.24	0.22	0	1

A particular focus is given to the role of credit market imperfections in limiting the capacity of households to produce at higher levels of technical efficiency. One of the main sources of liquidity is off-farm income which plays an important role in household income formation. Therefore, the share of off-farm earnings on overall household income is included as an additional explanatory variable. Another important source of liquidity considered in the analysis is the receipt of a pension. The OAP in South Africa provides an unconditional cash transfer to all women over age 60 and all men over age 65. The program has been found to be effective in reaching poor households in rural areas and constitutes the basis of credit facilities in local markets (Ardington et al., 2009). The transfer is expected to have a relevant impact on household behaviour[13] given its generosity. In Case

[13]Several studies have investigated the effects of the OAP on children health (Duflo, 2003), household structure (Edmonds et al. (2005), Maitra and Ray (2003)) labour supply (Bertrand et al., 2003; Posel et al., 2006; Ardington et al., 2009) and education (Edmonds, 2006).

Table 2.5: Descriptive statistics on pension receipt and eligibility

Age groups	% receiving the pension	Age groups	% non receiving the pension
Male members			
50-55	1.61	65-70	41.67
55-60	2.13	70-75	19.23
60-65	12.00	over 75	20.00
over 65	72.86		
Female members			
45-50	1.35	60-65	22.64
50-55	3.80	65-70	15.25
55-60	4.17	over 70	5.95
over 60	86.54		
Households with an eligible member			39%
Households with an eligible man			11%
Households with an eligible woman			34%

Source: author's calculation from 2004 KIDS Survey

and Deaton (1998) the authors find that the transfer is about twice the median per capita income of an African household. The baseline model for the analysis of technical efficiency is the following:

$$TE_i = \alpha + \beta\mathbf{X} + \delta P_i + \gamma O_i + \epsilon_i,$$

where TE_i indicates the technical efficiency scores estimated using the DEA method, $\mathbf{X}$ is the vector of contextual variables described above, O_i represents the share of off-farm earnings on total income and P_i indicates that there is a person receiving a pension in the household.

Because pension take-up could be an endogenous household decision it generates a potential source of endogeneity. Therefore, I consider pension eligibility rather than actual pension receipt. The current South Africa Old Age Pension program is the result of the extension to the black population of the white social pension system established during the apartheid. The means test applied to the pension does not exclude most of the African households. The monthly pre-means test transfer in 2004 is of 740 RAND. Individuals in the sample receive an average pension transfer of about 719 RAND which suggests that, in most cases, the means test is not effective. Moreover, because it is not based on family income but only on recipient wealth there are no incentives to pre-pension arrangements.

This implies that pension eligibility depends only on the age of the recipient rather than on past earnings or household composition. Household members are eligible at age 60 if female and 65 if male. About 40% of the households in the sample has a pensioner member. The take-up rate is around 87% for women and 73% for men as reported in table 2.5. This ensures that the eligibility indicator is a good approximation of pension receipt. The estimated equation is therefore:

$$TE_i = \alpha + \beta \mathbf{X} + \delta HE_i + \gamma O_i + \epsilon_i,$$

were HE_i indicates the presence of an eligible member in the household. This eligibility indicator, however, could also capture age trends or differences in background which could intensify or vanish the actual effect of the pension. This chapter allows for differences in household technical efficiency with the age structure of the household by controlling for the age of the oldest man and woman in the household and for the presence of adult male and female members close to the eligibility age. This is done by including dummies indicating the presence of female and male members over age 50 and 55 and male members over 60[14]. Pensions in South Africa have been found to affect household composition. Edmonds et al. (2005), exploiting the age-discontinuity in the structure of the pension program, finds and increase in the number of children aged 0 - 5 and in the number of women aged 18 - 23 and a decrease in the number of women aged 30-39 associated with pension receipt. In this chapter, a higher number of children in the household, for example, could lead to a lower household technical efficiency since more time is needed for children rearing and could therefore offset the possible benefits of having a pensioner in the family. To control for household living arrangements due to pension receipt the regressions include variables representing household size by age categories.

[14] This strategy has also been used in Duflo (2003) and Edmonds (2006).

Additionally, because pension take-up differs from pension eligibility and varies between men and women the effect of a pension receipt could be underestimated. To address this issue I also report the results obtained by instrumenting the variable indicating the presence of a pensioner $\widehat{PENS}_i$ using the number of eligible female and male members in the household as reported below:

$$TE_i = \alpha + \beta \mathbf{X} + \delta \widehat{PENS}_i + \gamma O_i + \epsilon_i.$$

The model has been estimated using a standard linear regression model and a two stage least squares estimator. The choice of this estimator, in contrast with the wide use of tobit models for the analysis of the determinants of efficiency, is motivated by the fact that technical efficiency scores should not be considered as censored values since they are not supported by a latent model. These efficiency indexes are better described as the result of a normalisation process imposed to ensure an unique solution to the linear programming model[15]. Efficiency scores are therefore not truly censored data and are better defined as fractional data (McDonald, 2009). The causal interpretation of tobit results relies on distributional assumptions, namely normality and homoschedasticity of the error term of the latent model. Because of the fractional nature of the technical efficiency variables, the variance of the error term depends on the limit of the dependent variable ($TE_i = 1$)[16] and therefore on the regressors (McDonald, 2009). This implies that the error term is heterosckedastic and White's standard errors need to be computed for valid hypothesis testing using ordinary least squares (ols). In contrast, tobit estimates are inconsistent in the presence of heterosckedasticity. Moreover, Hoff (2007), comparing

[15]The use of a tobit is also justified when the outcome is a corner solution which, however, is not the case when considering efficiency scores.

[16]This is because no households scores zero in terms of technical efficiency. Considering the following linear model: $y_i = x_i\beta + \epsilon_i$ where y cannot exceed 1, when $y = 1$ then $\epsilon = 1 - x_i\beta$ with probability $p(y = 1)$. Therefore the variance of ϵ involves a term related to the probability that $y = 1$ and therefore depends on x_i. The error term will usually be heteroschedastic.

tobit and linear regression results, finds that the latter is sufficient to represent second step DEA models[17]. In the instrumental variable estimation (last specification), a dummy variable indicating the presence of a pensioner in the household is instrumented using a two stage least squares procedure. The use of a probit or logit model in the first stage would lead to inconsistent results unless the first stage model is correctly specified. On the other side, conventional two stage least squares models are consistent independently of the non-linearity of the first stage (Kelejian (1971), Angrist (2001)). Therefore, I opted for the use of a standard two stage least squares estimator that relies on a broader set of assumptions.

2.6.1 Results

The analysis of the determinants of household technical efficiency shows a positive effect of income diversification (table 2.6). This result is robust throughout all the specifications. A rise of 10 percentage points in the share of off-farm income produces an increase of 0.02 to 0.04 units in technical efficiency. In other terms, the rise can cause a household in the bottom of the distribution to lift to a higher decile of the distribution of technical efficiency. This effect can indicate an easing of the liquidity constraint that allows the household to undertake efficiency enhancement purchases or to overcome the entry barriers to more profitable activities. At the same time, it could also signal the presence of positive knowledge spillovers from off-farm to farming activities. These positive effects can be reduced if off-farm opportunities subtract time for farm management preventing the adoption of management-intensive innovations that could explain the relative small magnitude of the coefficients. At this stage of the analysis it is not possible to disentangle the alternative mechanisms through which income diversification affects technical efficiency.

[17] This is done by evaluating the Spearmans rank correlation coefficients between the predicted and actual efficiency scores using alternative methods including ols and tobit models.

Neverthless, the positive sign on the pension eligibility coefficient confirms the presence of a positive liquidity effect. This effect is relevant. Having a pension eligible person in the household causes an increase in technical efficiency between 0.097 and 0.19 units that is quite large if compared to the fact that 25% of the households report a technical efficiency score lower than 0.10. Because the estimates of household technical efficiency consider household members (in terms of adult equivalent) as inputs of production, in the presence of a non binding liquidity constraint, a positive income effect would induce the household to consume more leisure and reduce labour supply leading to a lower overall household technical efficiency. In the presence of a binding liquidity constraint, instead, access to liquidity can, for example, allow farmers to adopt new technology packages that can shift the production surface (Carter, 1989). Because differences in the quality of inputs contributes to the overall technical efficiency, the purchase of more costly high-yielding seeds can shift the entire input-output relationship and lead to higher efficiency.

Column 4 reports the results of the two stage least squares estimation where the variable indicating the presence of a pensioner is instrumented using the number of eligible male and female members. The first stage regression, reported in the appendix, is strong with a very large t statistic (above 80). The results reported in column 4 (table 2.6) confirm the findings when differences between pension take-up and eligibility are taken into account. The effect of pension receipt is larger than that previously found, suggesting that the effect of the pension could have been underestimated due to the difference between pension eligibility and actual pension receipt. The results also show a positive effect of the employment rate at district level. This indicates that better access to job opportunities can improve the efficiency of households.

Although remittances may constitute an additional source of liquidity, they are excluded from this analysis. A potential omitted variable bias problem could arise because of a correlation between remittances and the receipt of the pension. If, for example, pen-

Table 2.6: Analysis of household technical efficiency

	OLS			IV
	(1)	(2)	(3)	(4)
Share off-farm income	0.002***	0.002***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Household eligibility	0.097*	0.096*	0.098*	0.188**
	(0.054)	(0.056)	(0.057)	(0.078)
Gender of household head (male)	0.007	0.034	0.089*	0.093**
	(0.032)	(0.042)	(0.046)	(0.045)
Age of household head	-0.003	-0.002	0.002	0.000
	(0.002)	(0.002)	(0.002)	(0.003)
Education of household head	0.002	0.002	0.002	0.003
	(0.004)	(0.004)	(0.004)	(0.004)
Ratio of skilled over adult members	0.089	0.076	0.067	0.062
	(0.090)	(0.090)	(0.085)	(0.083)
Title on land	0.040	0.039	0.037	0.043
	(0.037)	(0.036)	(0.035)	(0.035)
Employment rate (district level)	0.003**	0.003**	0.003*	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)
Adults dummies		Yes	Yes	Yes
Age of oldest members			Yes	Yes
District dummies	Yes	Yes	Yes	Yes
Observations	547	547	547	547

Robust standard errors in parenthesis. All regressions include a constant, household size by age categories (0-5, 6-14, 15-29, 30-49 and over 50) and indicators of access to water and electricity. *, ** and *** indicate significance at the 10%, 5%, and 1% level of significance

sion receipt produces a crowding out effect on remittances, the coefficient of the pension eligibility variable would be downward biased. On the other hand, if the receipt of a pension facilitates migration, financing job searching, two possible effects are expected. If it does not result in an increase in remittances, then the omission of remittances from the analysis should not bias the results. If, instead, remittances do increase, the effect on technical efficiency is expected to be similar to that of the pension and can be interpreted as an indirect liquidity effect of the latter. Migrants sending remittances are equally distributed across households, with 41% of them in households with an eligible member. Therefore, the exclusion of remittances from the analysis, which are likely to be subject to a measurement error, does not significantly affect the interpretation of the results.

In the regressions presented so far the share of off-farm income on total household income has been considered an exogenous regressor. However, labour allocation decisions can be simultaneous to household efficiency that can influence the selection into off-farm

Table 2.7: Estimations dealing with the endogeneity of the income diversification index

	IV	OLS	
	(1)	(2)	(3)
Share off-farm income	0.004*		
	(0.002)		
Household eligibility	0.151*	0.142*	0.096*
	(0.078)	(0.084)	(0.056)
Gender of household head	0.027	0.118	0.070
	(0.078)	(0.085)	(0.046)
Age of household head	-0.004	-0.009**	-0.003
	(0.004)	(0.004)	(0.003)
Education of household head	0.006	-0.009	0.001
	(0.007)	(0.007)	(0.004)
Ratio of skilled members	-0.077	-0.033	0.041
	(0.139)	(0.167)	(0.093)
Title on land	0.066	0.093	0.049
	(0.053)	(0.063)	(0.037)
Employment rate (district level)	0.004*	0.004*	0.003**
	(0.002)	(0.002)	(0.001)
Second quartile			0.016
			(0.045)
Third quantile			0.08*
			(0.044)
Fourth quantile			0.182***
			(0.048)
Adults dummies	Yes	Yes	Yes
Age of oldest members	Yes	Yes	Yes
District dummies	Yes	Yes	Yes
Observations	193	235	547

Robust standard errors in parenthesis. All regressions include a constant, household size by age categories (0-5, 6-14, 15-29, 30-49 and over 50) and indicators of access to water and electricity. *, ** and *** indicate significance at the 10%, 5%, and 1% level of significance.

activities. Moreover, because off-farm income is also used to compute household technical efficiency estimates, a potential measurement error in reporting off-farm earnings could lead to a spurious correlation between the two variables. I deal with this potential endogeneity problem using instrumental variable technique. Specifically I exploit the information on the share of off-farm income in 1998 for those households observed in both waves of the KIDS survey. Using this instrument, the sample size notably reduces. The results are reported in the first column of table 2.7 and confirm the previous results. However, the potential presence of serial correlation in the error term challenges the validity of this instrument. Although statistically valid and relevant, past participation in off-farm activities can, for example, be correlated with current managerial skills and still leave the problem unresolved. Unfortunately, no better instruments are available. When households participating in off-farm activities are excluded from the analysis, the positive liquidity effect of the pension receipt is still evident, confirming that the potential endogeneity of the income diversification indicator has not affected the other results (table 2.7, second column). Because of self selection issues, however, these latter results are not used for further inference.

Table 2.8: Additional robustness checks

	(1)	(2)	(3)
Share off-farm income	0.002***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)
Household eligibility	0.117*		
	(0.060)		
Gender of household head	0.084*	-0.002	0.034
	(0.047)	(0.032)	(0.039)
Age of household head	-0.004	-0.001	-0.003
	(0.003)	(0.002)	(0.002)
Education of household head	0.001	0.003	0.002
	(0.004)	(0.004)	(0.004)
Ratio of skilled members	0.06	0.079	0.076
	(0.093)	(0.090)	(0.090)
Title on land	0.053	0.041	0.005
	(0.037)	(0.036)	(0.037)
Employment rate (district level)	0.003*	0.003**	0.003**
	(0.002)	(0.001)	(0.001)
Person above eligibility		0.084*	
		(0.045)	
Person below eligibility		0.018	
		(0.045)	
Woman age 45-50			0.120**
			(0.054)
Woman age 50-55			0.087
			(0.063)
Woman age 55-60			0.047
			(0.064)
Woman age 60-65			0.130*
			(0.076)
Woman age 65-70			0.120*
			(0.068)
Woman age 70 and over			0.133**
			(0.072)
Adults dummies	Yes		
Age of oldest members	Yes		
District dummies	Yes	Yes	Yes
Observations	494	547	547

Robust standard errors in parenthesis. All regressions include a constant, household size by age categories (0-5, 6-14, 15-29, 30-49 and over 50) and indicators of access to water and electricity. *, ** and *** indicate significance at the 10%, 5%, and 1% level of significance.

Finally, in the last column of table 2.7, the share of off-farm income does not enter directly in the regression. Instead dummy variables indicating which quartile, in terms of the distribution of the shares, the household belongs to are included. The results show that when the share of off-farm income is above 35%, non-farming earnings have a positive effect on household technical efficiency. Moving from zero off-farm income to earning at least 35% of total income from off farm activities has a significant impact on household efficiency comparable to that of having a pension beneficiary in the household.

Additional checks have been made to further address the concerns about the discrepancy between pension take-up and eligibility. One of the reasons explaining these divergences lies in the potential misreporting of age. It is possible that interviewees report their age, or the age of their family members in rounded decades. If this is the case, it could be particularly problematic since pension eligibility for women starts at age 60. To analyse the influence of a potential measurement error on previous results I run the above sets of regressions excluding those households with women aged 60. The results are reported in the first column of table 2.8 and confirm previous findings. The coefficient of the pension eligibility variable is higher, indicating that the effect of the pension on household technical efficiency could have been underestimated due to a measurement error in the reported age of the women in the household.

To provide additional support to previous results, the age-discontinuity in the pension program structure is recalled to further address the issues of possible confounding effects between pension receipt and age trends. In column 2, the effects of the presence of a household member close to the age of eligibility, namely a man aged between 50 and 64 or a woman aged between 50 and 59, is compared to the effect of having an eligible man aged between 65 and 75 and a woman aged between 60 and 75 in the household. The results show a significant impact of those age groups above the eligibility age, while no effect is found for the presence of adult members below eligibility. Finally, dummy variables indicating the presence of a woman in different age groups - 45-50, 50-55, 55-60, 60-65, 65-70 and over 70 - are included in the regressions. The presence of elderly men in the household is not considered since there are very few male pension beneficiaries. The results show that the effect of an adult woman in the household decreases with her age. However a sharp increase in the size of the coefficient is observed for the 60-65 age group and for the others above the age of eligibility. This non-linearity in the age of the woman cannot be explained by an age effect and is, instead, in line with the fact that a woman becomes eligible at age 60.

2.7 Conclusions

The estimation of household technical efficiency, using a sample of 547 farm households in the KwaZulu Natal Province, has revealed the presence of large inefficiencies. The analysis has been conducted at household-level and off-farm activities have been considered as additional outputs of production. This is motivated by the presence of market imperfections and technical interdependencies between farm and off-farm activities. Household strategies to deal with market imperfections, such as the lack of credit and the presence of transaction costs, are captured in the household-level analysis and contribute to a more comprehensive estimation of technical efficiency. Income diversification is found to increase household technical efficiency. Although it is not possible to establish a prevailing explanation, this can partly be attributed to a liquidity effect. The positive effect of the receipt of a pension from the OAP, in fact, confirms the presence of a binding liquidity constraint. Access to liquidity enables farmers to undertake efficiency enhancement investments and overcome entry barriers to the labour market. These results suggest that institutional reforms to improve the access to the labour and credit markets in the KwaZulu Natal Province could allow a more efficient use of farm household resources.

Chapter 3

Analysing the welfare-improving potential of land in the former homelands of South Africa

3.1 Introduction

South Africa has a large rural population mostly residing in the former homelands. Statistical records reveal that in 1997 about 12.7 million households, that is 31% of the total population, were living in rural areas in the former homelands. Despite the large share of the rural population, 86% of arable land was controlled by large commercial farms while 50% of landed African households had access to less than 1 hectare of land. Although off-farm activities and government transfers are very important sources of income for the rural economy, land-based activities can highly contribute to the overall well-being of rural population by providing a return to family uneducated labour (Carter and May, 1999) and goods and services for home consumption. This argument gains more importance when placed in a broader economic context. According to Lipton and Lipton (1993), South Africa's large endowment of labour suggests the need for more labour-intensive agricultural production that requires a movement toward small scale labour-intensive farming. Eswaran and Kotwal (1986), in fact, show that the utility of landless households is also increased when the distribution of land is moved from a highly unequal setting, with few very large farms, to a scenario characterised by a more egalitarian distribution of smallholders.

Although the effects of land holding extend beyond those on the direct beneficiaries, the analysis proposed here is limited to the relationship between land endowments and household welfare in the rural former homelands. An asset index is used to measure household welfare using two different datasets collecting information on rural farm households. The choice of the indicator is driven mainly by the availability of the data. Nevertheless, the asset index has some advantages over other measures of welfare, which will be explained during the analysis. Moreover, the asset index, constructed using principal component analysis (Filmer and Pritchett, 2001), leads to a welfare distribution that is coherent with that obtained using alternative measures such as income and consumption per capita.

The economic theory of the farm household provides support to a positive relationship between land and household welfare. However, little empirical evidence is available mainly due to the difficulties in identifying the causal relationship between land and a measure of household welfare. This paper investigates the relationship between land endowments and household welfare exploiting historical data on migration to the former homelands. The identification strategy relies on the fact that, since the introduction of the Native Land Act in 1913, African households have been forcibly relocated to the homelands. The year of arrival in the current location is used as instrument for land endowments since later incomers were less likely to have access to land and to larger plots of land given the increasing population pressure in these areas.

Results show the positive effect of land access on household welfare. Land size is also positively related to household welfare so that an increase of 1 hectare is expected to lift the household into a higher decile of the welfare distribution. A set of alternative specifications control for the presence of confounding effects produced by the potential correlation between the year of arrival and the location of the household, the displacement costs occurring in the early years after arrival and the quality of the land. Further checks ensure that the results are robust to the choice of the welfare indicator and of the historical sub-periods characterising the process of segregation of the African population since 1913. The impossibility of distinguishing between voluntary and forced movements, in particular within the homelands where the first are more likely to occur, however challenges the validity of the instrument. This issue is partially addressed with the support of information on the district of previous residence contained in one of the survey adopted and in the 1996 population census.

This paper proceeds as follow. Section 2 provides an overview of the existing literature investigating the relationship between land endowments and household welfare from both a theoretical and empirical perspective. This is followed by the description of the two datasets used in the analysis and a discussion of the main characteristics of the households in the sample. Section 4 introduces the historical setting underlying the identification strategy proposed in this paper with a focus on the main events and aspects characterising the massive forced removals of the African population conducted during the apartheid legacy. Section 5 outlines the empirical strategy adopted and section 6 and 7 discuss the results. Finally section 8 concludes.

3.2 Land and household welfare: theory and existing empirical evidence

Several authors have highlighted the importance of land in contributing to the livelihoods of the rural South African population in both financial and social terms. Most households, for example, derive a direct use value from land-based activities from the provision of goods and services associated with livestock, foods harvested and natural resources for home consumption and for exchange with other goods and services.

The theoretical framework underlying the economic theory on land and household welfare is mainly based on the standard microeconomic theory of the farm household developed by Singh et al. (1986). The household farm is considered a unitary decision unit in which both the consumption and the production side are taken into account. The focus on the household rather than the farm unit is particularly relevant in the presence of market imperfections since consumption and production decisions are jointly determined. Eswaran and Kotwal (1986) and Finan et al. (2005), for example, use a farm household model with imperfect credit and labour market conditions and where access to credit increases with land size. This is based on the argument that larger farms have better access to credit through the collateral use of land. Eswaran and Kotwal (1986), show how household labour allocation decisions are determined by land endowments and that a transfer of working capital, including land, from larger to smaller farm households can be welfare and output improving. Finan et al. (2005) show how household income is positively affected by land endowments through a direct effect (the income generated by the increased production) and an indirect effect when labour and credit markets are imperfect. The magnitude of the overall effect varies across households, in particular, depending on whether the increased demand for inputs is matched by an increased availability of credit due to use of additional land as collateral. In the same vein, Burgess (2001) uses a theoretical household model where land generates a twofold effect on household welfare. Considering imperfections in land and food markets, the author shows that land has the potential to increase household consumption through an income effect, due to increased production, and by providing a cheaper source of food to the household.

Although the economic theory of the farm household gives support to the positive relationship between land and household welfare, with heterogeneous features across households, there is little empirical evidence mainly due to the difficulties in identifying the causal relationship between land and a measure of household welfare. Finan et al. (2005) analyse the impact of land on household welfare, measured using an asset index, using data on Mexican rural households for the period 1997-1998. They propose a linear and a non-parametric specification to capture the non-linearities in the relationship between land and household welfare. Although the study provides an extensive and rigorous analysis of the heterogeneous correlation between land and welfare across households, little attention is paid to the identification of the causal relation between the two. The authors find that land has a very high marginal welfare value for small farms (less than 1 hectare), and that this effect decreases with land size. Burgess (2001), using data on Chinese households, investigates the relationship between land size and household welfare measured by food consumption expenditure and calories intake. Considering a standard food demand equation, once per capita income has been controlled for, the additional effect of land per capita is to be attributed to its role in providing cheaper food to the household. The effect of land is identified by the fact that land in China is distributed on the basis of household nutritional needs, and therefore of household composition, that is beyond households' discretion given the strict family planning policies.

The majority of papers look at the impact of land transfers obtained through the implementation of land reforms. Besley and Burgess (2000), for example, using a panel dataset on sixteen Indian states for the period 1958-92 found that post-independence land reforms positively contributed to poverty reduction. The potential endogeneity of the land reform variable is addressed by using the composition of past political legislatures as instruments for land reform transfers. Other papers are particularly relevant for the present study given their focus on South Africa land reforms implemented since 1997. Keswell et al. (2010) exploit the quasi-experimental setting of the Land Redistribution and Agricultural Development (LRAD) program, introduced in 2001, and find a positive effect on household consumption for the beneficiaries. The impact is identified by comparing households still in the process of being granted the land transfer and households that have already received it. A previous paper by Valente (2009) looks at the impact of LRAD program on household food security. The results show that the land reform has not been successful in reducing the food insecurity of the beneficiaries. This is mainly attributed to the high displacement costs, since in most cases the assigned land is located far from household's current location, and the lack of organisation. The author uses alternative techniques to deal with observed and unobserved variable biases, although no suitable instruments were available to fully address the endogeneity of the land reform variable.

The existing empirical literature confirms the difficulties in identifying the causal relationship between land and household welfare given the non-random allocation of land and the lack of suitable instruments. In this paper, I will attempt to address this empirical issue by exploiting historical data on household migration to the homelands.

3.3 Data

The data used in this analysis are drawn from two different datasets: the KwaZulu-Natal Development Indicators Household survey (KZN-DIHS) of 1996 and the Rural Survey of 1997. These datasets are the only available datasets that provide information on both land and migration history. I opted for the use of two surveys mainly because neither of them provides exhaustive information for the purpose of this analysis. The Rural Survey 1997 provides data on the amount of land available to the household, the initial focus of this analysis, and detailed information on farming activities. However, it does not provide information on location (distance to the nearest town) and on the previous district of residence, although it does report the year of arrival. This latter information, in particular, is useful to narrow the focus of the analysis to provide further support for the use of the estimation strategy adopted in this study, as will be explained in the following sections. The KZN-DIHS 1996, instead, provides more detailed information on migration to the homeland but is confined to a much smaller sample and provides information only on whether the household has access to land with relatively less information on farming activities. The use of both surveys, therefore, allows me to conduct some exercises to support the instrumental variable strategy adopted in this study. However, because the information on land provided by the two surveys is now of two types, a binary variable indicating access to land and a continuous variable indicating the amount of land available, the analysis is conducted separately for each dataset and is described is section 3.6.1 and 3.6.2.

The KZN-DIHS has been conducted by the KwaZulu-Natal Provincial Government and Human Sciences Research Council (HSRC). The complete survey covers 6500 households across the province of KwaZulu-Natal, which incorporates the former homeland of KwaZulu. The sample size has been reduced to consider only the households living in rural areas. This cross section survey has been used mainly because it provides information on both the year of arrival and the previous district of residence, which makes it possible to establish whether a household has moved from a non-homeland area. However, it does not provide information on the amount of land owned by the household and only reports whether the household has access to land for agricultural purposes. The survey provides detailed information on household living conditions and asset ownership that are useful for the construction of a welfare index. It also provides information on household consumption that will be used in one of the empirical specifications proposed below.

The analysis of the impact of land size on household welfare uses the Rural Survey 1997 conducted by Statistics South Africa, which collected information on 6,000 rural households located in the 10 former homeland territories. This cross-section survey provides information on the hectares available to each household for farming purposes, although less detailed information is available as far as asset holding, income or consumption are concerned. Another drawback of this survey is the lack of information on the previous district of residence, as previously mentioned, so that it is not possible to distinguish between movements to and within the homelands.

Table 3.1: Descriptive statistics of the main variables of interest

	(1) KZN-DIHS 1996 Mean (sd)	(2) Rural Survey 1997 Mean (sd)
Land (dummy)	0.38	0.65
	(0.49)	(0.47)
Hectares of land (landed households)		1.41
		(3.57)
Education household head (dummy)	4.07	4.66
	(3.37)	(4.63)
Age of household head	49.25	56.25
	(14.23)	(16.22)
Gender of household head (dummy)	0.75	0.51
	(0.43)	(0.50)
Pension eligible members (dummy)	0.28	0.42
	(0.45)	(0.49)
Children	1.61	2.97
	(1.49)	(2.01)
Number of skilled members	0.80	1.08
	(1.15)	(1.22)
Number of unskilled members	2.20	2.00
	(1.45)	(1.31)
Average eduction of adult members	5.21	6.58
	(3.16)	(3.49)

Author's calculation using the KZN-DIHS and the Rural Survey.

The summary statistics of the variables used in the empirical analysis and reported in table 3.1 offer a general picture of the main characteristics of the households that are the focus of this study. According to the KZN-DIHS, 38% of the households living in rural areas in the KwaZulu Natal province have access to land. Among rural households residing in the former homelands, 65% of those interviewed by the Rural Survey have access to a plot of land. Plots are in general small with an average size of 1.41 hectares, so that as a consequence, only 10% of them produce farm products for sale while the majority work the land to provide food for home consumption. The average household size is between 4 and 5 members. Adult members have on average 5 to 6 years of education, less than the 9 years of compulsory education introduced in 1996.

3.4 Historical background

Segregation in South Africa started to take shape with the introduction of the Natives Land Act in 1913 stating that black Africans were no longer to be able to own or rent land outside designated reserves. During the apartheid era, which officially started in 1948, the reserves were converted to bantustans or homelands and later some of them into 'independent' states within South Africa. The population was classified into four racial groups ('black, 'white, 'coloured, and 'Indian). From 1958, the black population was deprived of its citizenship, legally becoming citizens of one of ten tribally based self-governing homelands: Lebowa, QwaQwa, Bophuthatswana, KwaZulu, KaNgwane, Transkei, Ciskei, Gazankulu, Venda and KwaNdebele. African people were only temporary resident in the remaining territories for as long as they offered their labour there.

Residential areas were segregated, often by means of forced removals. According to Desmond (1971), the governments object was to return 5% of the African population from the white areas to the homelands every year. Several laws regulated the movements of the African population. The Pass law, introduced in 1923, stipulated that the black population should carry pass books when outside the designated homelands. Several influx controls were introduced to limit the number of African people allowed to live and work in white areas (Platzky and Walker, 1985).

People were relocated from 'white farms, from 'black spots' (area of black settlement surrounded mainly by 'white farmers), from small town locations and from metropolitan areas. Removals were initially conducted by direct intervention of government authorities also through arbitrary searches and checks. Later, after 1980, the public emphasis was on people moving 'voluntarily. Removals were, however, the results of indirect coercion by the authorities and the security police through intimidations and threats of arrest and detention (Platzky and Walker (1985), pp 152-76). In many townships and rural areas, for example, new construction was frozen; hospitals, schools and other public facilities for the black population were relocated to the homelands. This was a deliberate tactic to enforce voluntary removal to the homelands (Murray, 1987).

There are no official records of removals and often statistical data were deliberately hidden. However, according to Platzky and Walker (1985), the process of forced removals affected some 3.5 million people in the period 1960-1982 excluding those households forcibly removed within the homelands due to the implementation of the betterment plans described below. Desmond (1971) provides the first attempt to document forced relocations, his narrative description of removals is the results of months spent travelling throughout the country. Simkins (1983) provides some quantitative estimates of population changes and movements for the year 1950, 1960, 1970 and 1980 and estimates a net inflow to the homelands of about a million people in the decade 1960-1970 that had its counterpart outflow mostly in the rural areas outside the homelands.

Figure 3.1: Distribution of arrivals in the homelands

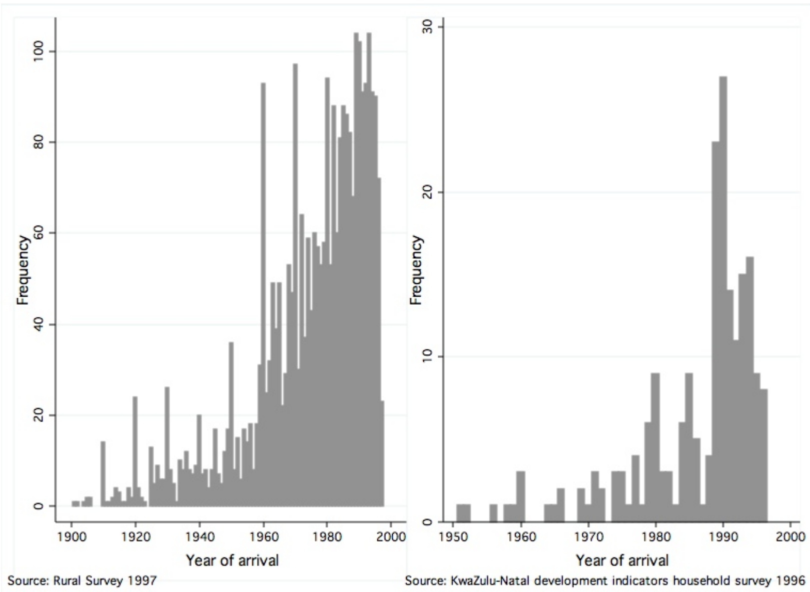


Figure 3.1 plots the frequency of arrivals in the current location for households living in the homelands using information from the two household surveys. While it shows that the date of arrival is not always accurately reported, given the high frequency of rounded decades (this issues will be addressed later in the empirical analysis), it shows an acceleration

of movements in the early 1990s. This is in line with the fact that evictions accelerated in this period partly in response to commercial farmers' concerns about legislation intended to improve the security and working conditions of their workers (Lyne and Darroch, 2004).

The process of forced relocations also continued within the homeland territories. According to Freund (1984) the initially scattered structure of the homelands, appearing as demarcated islands within South Africa, required a consolidation program that produced another massive wave of removals. Even after this process, because the population within the homelands was of heterogenous ethnic background, an additional reshuffling of people was conducted. Moreover, a series of 'betterment plans' were implemented from the 1930s onwards to control land usage, which are considered to have produced the numerically largest and most widespread form of resettlement in South Africa. de Wet (1994) argues that if within-homeland relocations are considered, at least seven million African have been resettled for political purposes since 1913. Under this program, designated areas were divided into distinct land use zones: residential, arable and grazing areas. Land regarded as unsuitable for cultivation was removed from use, so that in some areas people were left with less arable land than they had before or they lost their arable land altogether (de Wet, 1987). Finally, households were also removed for strategic and infrastructural reasons, for example to make space for dam projects (Woodstock and Upper Tugela) or for the clearance of South African borders (Freund, 1984). Finally, it is worth noting that forced removals "did not follow a pre-determined and predictable blueprint. Potential victims could not entirely count on the next move of the state" (Freund, 1984) since government removals plans often appeared in contradictory forms in different official publications.

3.4.1 Forced removals and access to land

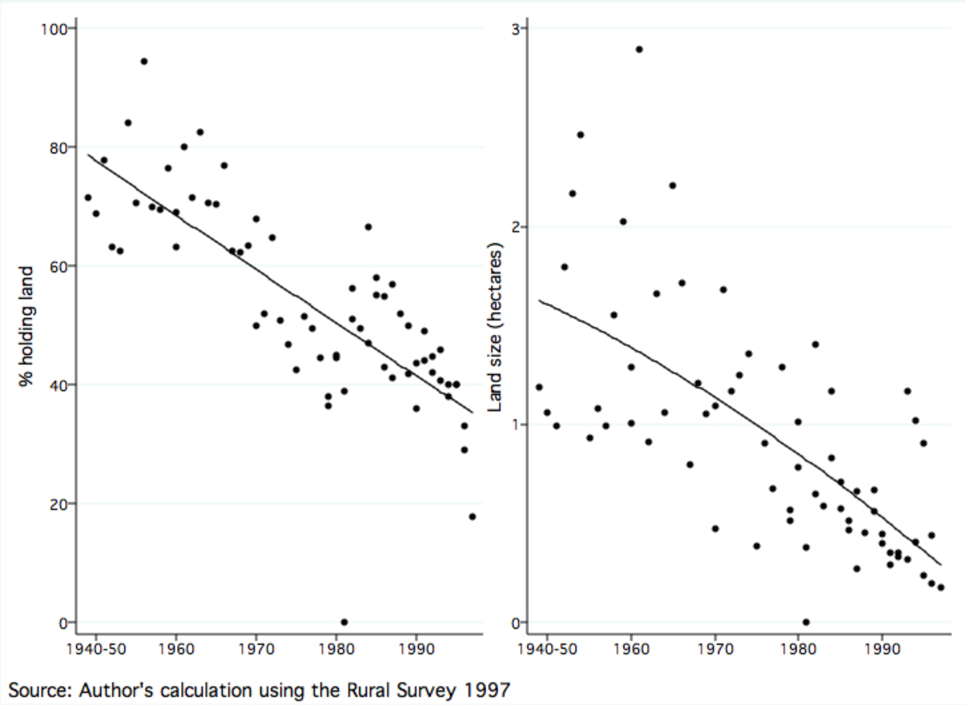
The relationship between removals and access to land is to be found in the increasing population density in the former homelands. The total population density for South

Africa almost doubled between 1970 and 1995, from almost 19 people per square km in 1970 to 34 people per square km. The situation was more dramatic in the homeland areas that constitute less than 14% of African territory and hosts a large share of South African population. According to Simkins (1983), while 39% (of a total 'black' population of 11 million) were living in the homelands in 1950, 53% (of 21 million) were in the homelands in 1980. The forced removals and settlement planning were major contributors to the overcrowding in the homelands. In the Qwaqwa homeland in 1983, for example, after a period of massive relocation of people, its population density was estimated to be over 1,000 people per square km, from a population of 24,000 in 1970 to a population of 400,000 in 1983 (de Wet, 1994).

The increased population density in the homeland areas inevitably led to increasing pressure on the available land for farming and residential purposes so that those arriving later in the homelands were less likely to have access to land and particularly to larger plots of land. Using data from the Rural Survey 1997 it is possible to see these patterns in land endowments. Figure 3.2 shows the negative relationship between both land access and size and the date of arrival in the current location. This negative correlation is at the basis of the identification strategy adopted in this study.

The two surveys were conducted in 1996 and 1997, two and three years, respectively, after the end of the apartheid. Although land distribution has been a major concern since 1995, the first period was mainly characterised by policymaking, consultation and the building of institutions for the delivery of a land reform. Government strategies for reconstruction and development became part of South Africa's Constitution later in 1996 and the final policy framework, the White Paper on South African Land Policy, was implemented in 1997. The available data on land from the two surveys used in this study, are therefore most likely to be unaffected by post-apartheid land reforms. This constitutes an advantage for this analysis since before the implementation of the land reforms, land

Figure 3.2: Percentage of household holding land and land size by year of arrival



endowments can be better predicted using historical information on migration to the homelands.

In general, movements to the homelands after 1913 can be attributed to forced removals through coercive actions, intimidation and pressure by the public authorities and security police. Case studies discussed in Platzky and Walker (1985), the narrative evidence reported in Desmond (1971) and other anecdotal evidence, in fact, suggest that no households would voluntarily move to the overcrowded and unpleasant homelands. The homeland of residence is also determined by the government according to the language spoken or the ethnic group to which the people apparently belong (Platzky and Walker, 1985), and it is, therefore, excluded from household decision-making. An important distinction need to be made between the timing of relocations and the fact of being removed. The empirical analysis proposed in this chapter considers only those households that report to have moved to the current location during the period 1913-94, therefore, although the results may not be generalised to the entire population of the homelands they are not driven by systematic differences between original inhabitants and new incomers. As far as the timing of relocations is concerned, a specific time pattern cannot be identified since relocations from white rural areas overlapped with removals from urban areas, black spots, sites allocated to strategic infrastructures and for 'betterment planning'. Therefore, the year of arrival in the homelands cannot be associated to specific causes or conditions. Moreover, because unobservable characteristics were also likely to be unknown to the authorities that enforced the relocations they are likely to be uncorrelated with the timing of arrivals. These circumstances provide a useful setting to analyse the relationship between land and household welfare exploiting the exogeneity of the year of arrival in the homeland to households' welfare-generating ability and its correlation with land access and size.

3.5 Measuring household welfare using principal component analysis

Household welfare is measured using an asset index. This approach is used mainly to construct a similar measure of welfare across the two surveys, although employing different type of assets, since information on consumption or income is not available in the Rural Survey 1997. Although the choice has been mainly driven by the availability of data, this approach has some advantages. An asset index captures aspects of household welfare that are usually neglected using monetary measures, for example access to basic services such as water and electricity. Moreover, because ownership of assets is easily verified it is expected to be more accurate than consumption expenditure data, for example, which are usually recorded using retrospective recall of information. Given the data available a possible alternative approach would have been the use of the number of assets owned by the households. However this approach give equal weights to all assets and does not taken into account differences in quality. Alternatively, the asset position of the household could also be measured using the value of the assets owned, however asset price are not available in the two surveys used in this study.

The asset index has been constructed using principal component analysis. This approach has been evaluated by Filmer and Pritchett (2001) who demonstrate its suitability for measuring household welfare. Because ownership of different assets is highly correlated across households it is advantageous to collapse information on specific asset ownership into a single new variable (McKenzie, 2005). This artificial variable, W_1, is obtained as the weighted sum of a set of correlated variables, in this specific case variables indicating asset ownership. Given the vector of asset indicators $(x_1, ...x_N)$ where each vector x_n contains observations on each of the N assets for the H household in the sample, the asset index is represented by the following linear combination:

$$W_1 = f_1\Big(\frac{x_1 - \overline{x}_1}{s_1}\Big) + ... + f_N\Big(\frac{x_n - \overline{x}_N}{s_N}\Big), \quad n = 1...N, \tag{3.1}$$

where, $\overline{x}_n$ and s_n are the mean and standard deviation of each asset over all households, therefore, the variables are standardised to have zero mean and unit variance. Weights, f_n, are chosen so that this linear combination has the greatest sample variation. In doing so it maximises the heterogeneity across households so that assets which all or none of the households hold receive small weights, since they do not explain the variation in welfare across households. In practise, this is done by computing the eigenvectors and eigenvalues of the covariance matrix of the standardised x, that is the correlation matrix of the x, defined as C. This can be seen from the maximisation, here reported in matrix form, of the variance of the vector W_1 given a normalisation constraint, $f'f = 1$ where $f = (f_1...f_n)'$. The Lagrangean of this maximisation problem is reported below:

$$L = f'Cf - \lambda(f'f - 1), \tag{3.2}$$

where λ is the Lagrangean multiplier associated with the normalisation constraint. The variance of W_1 is, in fact, given by:

$$\sum_{n,m=1}^{N} f_n f_m C_{m,n} = f'Cf. \tag{3.3}$$

By differentiating equation 3.2 we obtain the following first order conditions:

$$Cf - \lambda f = 0, \quad \rightarrow \quad (C - \lambda I)f = 0, \tag{3.4}$$

Solving the first order conditions is equivalent to finding the eigenvector, f, of the correlation matrix C and the corresponding eigenvalues, λ. The linear combination computed using the eigenvector so obtained is called the first principal component, W_1. Principal component analysis also computes the subsequent $N-1$ components, and the corresponding eigenvectors and eigenvalues, that account for the remaining variance in the sample. They are obtained by solving $N-1$ similar maximisation problems where the variance of alternative linear combinations is maximised subject to the constraint that each combination is orthogonal to the previous, that is: $Cov(W_n, W_m) = 0$ with $n < m$. Commonly, only the first component is retained and is used in this analysis as a measure of household welfare.

Table 3.2: Scoring factors and summary statistics

	(1) KZN-DIHS 1996			(2) Rural Survey 1997		
	Score f	Mean	Sd	Score f	Mean	Sd
Electricity (dummy)	0.27	0.37	0.48	0.22	0.26	0.44
Near water (dummy)	0.14	0.67	0.47	0.04	0.36	0.48
Flush toilet (dummy)	0.34	0.16	0.37	0.03	0.01	0.08
Pit latrine (dummy)	-0.27	0.75	0.43	0.46	0.71	0.45
Other toilet (dummy)	-0.02	0.09	0.29	-0.47	0.28	0.45
Brick structure (dummy)	0.31	0.19	0.39	0.45	0.47	0.50
Traditional house (dummy)	-0.28	0.66	0.47	-0.46	0.50	0.50
Rooms per person	0.14	0.63	0.38	0.22	0.93	0.71
Number of rooms	0.02	2.45	1.23	0.24	4.75	2.52
Own washing machine	0.29	0.23	0.42			
Own washing machine	0.33	0.05	0.23			
Own vacuum cleaner	0.33	0.05	0.23			
Own microwave	0.35	0.07	0.26			
Own car	0.31	0.13	0.33			

The percentage of the covariance explained by the first principal component is 33% in (1) and 29% in (2).

In this analysis, the vector of asset indicators contains dummy variables for the ownership of specific assets (fridge, washing-machine, vacuum cleaner, microwave and car), not available in the Rural Survey 1997, characteristics of the house (brick structure, traditional, type of toilet) and access to utilities (electricity and water), and some numerical variables such as the number of rooms in the house (table 3.2).

The first principal component explains 33% and 29% of the total variance in the data for the Kwazulu-Natal and Rural Survey respectively. For dummy variables the scores reported in table 3.2 can be easily interpreted. A movement from 0 to 1 in one of the asset indicator changes the index by the score divided by the standard deviation. A positive score indicates that the ownership of the asset leads to a higher welfare index. For example, a household owning a fridge has an asset index that is by 0.69 higher than that of a household without a fridge. This is in line with low quality assets being attributed a negative score, as in the case of the traditional-type houses and toilets of different types not connected to the sewer system.

Table 3.3: Descriptive statistics of the asset index by food consumption and income per capita

KZN-DIHS 1996		Rural Survey 1997	
Consumption pc Quartile	Welfare index Mean (sd)	Income pc Quartile	Welfare index Mean (sd)
1	-0.839	1	-0.393
	(1.092)		(1.697)
2	-0.719	2	-0.030
	(1.176)		(1.689)
3	-0.241	3	0.011
	(1.645)		(1.670)
4	2.475	4	0.451
	(3.492)		(1.693)

Regarding the Rural Survey 1997, although the approach uses a reasonable range of assets, the absence of information on household non-agricultural assets such as television, car etc. could lead to an incomplete representation of household living standards. Nevertheless, the asset index constructed using the Rural Survey seems to perform well when compared to an income based measure of welfare. This is reported in the second column of table 3.3 that shows how higher values of the asset index are associated with higher income per capita. Because information on income is provided only by categories, a better check would be against household expenditure, as it is done for the KZN-DIHS. Nevertheless, the comparison still adds some confidence in the use of this asset index as a measure of household welfare.

3.6 Empirical specification

This section outlines the empirical procedure for the estimation of the relationship between land endowments and household welfare. The base empirical specification is the following:

$$w_i = \alpha + \beta A_i + \theta X_i + \epsilon_i, \tag{3.5}$$

where w_i represents the asset index estimated using principal component analysis and X_i is a set of household and district level characteristics that are expected to affect household welfare. These variables include the characteristics of the household head: gender, age and education. The latter two variables, together with variables indicating the highest level of education in the household and the number of skilled members, are expected to capture the human capital contribution to household welfare. Because the pension transfer received from the Old Age Pension Program (OAP) is quite generous for African household and could bias the results if omitted, I also control for the presence of pension eligible members to avoid the potential endogeneity of actual pension take-up. This is done accordingly to the age eligibility criteria of the pension program by including a dummy variable taking the value of one when there is a woman over age 60 and a man over age 65 in the household. Additional controls include the number of children in different age categories, the number of unskilled members and magisterial district level characteristics such as population density, the employment rate and the share of household with access to telephone to measure the level of development of local infrastructures. Further variables are added to address specific empirical issues and will be discussed in the next section. The variable A_i captures land endowments and can be either a dummy variable indicating whether the household has access to land (obtained from the KZN-DHIS 1996) or a continuous variable representing the amount of land available to the household (using the Rural Survey 1997). These two alternative specifications are discussed below.

3.6.1 Dealing with an endogenous dummy variable

When analysing the impact of land access, the variable A_i indicates a dummy variable that takes the value 1 when the household has access to land and is obtained from the KZN-DHIS 1996. Given the binary nature of this variable, the average effect of land on household welfare, conditional on other covariates, can be written as follows:

$$E[w_{1i}|X_i, A_i = 1] - E[w_{0i}|X_i, A_i = 1] = E[w_{1i} - w_{0i}|X_i, A_i = 1],$$

where w_{1i} denotes the welfare of the household if it has access to land and w_{0i} represents household welfare otherwise. Because the second term on the left of this equation is not observed, the above effect, usually defined as treatment effect on the treated, cannot be estimated. Thus a comparison of outcomes between treated and untreated is necessary and is reported below:

$$\begin{aligned} E[w_{1i}|X_i, A_i = 1] - E[w_{0i}|X_i, A_i = 0] &= E[w_{1i} - w_{0i}|X_i, A_i = 1] \\ &+ E[w_{0i}|X_i, A_i = 1] - E[w_{0i}|X_i, A_i = 0]. \end{aligned}$$

This effect involves a bias term $E[w_{0i}|X_i, A_i = 1] - E[w_{0i}|X_i, A_i = 0]$ that differs from zero when certain type of households are more likely to have access to land than others.

This is the case, for example, if households that have experienced relatively unfavourable circumstances in the labour market are more likely to access land. The bias disappears when access to land is independent of household's ability to generate welfare.

One possible option to correct this bias is the use of an instrument that is correlated with the endogenous dummy variable and independent of household welfare potential, conditionally on the other included covariates. The instrument used in this analysis is the year of arrival in the homelands. As reported in section 3.4.1, removals conducted by the government were mostly unpredictable and produced a massive movement of people to the homelands. Given the increasing population density in the homelands following the continuous inflows of relocated households, later incomers were less likely to be given access to land (figure 3.2). This argument underlies the causal relationship between access to land and year of arrival.

Table 3.4: Descriptive statistics of household age structure by decade of arrival

	KZN-DIHS 1996		Rural Survey 1997	
Decade	Household head	Oldest member	Household head	Oldest member
1910			62.333	67.364
			(17.947)	(13.313)
1920			63.543	65.371
			(12.312)	(10.834)
1930			60.734	64.298
			(11.673)	(10.569)
1940			58.105	61.581
			(13.460)	(12.840)
1950	68.571	68.571	59.684	62.538
	(11.013)	(11.013)	(16.102)	(15.584)
1960	57.120	60.080	58.263	61.053
	(11.805)	(13.982)	(15.283)	(14.632)
1970	51.831	52.442	55.989	58.160
	(12.557)	(13.689)	(14.309)	(14.636)
1980	46.373	46.906	51.443	53.770
	(12.257)	(12.856)	(15.342)	(15.708)
1990	41.473	42.068	45.736	47.539
	(13.313)	(14.189)	(15.523)	(16.127)
Total	44.657	45.296	53.454	55.832
	(13.629)	(14.477)	(15.892)	(16.077)

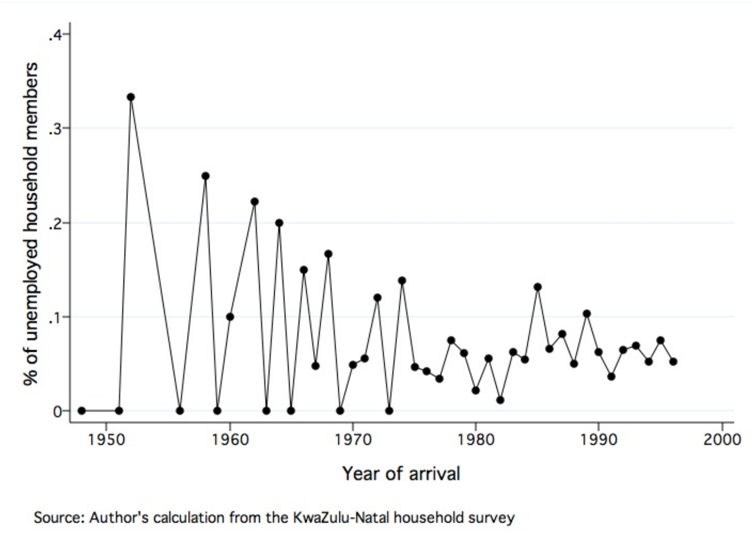
Standard errors are reported in parenthesis.

The independence assumption requires two conditions: that the instrument is exogenous and that it does not affect welfare other than through its effect on land. First the year of arrival in the homelands can be reasonably considered uncorrelated with unobservable household characteristics, since removals where enforced by official authorities with the aim of segregating the entire black population. Moreover, because household unobservable characteristics were also likely to be unknown to the authorities that enforced the relocations they are likely to be uncorrelated with the pace and timing of the arrivals. In addition, because forced removals were mostly unpredictable no pre-moving arrangements could be undertaken. The date of arrival can, therefore, be considered independent of households' welfare potential. Second, the year of arrival seems not to affect household welfare through other channels rather than access to land, once additional control variables, described below, are included. Table 3.4 reveals that households that have moved more recently are in general younger than those that arrived earlier. Because the age structure of the household could itself affect welfare and could be captured by the instrument, regressions include controls for the age of the household head and a polynomial of the age of the oldest member of the household. In addition, the time of arrival could also have affected the location in which the household resides, which could itself have an influence on household welfare. This mechanism can be ruled out by controlling for household road distance to the nearest town as reported in column 6 of table 3.5. An additional issue arises if, for example, later incomers had access to fewer job and business opportunities given the increasing population pressure in the homelands. Because this is likely to affect the probability of finding a job, it could lead to a potential correlation between the year of arrival and households' unobservable ability to generate welfare. Unfortunately it is not possible to control for household-specific employment opportunities. However, when plotting the current average share of unemployed members per household by year of arrival, using the KZN-DIHS data (figure 3.3), there does not seem to be a correlation between the two, suggesting that later arrivals are not worse off in terms of job opportunities[1].

[1]The larger variance for period 1950-70 is due to the lower number of observations.

This can also be explained by the fact that most of the people arriving in the homelands became cross-border commuters, living in the homelands and commuting daily to work in "white" areas (Murray, 1987). Additional controls for local population density and employment rate should also capture the availability of job opportunities in the district of residence. Additional controls are introduced to limit the presence of confounding effects and are discussed in the next sections.

Figure 3.3: Average share of unemployed people in the households by year of arrival (KwaZulu-Natal province, 1996)



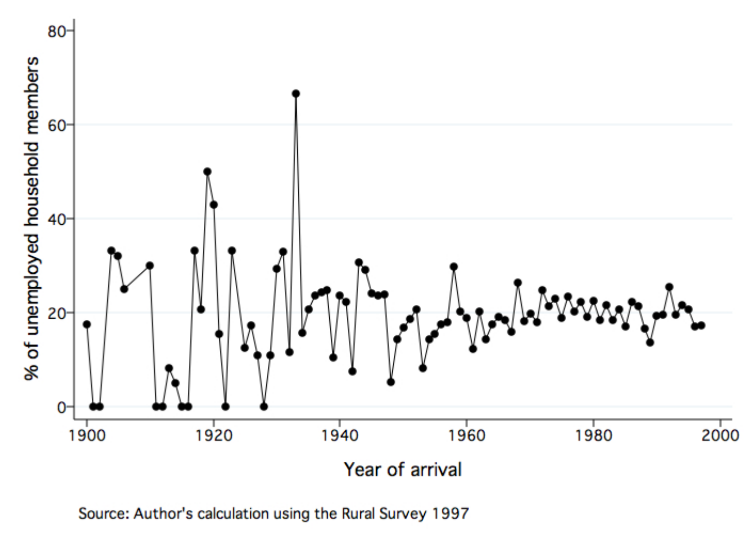
Given the binary nature of the land variable, the first stage regression could be estimated using a nonlinear method such as logit or probit. Angrist (2001) argues that the second-stage estimates will be inconsistent if the the first stage model is incorrectly specified while a standard two-stage least squares (2SLS) procedure can avoid this inconvenience. The model is therefore estimated using 2SLS. It is worth noting that with a dummy endogenous variable, instrumental variables procedures estimate causal effects for those households whose behaviour is affected by the instrument. This is usually known as local average treatment effect (LATE) (Angrist, 2001). That is, the effect is estimated for those households that obtained land because they arrived early but would not have received it if they arrived later, and for those households that did not receive land because they arrived late in the homelands, but would have obtained it if they had arrived earlier (these two groups are known in the literature as compliers). This means that the results are not informative about the effect on those households that would have never had access to land, or those that would have accessed the land independently of their year of arrival. Unfortunately, in practice, it is not possible to establish the size of this subpopulation of compliers.

3.6.2 Dealing with a positive continuous endogenous variable

When analysing the impact of land size on household welfare, A_i is a positive continuous variable indicating the amount of land available to the household that is obtained from the Rural Survey 1997. The majority of the households surveyed received the land from the local or tribal authority and about 82% do not possess title deeds. The absence of a land market could lead to the conclusion that land should be considered as an exogenous variable since households cannot easily adjust the amount according to their needs. However, the presence of unobservable household characteristics that could be potentially correlated with both land endowments and household welfare challenges the exogeneity of the land variable. Unobserved land quality, social status, habits and attitude toward agriculture, for example, are likely to be correlated with both household welfare and access to land. To address this issue I use, also in this specification, the date of arrival in the current location as an instrument for land endowments. Similarly to the argument used above, given the increasing population pressure in the homelands, later incomers were more likely to receive smaller plots of land. The correlation between land size and year of arrival can be observed in the right panel of figure 3.2 where average land size by year of arrival is plotted using data from the Rural Survey 1997. In line with the previous discussion on the independence of the instrument, the year of arrival can be reasonably considered uncorrelated with unobservable household characteristics, since the timing of forced removals cannot be related to households' habits, attitude toward agriculture or welfare-generating ability. By controlling for the age of the household head and of the oldest members in the household, as previously suggested, it is possible to isolate the effect of the instrument on land from that on the age structure of the household. The year of arrival, however, could also be negatively correlated with the quality of the land plot since better quality land could have been assigned to earlier incomers. In this case the instrument would capture both land size and quality inducing a bias in the estimates. Unfortunately, information on specific plot quality is not available, but this issue is partially addressed by including a proxy for land quality at district level, constructed using average maize production per hectare. This specification is reported in table 3.6 column 4.

An additional channel through which the year of arrival could affect household welfare is by capturing displacement costs. Specifically, households, in the first years after arrival, may incur high transaction costs, such as search costs and resettlement costs, in the new living area that could affect their ability to generate welfare. This issue is partially ruled out, since I consider only households that arrived before the end of the apartheid, that is only households that have resided in the current location for at least 3 years. This applies also to the previous specification that deals with access to land. Further support is provided by considering only households that had been living in the current location since 1990. This specification is reported in table 3.6 column 6. Finally, similarly to what has been mentioned in the previous section, when using data from the Rural Survey 1997, the year of arrival does not appear to be correlated with household employment conditions (figure 3.4), suggesting that the instrument does not affect household welfare through its correlation with the availability of local business and job opportunities. Additional robustness checks are conducted and discussed in the next sections.

Figure 3.4: Average share of unemployed people in the households by year of arrival



The model is estimated using 2SLS and the reduced form is the following:

$$w_i = \alpha + \beta \hat{A}_i + \theta X_i + \epsilon_i,$$

where $\hat{A}_i$ is obtained from a first-stage regression where the year of arrival in the current location is used as an instrument. Although the censored nature of the land size variable may suggest the use of a nonlinear first-stage regression, a conventional two-stage least squares model is consistent independently of the non-linearity of the first stage (Angrist, 2001) and is therefore preferred because it gives consistent estimates under a broader set of assumptions.

3.7 Estimation results: access to land and household welfare in the KwaZulu-Natal province

This section reports and discusses the effects of having access to land on household welfare in the KwaZulu-Natal province. The results have been obtained using an initial subsample of 4,368 African rural households provided by the KZN-DIHS 1996 and are reported in table 3.5. The subsample is further reduced once the instrument is used in the 2SLS procedure. The first column reports the results of the ordinary least squares estimation of equation 3.5 and shows a positive correlation between land holding and the welfare level of the household.

As discussed above, a potential endogeneity bias could be driving these results. The other columns, therefore, report instrumental variable (IV) estimates. The instrument for the land variable is the year of arrival in the current location in all the specifications. Households that moved after 1994, i.e. after the end of the apartheid, are excluded from the analysis. The households considered are, therefore, those that moved between 1948 and 1993 since no households reports a year of arrival earlier than 1948. Unfortunately, when using this instrument the sample size is reduced noticeably to about 700 households. The first-stage regressions are strong with an F statistic above 10 (reported at the bottom of the table). In all IV specifications the first stage estimates show that the year of arrival is negatively associated with access to land[2]. The first-stage regression associated with column 2 is reported in table 10 in the appendix. Column 3 includes additional controls for the age of the oldest members of the household. This is done to ensure that the instrument is not capturing the effect of differences in the age structure of the household and are included also in the subsequent specifications. All regressions include district council dummies to control for differences in environmental and local conditions.

[2]Given the availability of only one instrument, it is not possible to test for overidentifying restrictions. However, because any function of the instrument can potentially be a suitable instrument, using both the year of arrival and its square as instruments the model is overidentified. The overidentification test statistic reveals that the null hypothesis of joint validity cannot be rejected and add increases the confidence in the instrument. However it is worth noting that the overidentification test relies on the assumption that at least one instrument is valid. Therefore in this case, if this assumption does not hold for one of the instruments it necessarily does not hold for the other. This reduces the power of the test. Because first-stage regressions are better fitted with the year of arrival only, the results reported here consider only one instrument.

Table 3.5: OLS and 2sls regressions of the effect of access to land on household welfare

	OLS	IV						
	(1)	(2)	(3)	(4)[a]	(5)[b]	(6)[c]	(7)[d]	(8)[e]
Land (dummy)	0.226***	2.554**	2.542*	3.368**	3.806**	3.205*	122.431**	4.273**
	(0.066)	(1.298)	(1.304)	(1.681)	(1.823)	(1.700)	(51.736)	(1.669)
Education household head	0.187***	0.087***	0.093***	0.112**	0.107	0.127**	9.318***	0.038***
	(0.013)	(0.034)	(0.034)	(0.044)	(0.068)	(0.064)	(2.197)	(0.013)
Age of household head	0.025*	0.003	0.074	0.091	-0.066	-0.248	-6.496	0.020
	(0.013)	(0.041)	(0.149)	(0.192)	(0.288)	(0.259)	(7.363)	(0.029)
Gender of household head (dummy)	0.390***	0.387**	0.412**	0.034	-0.360	-0.160	-7.369	-0.011
	(0.061)	(0.190)	(0.190)	(0.323)	(0.682)	(0.636)	(15.851)	(0.147)
Pension eligible members (dummy)	-0.153	-0.148	0.292	0.324	2.078*	1.507	56.745*	-0.167
	(0.099)	(0.324)	(0.432)	(0.544)	(1.106)	(1.022)	(34.392)	(0.146)
Children age 1-5	-0.194***	-0.098	-0.092	-0.197	-0.230	-0.230	-20.046**	0.004
	(0.044)	(0.173)	(0.172)	(0.217)	(0.279)	(0.250)	(9.735)	(0.046)
Children age 6-17	-0.216***	-0.333***	-0.345***	-0.429***	-0.426**	-0.407**	-17.029***	-0.003
	(0.022)	(0.085)	(0.086)	(0.134)	(0.172)	(0.162)	(5.516)	(0.029)
Number of skilled members	0.009	-0.001	0.005	-0.019	0.113	0.231	-11.024	0.066
	(0.040)	(0.141)	(0.141)	(0.189)	(0.250)	(0.235)	(7.419)	(0.049)
Number of unskilled members	-0.354***	-0.526***	-0.505***	-0.512***	-0.810***	-0.746***	-27.983***	-0.114**
	(0.023)	(0.117)	(0.118)	(0.145)	(0.239)	(0.213)	(8.069)	(0.046)
Highest level of education	0.495***	0.450***	0.451***	0.438**	0.303	0.279	3.327	0.072***
	(0.055)	(0.142)	(0.143)	(0.173)	(0.262)	(0.245)	(10.236)	(0.022)
Employment rate (district level)	1.161***	1.851***	1.841***	1.103	-0.699	-1.718	-0.968	0.958***
	(0.186)	(0.580)	(0.582)	(1.076)	(1.642)	(1.549)	(61.172)	(0.208)
Population density (district level)	-0.000	0.003***	0.003***	0.004***	0.004***	0.004***	0.111**	-0.000
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.049)	(0.000)
Road distance						-0.089***	-0.963	
						(0.027)	(1.057)	
District council dummies	Yes	Yes	Yes	No	No	No	No	Yes (Prov)
Age of oldest members	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4372	694	694	496	199	199	198	2541
Kleibergen-Paap F statistics		14.918	14.866	11.033	10.865	10.820	11.064	12.452
Durbin-Wu-Hausman (p-value)		0.012	0.013	0.006	0.030	0.050	0.019	0.000

Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions include the squared age of the household head. Tests of overidentifying restrictions performed using both the year of arrival and its square do not reject the null hypothesis that the instruments are valid. [a] consider only households living in the homeland. [b] and [c] excludes households that moved within the homeland. [d] The dependent variable is per capita food consumption. [e] uses the Rural Survey 1997.

The coefficient of the land variable increases notably when land is instrumented with the year of arrival and the Durbin-Wu Hausman test suggests that 2SLS results are to be preferred to standard OLS. The downward bias of the OLS estimates could be explained by the fact that if more disadvantaged households, in particular those facing unfavourable circumstances in the labour market, choose to engage in agricultural activities, and therefore to access the land, the estimated effect of land on welfare would be small or even negative, as suggested in Carter and May (1999). Therefore neglecting this source of endogeneity would provide a more pessimistic view of the relationship between land and welfare. The instrumental variable estimates reveal that the effect of access to land on welfare is large. The increase in welfare of around 2.5 units is sufficient, on average, to shift a household from the lowest to the top quintile of the distribution of welfare.

In the IV specifications discussed so far, about 70% of the households considered live in a former homeland territory. Column 4 reports the results when only household living in the former homelands are considered. The KwaZulu homeland comprises a large number of non-contiguous parts spread throughout the KwaZulu-Natal province. The province created in 1994 incorporates the former homeland of KwaZulu and the surrounding province of Natal. Households in the sample are assigned to the former homeland on the basis of the magisterial district of residence. The identification of the magisterial districts belonging to the former homeland has been done using the information provided by Cox (2004) and the map that overlaps the KwaZulu homeland borders with magisterial district boundaries reported in Pauw (2005). The results reported in column 4 confirm previous findings. Although the sample size is further reduced the instrument maintains its explanatory power. This subsample, however, still considers both households that moved to and within the homeland, and therefore it may include households that voluntary changed location within the KwaZulu former homeland. This issue is expected to have a limited effect, since a large fraction of within-homeland movements are expected to be the result of government 'betterment planning'. According to Platzky and Walker (1985), in fact, more than a million people have been moved as a result of 'betterment plans' in KwaZulu from 1950 to 1985. To provide further support to the results, however, the estimates reported in column 5 are obtained by further restricting the sample to those households which migrated from non-homeland areas given the lower probability of encountering voluntary migration in this subsample. The coefficient of the land variable remains stable and significant and no relevant changes are observed for the other explanatory variables.

Given the availability of consumption data in the KZN-DIHS survey, column 7 reports the results of the same specification in column 6 where, however, the dependent variable is food consumption per adult equivalent computed using the OECD equivalence scale[3]. When using this alternative measure of household welfare, access to land is still found to have a positive effect. This specification offers the opportunity to provide an economic interpretation of the results. Obtaining access to land has a large effect on household welfare by generating an increase in per-adult equivalent food consumption close to its median value. Finally, the last column reports the results when the same specification is applied to the Rural Survey 1997 dataset. Results appear to be in line with the findings obtained using the KZN-DIHS although some issues related to the Rural Survey dataset need to be further addressed as it is done in the following section.

Although the main focus of this analysis is the impact of access to land on household welfare, some useful insights can be obtained by looking at the effects of the other covariates. Education plays an important role in contributing to household welfare. This is shown by the positive and significant effect, throughout most of the specifications, of the education level of the household head and of the highest educational attainment of the household members. It is also reflected in the negative effect of the number of unskilled

[3]The OECD equivalence scale assigns a value of 1 to the first member of the household, 0.7 to each additional adults and 0.5 to additional children in the household.

members in the households that is also likely to capture the effect of the lack of labour market opportunities for less educated household members.

3.8 Estimation results: land size and household welfare in the former homelands

In this section I explore the relationship between land size and welfare using the subsample of landed households provided by the Rural Survey 1997. I consider only households with access to land since the aim of the analysis conducted in this section is to analyse the welfare-improving effects of land size on households involved in land-based activities excluding those that could be potentially engaged into completely different livelihood strategies (Finan et al., 2005). In the first column of table 3.6, I report the estimates of the ordinary least squares estimation of equation 3.5. The dependent variable is the asset index constructed using data from the 1997 Rural Survey and summarised in the last column of table 3.3. The results show a positive correlation between the amount of land owned and household welfare. The remaining columns report the two-stage least squares estimates and consider only households that moved to the current location during the period 1913-1994. The first-stage regressions show the negative correlation between land size and the year of arrival. The first-stage regression associated with column 2 is reported in table 10 in the appendix. The F statistics reported at the bottom of the table confirms the relevance of the instrument. The results using instrumental variable show that an additional hectare of land, produces an increase of 0.610 in the welfare index, which is sufficient, on average, to cause a shift to a higher decile of the welfare distribution.

Table 3.6: OLS and 2sls regressions of land size on household welfare

	OLS	IV					
	(1)	(2)	(3)	(4)	(5)[a]	(6)[b]	(7)[c]
Land (hectares)	0.009*	0.608**	0.616**	0.612**	0.935**	0.420**	0.380**
	(0.005)	(0.240)	(0.254)	(0.249)	(0.425)	(0.213)	(0.160)
Education of household head	0.071***	0.063***	0.063***	0.063***	0.072**	0.054***	0.059***
	(0.007)	(0.016)	(0.016)	(0.016)	(0.028)	(0.013)	(0.014)
Age of household head	0.028***	0.006	0.003	0.004	-0.008	0.013	0.029
	(0.009)	(0.026)	(0.033)	(0.033)	(0.057)	(0.029)	(0.021)
Gender of household head (dummy)	0.050	-0.259	-0.367*	-0.370*	-0.600	-0.239	-0.385*
	(0.049)	(0.166)	(0.218)	(0.220)	(0.395)	(0.188)	(0.199)
Pension eligible members (dummy)	-0.055	-0.096	-0.065	-0.070	-0.099	-0.066	-0.154
	(0.069)	(0.134)	(0.146)	(0.146)	(0.244)	(0.129)	(0.133)
Children age 1-5	-0.025	0.022	0.024	0.024	0.091	0.019	0.041
	(0.024)	(0.052)	(0.053)	(0.053)	(0.087)	(0.046)	(0.046)
Children age 6-17	-0.013	0.001	-0.001	-0.000	0.039	-0.019	0.004
	(0.016)	(0.032)	(0.032)	(0.032)	(0.058)	(0.025)	(0.028)
Number of skilled members	0.101***	0.067	0.063	0.063	0.035	0.059	0.035
	(0.026)	(0.050)	(0.055)	(0.055)	(0.087)	(0.047)	(0.047)
Number of unskilled members	-0.057**	-0.052	-0.057	-0.057	-0.046	-0.078**	-0.103**
	(0.023)	(0.039)	(0.040)	(0.040)	(0.068)	(0.038)	(0.040)
Highest level of education	0.068***	0.030	0.029	0.029	-0.004	0.053***	0.050*
	(0.011)	(0.034)	(0.034)	(0.034)	(0.059)	(0.019)	(0.029)
Labor market dev index (district level)	0.873***	1.774***	1.786***	1.711***	1.469**	1.426***	1.382***
	(0.110)	(0.479)	(0.495)	(0.421)	(0.655)	(0.363)	(0.318)
Population density (district level)	0.000**	0.001*	0.001*	0.001*	0.002*	0.001	0.001
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Land quality index				0.070	0.323	0.061	0.057
				(0.127)	(0.224)	(0.124)	(0.094)
Province dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age of oldest members	No	No	Yes	Yes	Yes	Yes	Yes
Observations	3775	2736	2736	2736	2136	2359	2078
Kleibergen-Paap F statistics		12.398	11.016	11.381	6.314	12.898	16.647
Durbin-Wu-Hausman (p-value)		0.000	0.000	0.000	0.000	0.021	0.004

Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions include the squared age of the household head. Tests of overidentifying restrictions performed using both the year of arrival and its square do not reject the null hypothesis that the instruments are valid. [a] excludes the homelands of Venda and Transkei, [b] Period 1913-90, [c] excludes rounded decades.

Column 3 reports the results when additional controls for the age structure of the household are included, namely a polynomial of the age of the oldest man and woman in the household. This is also included in the subsequent specifications. All regressions include province dummies to control for differences in environmental and other local conditions. In column 4 a variable capturing the variation in land quality across districts is included and is intended to control for the potential correlation between the year of arrival and the quality of the land, although it does not capture plot-specific quality. The results are in line with previous findings and the coefficient of the land quality index is not significant. This can be explained by the fact that the land in the former homelands is in general of poor quality with little variation within the territory (Desmond, 1971).

Because the reasons for the move are unknown, one of the main concern is that forced removals could be confounded with voluntary migration. Voluntary relocations were more likely to occur within the homelands since, as previously mentioned, conditions in the homelands were extremely unfavourable and the available descriptive evidence suggests that no households would voluntary move to these overcrowded and unpleasant areas. Unfortunately, it is not possible to establish whether the household moved to or within the homeland of current residence since the 1997 Rural Survey does not provide information on the previous place of residence. It is worth noting that massive forced relocations were also implemented within the homeland territories, often motivated by 'betterment plans' implemented since 1930. Therefore movements within the homelands are also likely to be the result of coercive government policies although no direct evidence is available. To further address this problem I use the 1996 South Africa population census which provides information on the year of arrival in the current location and the district of previous residence. Using this information it is possible to exclude from the analysis those areas with the highest percentage of within-homeland movements and, therefore, potentially with the highest probability of voluntary movements. Table 3.8 reports the distribution of movements by homelands and distinguishes between "within-" and "to-" homeland migration. According to these figures, the two former homelands of Transkei and Venda have the highest percentage of within-homeland movements, since 89% and 86% of the households that arrived in the current location during the period 1913 - 1994 were previously residing within the same homeland. In column 5 of table 3.6, households living in these two homelands are excluded from the sample. The estimates reported confirm previous findings although the F statistic of the first-stage regression is now lower due to the reduced sample size. Similar results, reported in table 3.7, are also found when different sub-periods are considered, in particular when households that moved before 1930, 1948 and 1958 are subsequently removed from the sample. These dates correspond to the main events that affected the process of forced removals: 'betterment plans' were introduced in 1930, while apartheid officially started in 1948 and 'black' people were officially assigned to a homeland territory in 1958.

Another potential problem may arise if forced removals have a direct welfare cost, independent of land endowments. As previously mentioned, this issue is partially ruled out by considering only households that arrived before the end of the apartheid, that is only households that have resided in the current location for at least 3 years. A further exercise is conducted by excluding from the sample those households that arrived in the current location before 1990. Although post-arrival tangible and intangible displacement costs can affect household performances and consequently household welfare, it is reasonable to expect that after at least 7 years of residence in the same location the household can overcome the initial difficulties. The results reported in column 6 confirm previous findings. The coefficient of the land variable, although reduced in size is still positive and significant. This confirms that the instrument is not capturing the effect of the displacement costs associated with the length of the residence in the current location.

Because figure 3.1 reveals that households tend to report the year of arrival in rounded

Table 3.7: Additional results on the effect of land size on household welfare

	1930-1994 (1)	1948-1994 (2)	1958-1994 (3)	1913-1994 (4)
Land (hectares)	0.609**	0.537**	0.492**	0.628**
	(0.248)	(0.235)	(0.198)	(0.259)
Education of household head	0.063***	0.062***	0.069***	0.065***
	(0.016)	(0.016)	(0.016)	(0.016)
Age of household head	0.009	0.017	0.020	0.005
	(0.033)	(0.031)	(0.030)	(0.034)
Gender of household head (dummy)	-0.368	-0.385*	-0.324	-0.375*
	(0.227)	(0.229)	(0.204)	(0.226)
Pension eligible members (dummy)	-0.057	-0.049	-0.043	-0.057
	(0.149)	(0.145)	(0.143)	(0.147)
Children age 1-5	0.024	0.029	0.031	0.027
	(0.055)	(0.054)	(0.052)	(0.056)
Children age 6-17	0.005	-0.010	-0.012	-0.002
	(0.033)	(0.032)	(0.031)	(0.032)
Number of unskilled members	-0.044	-0.050	-0.049	-0.060
	(0.041)	(0.042)	(0.042)	(0.041)
Number of skilled members	0.069	0.052	0.040	0.061
	(0.056)	(0.054)	(0.054)	(0.056)
Highest level of education	0.030	0.035	0.035	0.029
	(0.035)	(0.034)	(0.032)	(0.035)
Labor market dev index (district level)	1.741***	1.721***	1.621***	1.744***
	(0.432)	(0.414)	(0.369)	(0.439)
Population density (district level)	0.001*	0.001*	0.001**	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Land quality index	0.065	0.021	0.018	0.103
	(0.126)	(0.126)	(0.118)	(0.154)
Independence (dummy)				-0.281
				(0.307)
Observations	2649	2479	2328	2736
Kleibergen-Paap F statistics	10.872	9.576	10.975	10.829
Durbin-Wu-Hausman (p-value)	0.000	0.001	0.000	0.000

All regressions include also the square of the age of the head of the household. Tests of overidentifying restrictions performed using both the year of arrival and its square do not reject the null hypothesis that the instruments are valid.

decades, I conduct a further exercise by dropping those households that could be potentially misreporting the year of arrival, i.e. those households that arrived in 1920, 1930, 1940, 1950, 1960, 1970, 1980 and 1990. The instrument improves the ability to predict the amount of land held by the household, so that the F statistic of the first stage is now above 16 and the results are again similar to previous findings. Finally, because those homelands that obtained independence, namely Transkei in 1976, Bophuthatswana in 1977, Venda in 1979 and Ciskei in 1981, were rewarded by the government with the building of roads, shopping centres and hotels (Platzky and Walker (1985), p 23) I also run an additional specification (reported in table 3.7) including a dummy variable indicating whether or not the homeland obtained independence and the results are almost unchanged.

Table 3.8: Movements of households in former homelands

Former homelands	% moved within the homeland area	% moved from other areas
Kwazulu	56	44
Bophuthatswana	62	38
KaNgwane	34	66
KwaNdebele	19	81
Transkei	89	11
Ciskei	43	57
Venda	86	14
Ganzankulu	63	37
Lebowa	73	27
Qwaqwa	22	78

Source: author's calculation from the South Africa Census 1996

In line with the results reported in the previous section, the education of the household head positively affects household welfare. The number of unskilled members has a negative effect, although not always significant, probably signalling the presence of constraints in the labour market for less educated household members. The positive and significant effect of the employment rate at district level indicates how the development of the local labour market can positively influence household welfare. Finally, households with a male head are worse off in comparison to female-headed households. This could be related to the fact that, in rural areas, male heads usually tend to migrate to urban centres, and therefore their presence in the household could signal a lack of off-farm sources of income.

Although the paper finds a positive relationship between land endowments and welfare, it is not possible to identify how these effects are transmitted. The high share of households producing mainly for home consumption suggests that land can benefit them by providing a cheaper source of food. However, other mechanisms could be in action and cannot be disentangled without further investigations.

3.9 Conclusions

This paper explores the relationship between land endowments and household welfare. Although economic theory supports a positive relationship between land and welfare, little empirical evidence is available mainly due to the difficulties in identifying the causal relationship between land and a measure of household welfare. The potential endogeneity of the land variable is here addressed using historical data on migration to the former homelands. The availability of data on the year of arrival in the current location reveals a negative correlation between land endowment and arrival date that is in line with records on increasing population pressure, and therefore with increasing scarcity of land, in these areas since the introduction of the Native Land Act in 1913. The year of arrival is expected to be independent of households' unobserved ability to generate welfare. Movements to the homelands, in fact, can be attributed to the massive forced removals conducted by the central government with the aim of segregating the African population into different homelands according to their ethnic background. Movements within the homelands can also be largely explained by government 'betterment planning' for the reorganisation of the territory in the homelands.

The empirical specification adopted in this paper assumes a linear relationship between land size and household welfare that fails to capture the potential non-linear effects of land endowments. Finan et al. (2005), for example, argue that credit and labour market imperfections can affect the ability of the household to maintain production intensity when land area increases. Therefore the relationship between land and household welfare seems to follow a more complex pattern. Non-linear analyses, however, often require non-parametric techniques or non-linear specifications where the presence of potential endogenous explanatory variables requires the use of less conventional and more complex solutions, when possible. Nevertheless the relevance of such heterogenous effects leaves room for further investigations on the relationship between land and welfare across different dimensions of the farm household.

Results show the positive effect of land access on household welfare. Land size is also positively related to household welfare so that an increase of 1 hectare is expected to lift the household into a higher decile of the welfare distribution. A set of alternative specifications control for the presence of confounding effects produced by the potential correlation between the year of arrival and the location of the household, the displacement costs occurring in the early years after arrival and the quality of the land. Results are also robust to alternative specifications. The positive relationship here identified cannot, however, be attributed to one or more transmission mechanisms and again leaves room for further investigations. Nevertheless, these results suggest that reforms aimed at improving access to land, a major concern of post-apartheid governments, have the potential of improving household welfare. Moreover, because the households considered in this analysis are living in relatively disadvantaged and less fertile areas - the homelands - these results are likely to provide a lower bound for the positive effects of land access on household welfare.

Conclusion

This thesis has provided an analysis of household livelihoods in rural South Africa. The focus of the first two chapters is on how liquidity constraints and labour market imperfections affect households' allocation and use of resources. The first chapter explains household labour allocation decisions using a theoretical farm household model developed from that in Sadoulet et al. (1998). The model incorporates a liquidity constraint and transaction costs in the labour market to explain the emergence of three distinct household categories: small-scale peasants, self cultivators and hiring-in households. The model is tested empirically and labour allocation decisions are found to be influenced by liquidity constraints and market imperfections such as transaction costs and limited access to information. The second chapter analyses technical efficiency at household level considering off-farm activities as additional outputs of production. The empirical analysis reveals the presence of large inefficiencies that can be reduced by accessing sources of liquidity such as pension transfers and income diversification. Finally the third chapter explores the relationship between land endowments and household welfare. Results show the positive effect of land access on household welfare. Land size is also positively related to household welfare so that an increase of 1 hectare is expected to lift the household into a higher decile of the welfare distribution. In general, the results reported in this thesis suggest that households' decisions and performances are affected by the presence of market imperfections that limit the ability of the households to optimally allocate, use and accumulate resources. Improving access to the credit, labour and land market can improve households' allocative and technical efficiency as well as their welfare. The analyses conducted in this thesis present some limitations that, although discussed in the text, are here reiterated to restate the motivations that have led to particular methodological choices despite their shortcomings and, in particular, to provide some suggestions and indications for future research.

In the first chapter, the potential endogeneity of some explanatory variables, in particular household composition and land endowments, has not been fully addressed because of the lack of suitable instruments and of the particular estimation approach adopted: a generalised ordered logit that does not easily handle endogeneity issues. This particular methodology has been chosen for its adaptability to the specific theoretical propositions under analysis. This method allows, in fact, the modelling of household-specific exposure to market imperfections, relaxing the assumption of homogenous shadow wages that has usually been adopted in the literature. This choice was taken at the expense of a less conventional approach that cannot be easily adjusted to handle endogeneity problems, therefore leaving room for further investigations.

In the second chapter, the use of a non-parametric technique (DEA) for the estimation of household technical efficiency might raise some concerns about its deterministic nature and its sensitivity to sample variation. There is an ongoing debate on the advantages and disadvantages of this method and its parametric counterpart, the stochastic frontier approach, and the conclusions are mixed and contrasting. The decision in favour of a non-parametric technique has been driven mainly by the characteristics of the situation under analysis, namely a multi-output multi-input production process, and of the type of data considered: input and output data with zero entries. Although beyond the scope of the analysis conducted in the chapter, a comparison of the results from the parametric and non-parametric approaches could be the object of future research. An additional concern arises from the use of off-farm income as a covariate in the analysis of technical efficiency, although not directly but as a component of the income diversification indicator. Its omission from the analysis could have led to biased estimates since labour allocation decisions are likely to be correlated with pension receipt and household composition. Moreover the lack of detailed information on household time allocation has prevented the use of a less problematic indicator of income diversification. The problem was partially addressed using instrumental variable techniques whose validity is, however, vulnerable to the potential presence of serial correlation. This problem could be further investigated using alternative surveys with more detailed information on household time allocation or other community level information to be used as instruments for the income diversification indicator.

In the last chapter, the major concern arises from the use of the year of arrival as an instrument for land access and size. It is argued that the year of arrival is exogenous since, according to extensive narrative evidence, the timing of forced removals cannot be associated with household-specific unobserved characteristics, yet not all movements recorded by the surveys might be the result of forced removals. Because the reasons for moving are unknown, one of the main concerns is that forced removals could be confounded with voluntary migration challenging the exogeneity of the instrument. Voluntary relocations were more likely to occur within the homelands since conditions in these areas were extremely unfavourable and the available descriptive evidence suggests that no household would voluntarily move to the overcrowded and unpleasant homelands. In the chapter, the problem is partially addressed by limiting the analysis to those households that moved from non-homeland areas given the lower probability of encountering voluntary migration in this subsample. Nevertheless, the lack of information prevented me from fully addressing this problem and calls for a more detailed record of information on household migration history. A second concern arises from the assumption of a linear relationship between land size and household welfare, which fails to capture the potentially non-linear effects of land endowments. Non-linear analyses, however, often require non-parametric techniques or non-linear specifications where the presence of potential endogenous explanatory variables require the use of less conventional and more complex solutions, when possible. Nevertheless the relevance of such heterogenous effects leaves room for further investigation of the relationship between land and welfare across different dimensions of the farm household.

Finally, the three essays base their theoretical framework on a standard static farm household model where the household is assumed to act as a unitary decision unit and the process by which resources are distributed within the household is not taken into consideration. Inter-temporal modelling is needed to introduce uncertainty and to model households' behaviour toward risk. Although uncertainty and risk-aversion are likely to play an important role in shaping households' behaviours, these aspects have not been consider to avoid making the analysis of market imperfections overly complex. However, because the analysis of households' risk-management and risk copying strategies can add important insights into the analysis of rural households' livelihoods (Dercon, 2002), it leaves room for future research. As far as the unitary modelling of the household is concerned, the theoretical and empirical analyses proposed in this thesis do not sufficiently explore intra-household resource allocation and distribution. Although, for example, pension receipts might have different effects according to the gender of the recipient (as shown in Duflo (2003)), this distinction could not be made given the small size of the subsample of male pensioners. Collective household models have been increasingly adopted in the literature as an alternative to unitary models. These models explicitly consider the individuality of the household members and only assume that intra-household decisions are Pareto efficient (Chiappori, 1992). Empirical analyses conduced on the basis of the collective modelling of the household are, however, more demanding in terms of data disaggregation and often require specific surveys intended to capture intra-household task allocation and bargaining relationships. Having such data, it would be interesting to analyse whether, for example, household members are equally affected by liquidity constraints or whether the distribution of tasks within the household matters for household efficiency and welfare.

Bibliography

Ali, M. and Byerlee, D. (1991). Economic efficiency of small farmers in a changing world: a survey of recent evidence. *Journal of International Development*, 3(1):1–27. 58

Angrist, J. D. (2001). Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice. *Journal of Business and Economic Statistic*, 19(1):2–16. 66, 94, 97

Anriquez, G. and Daidone, S. (2008). Linkages between farm nad non-farm sectors at the household level in rural Ghana. ESA Working paper 08-01, FAO. 42

Ardington, C., Case, A., and Hosegood, V. (2009). Labor supply responses to large social transfer: longitudinal evidence from South Africa. *American Economic Journal: Applied Economics*, 1(1):22–48. 43, 62

Banker, R. D. (1993). Maximun likelihood, consistency and data envelopment analysis: A statistical foundation. *Management Science*, 39(10):1265–1273. 50

Banker, R. D. and Natarajan, R. (2008). Evaluating contextual variables affecting productivity using data envelopment analysis. *Operations research*, 56(1):48–58. 61

Bardhan, P. and Udry, C. (1999). *Development Microeconomics*. Oxford University Press, Oxford. 8

Barret, C. B. (1997). How credible are estimates of peasant allocative, scale, or scope efficiency? a commentary. *Journal of International Development*, 9(2):221–229. 40

Barrett, C. B., Reardon, T., and Webb, P. (2001). Nonfarm income diversification and household livelihood strategies in rural Africa: concepts, dynamics, and policy implications. *Food Policy*, 26:315–331. 8

Basu, K., Narayan, A., and Ravallion, M. (2002). Is literacy shared within households? *Labor economics*, 8:649–665. 46

Baumol, W. J., Panzar, C., and Willig, R. D. (1982). *Contestable markets and the theory of Industry structure*. Harcourt Brace Jovanovich, New York. 44

Bedi, A. and Tunali, I. (2005). Testing for market imperfections: participation in land and labor contracts in Turkish agriculture. Conference paper 385, 6th European Historical Economics Society Meeting, Istanbul, Turkey. 21

BEPA (2004). Prudent environmental management: A catalyst for economic development. Focus on Key Economic Issues 54, BEPA. 27, 30

Bertrand, M., Mullainathan, S., and Miller, D. (2003). Public policy and extended families: evidence from pensions in South Africa. *The World Economic review*, 17(1):27–50. 42, 43, 62

Besley, T. and Burgess, R. (2000). Land reform, poverty reduction, and growth: Evidence from India. *The Quarterly Journal of Economics*, 115(2):389–430. 77

Burgess, R. (2001). Land and welfare: Theory and evidence from China. Working paper, Department of Economics and STICERD, London School of Economics. 76, 77

Cafiero, C., De Janvry, A., Sadoulet, E., and Vakis, R. (2004). Testing for separability in household models with heterogeneous behavior: A mixture model approach. Working paper 990, CUDARE, University of California, Berkeley. 8

Carter, M. (1984). Identification of the inverse relationship between farm size and productivity: an empirical analysis of peasant agricultural production. *Oxford Economic Papers*, 86(1):131–145. 40

Carter, M. (1989). The impact of credit on peasant productivity and differentiation in Nicaragua. *Journal of Development Economics*, 31:13–36. 48, 67

Carter, M. R. and May, J. (1999). Poverty, livelihood and class in rural South Africa. *World Development*, 27(1):1–20. 2, 7, 73, 100

Case, A. and Deaton, A. (1998). Large cash transfers to the elderly in South Africa. *The Economic Journal*, 108(450):1330–1361. 28, 29, 62

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2:429–444. 50

Chavas, J. P., Petrie, R., and Roth, M. (2005). Farm household production efficiency: evidence from the Gambia. *American Journal of Agricultural Economics*, 87(1):160–179. 4, 41, 42, 43, 46

Chiappori, P. A. (1992). Collective labor supply and welfare. *Journal of Political Economy*, 100:437–467. 113

Coelli, T. J., Rao, P., O'Donnell, C. J., and Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. Springer, New York, second edition. 40, 51

Cox, K. R. (2004). Urbanization in South Africa and the changing character of migrant labour. *South African Geographical Journal*, 86(1):7–16. 100

de Janvry, A., Fafchamps, M., and Sadoulet, E. (1991). Peasant household behaviour with missing markets: Some paradoxes explained. *The Economic Journal*, 101(409):1400–1417. 21

de Wet, C. (1987). Betterment planning in South Africa: some thoughts on its history, feasibility and wider policy implications. *Journal of Contemporary African Studies*, 6(1/2):85–122. 83

de Wet, C. (1994). Resettlement and land reform in south africa. *Review of African and Political Economy*, 61:359–373. 83, 84

Deere, C. D. and de Janvry, A. (1981). Demographic and social differentiation among northern Peruvian peasants. *Journal of Peasant Studies*, 8(3):335–66. 58

Dercon, S. (2002). Income risk, coping strategies, and safety nets. *The World Bank Research Observer*, 17(2). 49, 113

Dercon, S. and Krishnan, P. (1996). Income portfolios in rural ethiopia and tanzania: Choices and constraints. *Journal of Development Studies*, 36(2):850–75. 1, 7, 9, 48

Desmond, C. (1971). *The discarded people: an account of African resettlement in South Africa*. Penguin, Harmondsworth. 81, 82, 85, 104

Duflo, E. (2003). Grandmothers and Granddaughters: Old-Age Pensions and Intrahousehold Allocation in South Africa. *World Bank Economic Review*, 17(1):1–25. 29, 62, 64, 113

Eastwood, R., Kirsten, J., and Lipton, M. (2006). Premature deagriculturalisation? land inequality and rural dependency in Limpopo province, South Africa. *The Journal of Development Studies*, 42(8):1325–1349. 7

Edmonds, E. V. (2006). Child labor and schooling responses to anticipated income in South Africa. *Journal of Development Economics*, 81(2):386–414. 29, 62, 64

Edmonds, E. V., Mammen, K., and Miller, D. L. (2005). Rearranging the family? income support and elderly living in a low income country. *Journal of Human resources*, 40(1):186–207. 62, 64

Ellis, F. (1993). *Peasant Economics. Farm Households and Agrarian Development*. Cambridge University Press, Cambridge and New York. 126

Eswaran, M. and Kotwal, A. (1986). Access to capital and agrarian production. *The Economic Journal*, 96(382):482–498. 8, 17, 74, 76

Fare, R., Grosskopf, S., and Zelenyuk, V. (2004). Aggregation bias and its bounds in measuring technical efficiency. *Applied economics letter*, 11(10):657–60. 58

Fenwick, L. J. and Lyne, M. C. (1999). The relative importance of liquidity and other constraints inhibiting the growth of small-scale farming in Kwazulu-Natal. *Development Southern Africa*, 16(1):141–155. 8, 15, 35

Fernandez-Cornejo, J. (2007). Off-farm income, technology adoption, and farm economic performance. Economic Research Report 30, USDA. 42

Filmer, D. and Pritchett, L. H. (2001). Estimating wealth effects without expenditure data - or tears: an application to educational enrollments in states of India. *Demography*, 38(1):115–132. 74, 87

Finan, F., Sadoulet, E., and de Janvry, A. (2005). Measuring the poverty reduction potential of land in rural mexico. *Journal of Development Economics*, 77:27–51. 76, 77, 102, 108

Fletschner, D. (2008). Women's access to credit: does it matter for household efficiency? *American Journal of Agricultural Economics*, 90(3):669–683. 42

Fletschner, D. and Zepeda, L. (2002). Efficiency of small landholders in Eastern Paraguay. *Journal of Agricultural and Resourse Economics*, 27(2):554–572. 42, 49

Freund, B. (1984). Resettlement and the political economy of south africa. *Review of African Political Economy*, 29:49–63. 83

Goodwin, B. K. and Mishra, A. K. (2004). Farming efficiency and the determinants of multiple job holding by farm operators. *American Journal of Agricultural Economics*, 86(3):722–29. 49

Hoff, A. (2007). Second stage dea: comparison of approaches for modelling the dea score. *European Journal of Operational Research*, 181(1):425–435. 65

Jensen, R. T. (2004). Do private transfers "displace" the benefit of public transfers? evidence from south africa. *Journal of Public Economics*, 88(1-2):89–112. 43, 49

Kelejian, H. H. (1971). Two-stage least squares and econometric systems linear in parameters but nonlinear in the endogenous variables. *Journal of American Statistical Association*, 66(334):273–374. 66

Keswell, M., Carter, M., and Deininger, K. (2010). Poverty and land redistribution: Quasi-experimental evidence from south africa's lrad program. Mimeo, Stellenbosch University. 77

Klasen, S. (2008). Surviving unemployment without state support: unemployment and household formation in South Africa. *Journal of African Economies*, 18(1):1–51. 43

Kumbhakar, S. C. and Lovell, C. A. K. (2000). *Stochastic Frontier Analysis*. Cambridge University Press, Cambridge. 51

Lau, L. J. (1972). Profit functions of technologies with multiple inputs and outputs. *The Review of Economics and Statistics*, 54(3):281–289. 44, 45

Leathers, H. D. (1991). Allocable fixed inputs as a cause of joint production: a cost function approach. *American Journal of Agricultural Economics*, 73(4):1083–1090. 44

Leibbrandt, M., Poswell, L., Naidoo, P., and Welch, M. (2006). Measuring recent changes in south african inequality and poverty using 1996 and 2001 census data. In Bhorat,

H. and Kanbur, R., editors, *Poverty and Policy in Post-Apartheid South Africa*. HSRC Press, Pretoria. 1

Lipton, M. and Lipton, M. (1993). Creating rural livelihoods: some lessons for South Africa from experience elsewhere. *World Development*, 21(9):1515–1548. 2, 73

Long, J. S. and Freese, J. (2003). *Regression Models for Categorical Dependent Variables Using Stata (rev.)*. Stata Press, College Station, TX. 24

Lyne, M. and Darroch, M. (2004). Land redistribution in Kwazulu-Natal, South Africa: Five census surveys of farmland transactions, 1997-2001. Basis crsp, Department of Agricultural and Applied Economics. 83

Maitra, P. and Ray, R. (2003). Resource inflows and household composition: Evidence from south african panel data. *Journal of International Development*, 15(8):1037–1047. 62

Makhura, M. T. (2001). *Overcoming transactions costs barriers to market participation of smallholders farmers in the Northern Province of South Africa*. PhD thesis, Department of Agricultural economics, University of Pretoria. 34

McDonald, J. (2009). Using least squares and tobit in second dea efficiency analyses. *European Journal of Operational Research*, 197(2):792–798. 61, 65

McKenzie, D. J. (2005). Measuring inequalit with asset indicators. *Journal of Population Economics*, 18(2):229–260. 87

Moschini, G. (1989). Normal inputs and joint production with allocatable fixed factors. *American Journal of Agricultural Economics*, 71(4):1021–1024. 44

Murray, C. (1987). Displaced urbanization: South africa's rural slums. *African Affairs*, 86:311–29. 82, 94

Pauw, K. (2005). Forming representative household and factor groups for a south african sam. Technical paper series, PROVIDE Project. 100

Piesse, J., Von Bach, H. S., Thirtle, C., and Van Zyl, J. (1996). The efficiency of smallholder agriculture in South Africa. *Journal of International development*, 8(1):125–144. 36, 41, 58

Platzky, L. and Walker, C. (1985). *The Surplus People: Forced Removals in South Africa*. Rivan Press, Johannesburg. 81, 82, 85, 101, 106

Posel, D., Fairburn, J. A., and Lund, F. (2006). Labour migration and households: A reconsideration of the effects of the social pension on labour supply in south africa. *Economic Modeling*, 23:836–853. 43, 62

Pudney, S. and Shield, M. (2000). Gender, race, pay and promotion in the British nursing profession: estimation of a generalized ordered probit model. *Journal of Applied Econometrics*, 15(4):367–399. 23

Ranchhod, V. (2006). The effect of the south african old pension on labour supply of the elderly. *South African Journal of Economics*, 74(4):725–744. 42

Ravallion, M. (2003). Exernalities in rural development. evidence for china. Discussion Paper 2003/54, WIDER. 46

Roemer, J. (1982). *A General Theory of Exploitation and Class*. Harvard University Press, Cambridge. 8

Sadoulet, E., de Janvry, A., and Benjamin, C. (1998). Household behavior with imperfect labor markets. *Industrial Relations*, 37(1):85–108. 3, 8, 13, 15, 16, 21, 35, 110

Shumway, C. R., Pope, R. D., and Nash, E. K. (1984). Allocatable fixed inputs and jointness in agricultural production: implications for economic modeling. *American Journal of Agricultural Economics*, 66(1):72–78. 44, 45

Simar, L. and Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*, 44(1):49–61. 50, 60

Simar, L. and Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136:31–64. 61

Simkins, C. (1983). *Four essays on the past, present and possible future of the distribution of the black population in South Africa*. Saldru, Cape Town. 82, 84

Singh, I., Squire, L., and Strauss, J. (1986). *Agricultural Household Models*. Johns Hopkins University Press, Baltimore. 7, 76, 124

Valente, C. (2009). The food (in)security impact of land redistribution in South Africa: microeconometic evidence from national data. *World Development*, 37(9):1540–1553. 78

Van Zyl, J., Binswanger, H. P., and Thirtle, C. (1995). The relationship between farm size and efficiency in South African agriculture. Policy research Working paper 1548, World Bank. 13, 35, 41

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *The Stata Journal*, 6(1):58–82. 23

Wilson, P. W. (1993). Detecting outliers in deterministic nonparametric frontier models with multiple outputs. *Journal of Business and Economic Statistic*, 11(3):319–323. 52, 56

Wilson, P. W. (2007). Fear: A software package for frontier efficiency analysis with r. *Socio-Economic Planning Sciences*, 42(4):247–254. 58

Appendix 1

The on-farm marginal productivity of labour, w^*, in the case of labour-autarkic households is used as latent decision variable in the empirical section of chapter 1. To analyse how household characteristics and exogenous factors affect household membership in the three categories, it is useful to analyse the response of the on-farm marginal productivity of labour to changes in its exogenous determinants. Considering the self-cultivators category presented in the chapter, the following six equations are sufficient to determine $f_i^1, w^*, x, y, \widetilde{p_x}$ and λ:

$$
\begin{aligned}
&u_1 = u_y p q_L,\\
&p q_L = w^*,\\
&p q_x = p_x(1 + \lambda\alpha/u_y),\\
&y = pq(A, f_q^1, x) - p_x x + w_o^2 k f^2 + T - K,\\
&\alpha w_o^2 k f^2 + \varphi T - \alpha p_x x = 0,\\
&u_y p q_x - p_x u_y - \alpha p_x \lambda = 0.
\end{aligned}
$$

Following Singh et al. (1986) we totally differentiate the first order conditions arranging the results into a matrix framework:

$$\begin{pmatrix} a & u_{1y}(pq_x - p_x) - u_{yy}pq_l(pq_x - p_x) - u_y pq_{Lx} & 0 & 0 & 0 & 0 \\ -pq_{LL} & -pq_{Lx} & 1 & 0 & 0 & 0 \\ -pq_{Lx} & -pq_{xx} & 0 & 0 & 0 & 0 \\ -pq_L & p_x - pq_x & 0 & 1 & 0 & 0 \\ 0 & -\alpha p_x & 0 & 0 & 0 & 0 \\ b & u_{yy}pq_x pq_x + u_y pq_{xx} - p_x u_{yy} pq_x & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} df_q^1 \\ dx \\ dw \\ dy \\ d\lambda \\ d\widetilde{p_x} \end{pmatrix} =$$

$$\begin{pmatrix} (-u_{11} + u_{y1}pq_L)df^1 & cdf^2 & (-u_{1y}pq_A + u_y pq_{LA} + u_{yy}pq_A)dA & (-u_{1y} + u_{yy}pq_L)dT & (u_{1y} - u_{yy}pq_{Lx})dp_x \\ 0 & 0 & pq_{LA}dA & 0 & 0 \\ 0 & 0 & pq_{xA}dA & 0 & 0 \\ 0 & w_o^2 kdf^2 & pq_A dA & dT & xdp_x \\ 0 & -\alpha w_o^2 kdf^2 & 0 & -\varphi dT & -x\alpha dp_x \\ (u_{y1}pq_x - p_x u_{y1})df^1 & ddf^2 & (-u_{yy}pq_x pq_A - u_y pq_{xA} + p_x u_{yy} pq_A)dA & (-u_{yy}pq_x + p_x u_{yy})dT & (u_y + \lambda\alpha)dp_x \end{pmatrix}.$$

Where $a = -u_{11} + 2u_{1y}pq_L - u_{yy}pq_L pq_L - u_y pq_{LL}$, $b = u_{yy}pq_x pq_L - u_{y1}pq_x + u_y pq_{xL} - p_x u_{yy} pq_L + p_x u_{y1}$, $c = u_{yy}pq_L w_o^2 k + u_{y2}pq_L(1-k) - u_{y1}w_o^2 k$ and $d = -u_{y2}pq_x(1-k) - u_{yy}pq_x w_o^2 k + p_x u_{yy} w_o^2 k + p_x u_{y2}(1-k)$.

Recalling Cramer's rule, the marginal effect of the k-th exogenous variables, z_k, on the on-farm marginal productivity of labour can be determined as follow:

$$\frac{\partial w^*}{\partial z_k} = \frac{1}{D} D_{wk}.$$

where D is the determinant of the first matrix (left hand side) reported above and it is given by:

$$D = p_x(-u_{11} + 2u_{1y}pq_L - u_{yy}pq_L pq_L - u_y pq_{LL}) > 0,$$

where D_{wk} is the determinant of the matrix obtained substituting the k-th column of the last matrix into the third column of the first matrix. It is now possible to derive the sign of the derivatives with respects to the exogenous variables of the model:

$$\begin{aligned}
\frac{\partial w^*}{\partial f^1} &= \frac{1}{D}[pq_{LL}p_x\alpha(u_{y1}pq_L - u_{11})] < 0, \\
\frac{\partial w^*}{\partial f^2} &= \frac{1}{D}[\alpha w_o^2 k(2u_{1y}pq_L pq_{Lx} - u_{11}pq_{Lx} - u_{1y}pq_x pq_{LL} + u_{yy}pq_L pq_{LL}pq_x) \\
&\quad +pq_L pq_{LL}p_x\alpha u_{y2}(1-k)] \lesseqgtr 0, \\
\frac{\partial w^*}{\partial T} &= \frac{1}{D}[(pq_x + p_x\alpha(1-\varphi))(u_{yy}pq_L pq_{LL} - u_{1y}pq_{LL}) - pq_{Lx}pq_L\varphi(u_{yy}pq_L - 2u_{y1}) \\
&\quad -u_{11}pq_{Lx}\varphi] > 0, \\
\frac{\partial w^*}{\partial p_x} &= \frac{1}{D}[(pq_L pq_{Lx} - pq_{LL}pq_x)(u_{yy}pq_{Lx} - u_{y1}\alpha x) - 2u_{y1}pq_L pq_{Lx}\alpha x \\
&\quad +u_{11}pq_{Lx}\alpha x] < 0, \\
\frac{\partial w^*}{\partial A} &= \frac{1}{D}p_x\alpha[pq_A pq_{LL}(u_{yy} - u_{y1}) - u_{11}pq_A + pq_L pq_{LA}(2u_{u1} - u_{yy}pq_L)] > 0.
\end{aligned}$$

The ambiguity of the skilled labour endowment effect is now discussed. Considering the extreme instance in which all skilled members are unemployed ($k = 0$), the above mentioned effect becomes unambiguously negative. In this case the self-cultivators model falls into the classical Chayanov farm household model (Ellis, 1993) where the ratio between consumers and working members ($(f^1 + f^2)/f^1$) is negative correlated with w^*, such that a larger number of dependant skilled members lowers the marginal productivity of labour. However, when $k > 0$, the higher the amount of time skilled members devote to off-farm labour the lower is their dependance on the rest of the household and, at the same time, the higher are the positive income and liquidity effects induced by the extra off-farm earnings. These two opposite effects (the consumers-workers ratio effect and the income/liquidity effect) explain the ambiguity of the overall sign.

Appendix 2

Table 9: First stage regressions

	Pensioner (dummy)		Share off farm income
Eligible man (dummy)	0.324***	Share off farm income in 1998	0.288***
	(0.070)		(0.064)
Eligible woman (dummy)	0.644***	Household eligibility	-1.051
	(0.057)		(6.881)
Share off farm income	-0.0006		
	(0.000)		
Gender of household head	-0.047	Gender of household head	7.405
	(0.029)		(6.841)
Age of household head	0.015***	Age of household head	-0.034
	(0.002)		(0.295)
Education of household head	-0.004	Education of household head	-0.582
	(0.003)		(0.671)
Ratio of skilled members	0.051	Ratio of skilled members	11.220
	(0.058)		(12.290)
Title on land	-0.018	Title on land	-4.769
	(0.023)		(4.789)
Employment rate (district)	0.00131	Employment rate (district)	0.126
	(0.001)		(0.207)
Kleibergen-Paap F stat	88.574	Kleibergen-Paap F stat	14.882
Observations	505	N	193

Robust standard errors in parenthesis. Regressions include all the covariates used in the second stage analysis previously reported.

Appendix 3

Table 10: First stage regressions

	(1) Land (dummy)	(2) Land (hectares)
Year of arrival	-0.009***	-0.012***
	(0.002)	(0.003)
Education of household head	-0.003	0.003
	(0.007)	(0.003)
Age of household head	-0.002	0.030
	(0.010)	(0.036)
Gender of household head (dummy)	0.014	0.556***
	(0.045)	(0.155)
Pension eligible members (dummy)	-0.120	0.054
	(0.081)	(0.190)
Children age 1-5	0.060*	-0.085
	(0.035)	(0.064)
Children age 6-17	0.048***	-0.039
	(0.015)	(0.036)
Number of skilled members	0.063***	0.019
	(0.021)	(0.064)
Number of unskilled members	0.064***	-0.051
	(0.017)	(0.046)
Highest level of education	0.003	0.051
	(0.003)	(0.048)
Labor market dev index (district level)	-0.257**	-1.440***
	(0.111)	(0.004)
Population density (district level)	-0.000***	-0.002***
	(0.000)	(0.001)
Observations	695	2736

Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
All regressions include also the square of the age of the head of the household.