



A University of Sussex DPhil thesis

Available online via Sussex Research Online:

<http://sro.sussex.ac.uk/>

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details

Social Networks and Economic Life in Rural Zambia

Jennifer Leavy

Doctor of Philosophy (Economics)

University of Sussex

30th September 2011

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:.....

UNIVERSITY OF SUSSEX

Jennifer Julie-Anne Leavy

Doctor of Philosophy

Social Networks and Economic Life in Rural Zambia

SUMMARY

This thesis explores the relationship between social networks and economic life in rural Zambia. The motivation for the study lies in the crucial role played by social context and social networks in exchange behaviour in rural sub-Saharan Africa, and inherent difficulties in formalising market transactions in this context within a standard neoclassical economics framework.

The study examines the role of social networks in rural production systems, focusing on crop market participation. It is based on analysis of findings from social network research conducted by the author in three predominantly Bemba villages in Northern Province, Zambia. Data collected using quantitative and qualitative methods are used to map social networks of individuals and households. Variables are constructed capturing network characteristics, and incorporated into transactions cost models of commercialisation.

The overarching question is: do social networks play a role in determining farming success in settings with little variability between households on assets and endowments – land, labour, inputs – and where markets are incomplete or missing? Do social networks mediate market and resource access, helping to explain socio-economic differences between households?

The research finds rural life is characterised by diverse networks with multiple, overlapping functions. Much economic exchange takes place on reciprocal or kinship bases, rooted in social norms and reflecting community structures. How social networks are measured matters. Different network attributes are important for different people, and relationships between networks and outcomes depend on the measure used. Controlling for endogeneity, estimation results suggest larger networks have a negative effect on crop incomes whereas having a greater proportion of kin in the network has a positive effect, implying that in this context strong ties are key. Qualitative research suggests the nature of people's networks and their positions within them play an important role in the command over labour: "the famous always get their work done".

Table of Contents

Table of Contents	iv
Acknowledgements	vi
List of Tables	vii
List of Figures.....	viii
Chapter One The Context	1
1.1 Background	3
1.2 Commercialisation and Household Crop Market Participation	5
1.3 Social Institutions, Economic Decisions and Livelihood Strategies in Rural sub-Saharan Africa	8
1.4 The Study Area	12
1.4.1 Economic Development and Agricultural Commercialisation in Zambia	13
1.4.2 Northern Province	22
Chapter Two Modelling the Social Context: Social Capital and Social Networks	27
2.1 Defining and Conceptualising Socially Embedded Resources: Social Capital.....	28
2.2 Criticisms of the Social Capital Literature.....	30
2.3 Social Networks and Socially Embedded Resources.....	32
2.4 Operationalising social capital and social networks	34
2.4.1 Econometric studies of Social Networks in Economic Life	36
2.4.2 Criticisms of the social networks approach in economics	43
2.5 Social Network Analysis.....	45
2.5.1 Definition	45
2.5.2 Methods.....	45
2.5.3 Concepts and Measures.....	46
2.6 Conceptual Frameworks: Theoretical Approaches to Markets and Networks.....	47
2.7 Modelling Markets and Networks.....	51
2.7.1 Smallholder crop market participation and transactions costs	52
2.7.2 Social Networks and Commercialisation	60
Chapter Three Data Collection Methods.....	64
3.1 Study Site Selection	64
3.2 Ethical Considerations	66
3.3 The Need for a Mixed-Method Approach.....	68
3.3.1 Quantitative and Qualitative Traditions	68
3.3.2 Tensions between quantitative and qualitative approaches.....	70
3.3.3 Combined methods – transcending the tensions	71
3.3.4 Integrating quantitative and qualitative approaches.....	73
3.4 Design of Data Collection Instruments	77
3.4.1 Household survey instrument design	81
3.4.2 Qualitative instrument design	86
3.5 Data Collection	88
3.5.1 The Household Survey	89
3.5.2 Generating qualitative data	91
3.6 Data Entry and Data Cleaning	97
3.6.1 Survey Data Entry	97
3.6.2 Qualitative Data Entry	98
3.6.3 Social Networks Data Entry and Dataset Construction	98
3.6.4 Data Cleaning.....	99
3.7. Methodological Concerns	99
Chapter Four Social Networks in Three Zambian Villages.....	106
4.1 Egocentric Social Networks.....	106
4.2 Social Networks Measures.....	110
4.2.1 Network Size.....	111
4.2.2 Prestige and Influence.....	112

4.2.3 Diversity in Composition and Function	113
4.3 Correlations between different social networks measures	121
4.4 Network characteristics and respondent attributes.....	122
4.5 Determinants of Social Network Characteristics	124
4.6 Conclusion	129
Chapter Five Social Networks and Access to Economic Opportunities.....	131
5.1 Linking Social Networks to Outcomes	131
5.1.1 Social Networks Hypotheses	132
5.1.2 Social Networks Variables.....	134
5.1.3 Outcome Variables.....	134
5.2 Hypothesis 1: Bigger is Better	138
5.2.1 Social Network Measures: Network Composition.....	138
5.2.2 Network Size and Outcomes	138
5.3 Hypothesis 2: Friends in High Places	141
5.3.1 Social Network Measures: Prestige and Influence.....	141
5.3.2 Influence and Outcomes.....	142
5.4 Hypothesis 3: The Strength of Heterogeneous Networks	143
5.4.1 Social Network Measures: Diversity in Composition and Function.....	144
5.4.2 Network Diversity and Outcomes	145
5.5 Smallholder Commercialisation and Command over Resources.....	149
5.5.1. A Typology of Smallholder Farming Households	149
5.5.2 Cluster Determinants.....	154
5.5.3 Multinomial Probit.....	157
5.6 Conclusion	162
Chapter Six Empirical Model of Household Crop Marketing and Social Networks.....	165
6.1 Empirical Model	166
6.2 Challenges in Estimating the Empirical Model	168
6.2.1 Endogeneity	168
6.2.2 Separability	171
6.2.3 Identification	173
6.3 Methodology	174
6.3.1 Complex survey design:.....	174
6.4.2 Choice of models	174
6.3.3 Choice of variables.....	175
6.3.4 Regressor Endogeneity	182
6.4 Instrumental Variable Analysis.....	185
6.4.1 Instrument Validity	185
6.4.2 Diagnostic Tests.....	186
6.4.3 Potential instruments.....	188
6.5 Results and Analysis	193
6.5.1 GMM estimation.....	193
6.6 Conclusion	201
Chapter Seven Conclusions.....	204
7.1 Summary of the study	204
7.2 Challenges in undertaking the research	205
7.3 Summary of the findings.....	208
7.4 Broader implications	211
Bibliography.....	214
Appendices.....	240

Acknowledgements

This thesis owes everything to the cooperation of many people. The first and biggest thanks must go to the people of Ngulula, Kabila and Lufubu Resettlement Scheme, for their hospitality, generosity, and inspiration.

The research was designed alongside a DFID-funded research project on rural labour markets in sub-Saharan Africa directed by Dr Howard White, in collaboration with Professor Venkatesh Seshamani at the University of Zambia, and supported by Ngosa Chisupa at the Zambian Ministry of Labour and RuralNet Associates, Lusaka. While the research takes a different direction to the project, it would not have been possible without it.

In Northern Province, Mr William Shimumbi, Mr Freddie Mpundu and staff at the Misamfu Regional Research Centre, Kasama, eased our access to the communities, providing invaluable logistical support and insights. In Lusaka, many thanks to the RuralNet survey team, many of whom I now feel privileged to count among my very dear friends: Mwila Mulumbi, Kelvin Munjile, Jessie Mukwaila, Fison Mujenje, Stuart Nsana, Geoffrey Siulemba, Muyaka Kampango, Mupeta Sokontwe, Dennis Chiwele, Stephen Tembo, Davison MendaMenda, Ngosa Chisupa and Rhoda Chiwele and her team of data-entry people. The nature of this research means I could not have even scratched the surface of Zambian smallholder social networks without the energy and qualitative research skills of Mwila Mulumbi and Kelvin Munjile. It was a pleasure to spend extra time in the field with Mwila and Kelvin and learn from the wonderful way they have with people. Sam Gibson and Taylor Brown always provided a warm welcome and solid base in Lusaka on weary returns from the field. Arriving at their house always felt like coming home.

My supervisors Julie Litchfield and Howard White have been brilliant. This thesis has been far too long in the making, and I would not have got there in the end without their patience and support. I benefitted greatly from their knowledge, experience and good humour. I am very grateful to Howard for generously providing me the space to indulge my research interests under the aegis of the rural labour project, and for continuing to engage with the research long after he was obliged to. I am hugely grateful for Julie's excellent support and advice, especially during The Big Push, and for helping me to feel as though it's all been worth it in the end.

At the University of Sussex, Buzz Harrison provided excellent comments from an anthropology perspective on early drafts. Ann Whitehead was an inspiration; her outstanding qualitative methods course really spoke to me. Andy Newell in the Economics Department provided excellent feedback. Advice in the very early stages of the research came from Marlene Dekker, Joachim DeWeerd and Abigail Barr, some of which I followed, some I didn't but wish I had.

I have had tremendous support from my employer, The Institute of Development Studies, most notably Allister McGregor. Invaluable academic advice and moral support came from my mentor Naila Kabeer and colleagues in the Vulnerability and Poverty Reduction Team, past and present. I must also single out Milasoa Cherel-Robson, Naomi Hossain, Joanna Wheeler, Kay Sharp, Ed Anderson and Isabel Vogel, who have all been there and survived to tell uplifting stories over cups of tea.

Thanks to the many friends too numerous to mention who never tired of asking me how it was going, or if they did were kind enough not to show it. And for helping to keep my feet on the ground, Family, both sides, by asking "are you still doing that?" and "who's going to read it?"

Finally, Matt, Seth, Eliza: gold medals all round for tolerance. So much to thank Matt for: help with data entry; introducing me to philosophy and the delights of phenomenology; boot-camp style timetabling in the final stages; delicious dinners and for retraining as a journalist in good time for the write-up. You say Occam's Razor, I say parsimony.

List of Tables

Table 1.1 Nutritional Status of Children under five years of age by farmer category	6
Table 1.2 Typology of Agricultural producers in Zambia	16
Table 1.3 Growth Rates (per cent per annum) for Selected Crops Produced by Smallholder Farmers, Zambia , 1991 – 2004	19
Table 1.4 Selected Indicators 2002	22
Table 2.1 Transaction Cost Measures in Studies of Smallholder Commercialisation	56
Table 3.1 Survey sites	65
Table 3.2 Main features of qualitative and quantitative approaches	69
Table 3.3 Strengths and weaknesses of quantitative and qualitative methods	74
Table 3.4 Household Survey Modules	82
Table 3.5 Qualitative methods used	87
Table 3.6 Sampling frame	90
Table 3.7 Sample Attrition	91
Table 3.8 Qualitative Fieldwork – Total Number of Interviews	94
Table 3.9 People selected for Oral Life Histories – Round 2 of Fieldwork	95
Table 4.1 Social Networks/ Capital Measures Used in a Selection of Econometric Studies	107
Table 4.2 Network Size Variables	111
Table 4.3 Prestige Network Variables	113
Table 4.4 Composition Diversity Variables	115
Table 4.5 Kin as proportion of Network, by Community	118
Table 4.6 Functional Diversity Variables	120
Table 4.7 Social Networks Variables: Pearson Correlation Coefficients	122
Table 4.8 Dependent Variables in Social Networks OLS estimations	124
Table 4.9 OLS Results: Determinants of Social Networks Characteristics	126
Table 5.1 Outcome Variables: Summary Statistics N=136 Households	137
Table 5.2 Network Size Variables – Household Level	138
Table 5.3 Mean Household Network Composition and Outcomes	139
Table 5.4 Distribution of Network Size and Group Affiliation by Socio-Economic Status (% households)	139
Table 5.5 Distribution of Network size and Group Affiliation by Crop Market Participation (% households)	140
Table 5.6 Distribution of Network Size and Group Affiliation by Crop Sales (% households)	141
Table 5.7 Prestige Network Variables – Household Level	142
Table 5.8 Percentage of Outcome Group with Influential Alters	142
Table 5.9 Distribution of prestige networks by crop market participation (% households)	143
Table 5.10 Distribution of prestige networks by level of crop sales (% households)	143
Table 5.11 Composition Diversity Variables – Household Level	145
Table 5.12 Functional Diversity Variables	145
Table 5.13 Mean Network Diversity and Outcomes	146
Table 5.14 Distribution of age heterogeneity of alters by crop market participation and crop sales (% households)	147
Table 5.15 Value of crops sold, by community ‘000 Kwacha	150
Table 5.16 Farmer Clusters - Typology	152
Table 5.17 Farmer Clusters (number of households)	153
Table 5.18 Correlation Coefficients: Social Networks and Transactions Costs	156
Table 5.19 Descriptive Statistics: Dependent Variables and Regressors Multinomial Probit ..	159
Table 5.20 Multinomial Probit Results: Determinants of Farmer Cluster	160
Table 6.1 Sampling weights	174
Table 6.2 Descriptive Statistics: Dependent Variables and Potential Regressors	181
Table 6.3 Correlation Matrix dependent variable, endogenous regressors, instrumental variables	190
Table 6.4 Instrument Sets for Potentially Endogenous Regressors	192

Table 6.5 Summary results of diagnostic tests for instrument relevance and validity: GMM with social networks variables	196
Table 6.6 Estimation Results: GMM regressions with instrumental variables	198

List of Figures

Figure 1.1: GDP growth (%) – 1961-2010	14
Figure 1.2: GDP per capita – 1965-2010	15
Figure 1.3: Map of Zambia: Provinces	22
Figure 4.1: Egocentric Networks of Three Interconnected Egos	109
Figure 4.2: Egocentric Network with Two Multiplex Ties.....	110
Figure 4.3: Kernel Density Estimate – Network Size	112
Figure 4.4: Kernel Density Estimate – Age Heterogeneity of Network	116
Figure 4.5: Kernel Density Estimate – Sex Heterogeneity of Network	116
Figure 4.6: Kernel Density Estimate – Kin Composition of Network.....	117
Figure 4.7: Kernel Density Estimate – Network Multiplexity (proportion of ties).....	121
Figure 5.1: Kernel density estimate for crops sold, by community	151
Figure 6.1: Kernel Density Estimate – Value of crops sold.....	177
Figure 6.2: Kernel Density Estimate – Value of crops sold (log).....	177

Chapter One

The Context

This thesis explores the relationship between social networks and economic life in rural Zambia. The social-embeddedness of economic decision-making in rural sub-Saharan Africa is well documented. A person's social network can mediate access to markets and resources in environments characterised by an absence of functioning credit and insurance markets, cash constraints and a lack of assets. The motivation for the study lies in the crucial role played by social context and social networks in exchange behaviour and the inherent difficulties in formalising market transactions in this context within the standard neoclassical framework.

The empirical study examines the role of social networks in rural production systems in three predominantly Bemba villages in Northern Province, Zambia, focusing on agricultural commercialisation and market participation. Social network analysis techniques are used to construct variables capturing characteristics of people's social networks, from data collected during household surveys in three study sites in Northern Province, Zambia. These are incorporated into models of commercialisation based on a transactions cost approach (Bellemare and Barrett, 2006; Key et al, 2000; Goetz, 1992).

The main objective of the research is to uncover the ways in which smallholder farmers' social networks are important in enabling them to participate in crop output markets, in the context of remote rural areas in sub-Saharan Africa with restricted access to markets and institutions. Do networks facilitate smallholder farming households' engagement with crop output markets? Is there value in terms of economic outcomes of being able to access resources through personal social networks? Does this lead to better economic outcomes for households in the form of higher crop sales, through reducing transactions costs and mediating access to markets?

Social networks are analysed in relation to respondent attributes to explore whether or not the way we measure social networks matters. For example, do particular kinds of people have large networks, or networks containing many kin members, and what does this potentially mean in relation to what different kinds of networks may allow people

to access or achieve. For some, a large proportion of kin in the network may bring dividends for example in access to labour or other resources, for others this may constitute a higher level of obligation and more demands on resources. The social networks variables are included in a model of household agricultural commercialisation with transaction costs. The research is informed by in-depth qualitative material gathered during extended stays in the study sites. While many studies have been conducted of social networks in the context of technology, information diffusion, and in migration, social networks are still relatively under-researched in the area of agriculture and commercialisation¹. The thesis potentially contributes to the methodological literature: how we conduct research on social networks and how important it is to think about the nature of social networks when attempting to examine the effect of “connectedness” on economic decisions. The research also offers some interesting empirical insights into the extent to which connectedness matters.

The research finds the way social networks are measured matters: different kinds of people have different kinds of networks, and relationships between social networks and outcomes depend on the measure used. This has implications for the kinds of resources people can access through networks. While there are potential benefits of networks in terms of outcomes there may also be drawbacks or costs. Social networks can be an important conduit for accessing markets and resources, as well as social support and friendship. However, obligations to social network partners can act as a drain on a household’s resources or lead to non-optimal economic behaviour. Although the sample in the research is by no means nationally representative, the case studies presented here yield insights applicable to other smallholder farmers living in similar conditions and in similar ways. These contributions have implications for a number of research areas, for example in microfinance where social networks underpin lending models, in social capital and in migration and remittance research.

This chapter locates the thesis research in the context of agricultural development and commercialisation in Northern Province, Zambia, and the way social relations have played a role historically, and still feature prominently, in production systems, markets and decision-making of smallholder farmers. The chapter sets out first the choice of

¹ This will be examined in the literature review in Chapter 2, and returned to in Chapter 4.

location for the study (Section 1.1) before turning to agricultural commercialisation and the importance of markets for smallholder farming households in rural sub-Saharan Africa (Section 1.2). Section 1.3 describes the role of social institutions in people's economic decisions and livelihood strategies in rural sub-Saharan Africa. The economic, geographical and policy context for the study focusing on Zambian smallholder agriculture and that of Northern Province in particular are set out in Section 1.4.

1.1 Background

The background to the research is a project investigating rural labour markets and poverty in sub-Saharan Africa at the Institute of Development Studies, UK. The project was funded by the UK's Department for International Development and conducted in partnership with the University of Zambia, the Zambian Ministry of Labour and RuralNet Associates, a Lusaka-based rural development consultancy, who provided important logistical and fieldwork support. Motivation for the project stemmed from an apparent dichotomy between the importance of rural populations in poverty figures and the availability of data reflecting a true understanding of the workings of rural labour markets. This is a surprising gap given that a major way for poor rural people, who are not directly engaged in markets, to benefit from pro-poor growth is through working for those who *are* active in markets.² While the DPhil research was linked to the project, the study here, including methods, analysis and conclusions, is all my own.

The choice of country in which to conduct the DPhil research was determined by coverage of the rural labour markets project. Zambia was chosen as the country of interest for a number of reasons. First, there were sufficient links with the University of Zambia to allow ease of collaboration with in-country partners. Second, Zambia has a significant smallholder agriculture sector, with labour the most important factor of production on smallholder farms, but differs from other countries in the region in that the former dominance of copper mining encouraged much internal migration suggesting

² See: White H, J Leavy, M Mulumbi, G Mulenga, Venkatesh Seshamani (2005) *Rural Labour Markets in Africa: a channel for pro-poor growth? Preliminary findings from a research project in Northern Province, Zambia*. Discussion Draft. IDS Mimeo; White, H, J Leavy, J and

that remittances may have increased monetisation of the rural economy. Northern Province in particular had traditionally been thought of as a labour reserve for the Copperbelt mines, with much circular migration between the two regions. Third, there was a clear information gap – the most recent *Zambian Living Conditions Monitoring Survey* at the time of the research (LCMS III, 2002), contained no detailed information about rural labour exchange patterns and behaviour; neither are there any explicit social networks data for Zambia³. Fourth, a long anthropological tradition in the Central-Southern African region provided a rich secondary source of documentation about the lives of rural people, from the work of earlier anthropologists such as Audrey Richards' famous study "Land, Labour and Diet in Northern Rhodesia" (1939), to monographs of the 'Manchester School' of anthropologists and the Rhodes-Livingstone Institute⁴, encompassing the work of J. Clyde Mitchell on social networks, A. L. Epstein (1969; 1981) on organisations, kinship and social relationships, Norman Long (1968) on social change, and more recent studies, for example Moore and Vaughan's 'restudy' of Richards' seminal work (1994).⁵

A single province, Northern Province, was chosen as a further focus, first, due to its location. It was felt to be important to go further afield than areas close to the capital that tend to be easier to access and therefore more researched (so-called 'pro-road bias'). Second, again there are many classic and modern anthropological studies of this region and its people, particularly the Bemba, the dominant tribe, which provide a good back-drop to rural economic life (see Richards, 1939; Moore and Vaughan, 1994; among others). This wealth of secondary data helped to inform the conceptual framework of the thesis research and to design the data collection instruments, as well as in contextualising and deepening understanding in subsequent analyses.

V Seshamani (2005); 'Agricultural Development in Zambia's Northern Province: Perspectives from the Field Level', *IDS Bulletin* 36.2:132-138, Brighton: IDS.

³ There have been three subsequent rounds of the LCMS, in 2004, 2006 and 2010.

⁴ Renamed 'The Institute for African Studies of the University of Zambia' following Zambian independence.

⁵ See Magubane (1971), Mafeje (1976), among others, for criticisms of early anthropological work in the region. See also Poewe (1978, 1979).

1.2 Commercialisation and Household Crop Market Participation

The agricultural commercialisation literature offers a number of definitions of ‘commercialised’ in relation to smallholder farming households. At the centre of many of the definitions of agricultural commercialisation is the degree of participation in output markets, focusing on cash incomes. These range from simple measures of volumes or values of crops sold to indices of crop sales, for example measured as a proportion of all output (see for example Strasberg et al, 1999). Commercialisation can also mean commercialisation in input use, captured as the degree to which the household uses purchased inputs or non-household labour, or at its broadest, the degree to which the household is integrated into the cash economy, thus going beyond purely agricultural activities to consider all market activity (for more discussion see: Leavy and Poulton, 2007; Pingali, 1997; von Braun, 1995; von Braun and Kennedy, 1994). In the context of the research, agricultural commercialisation means that household crop production does not just meet family subsistence needs, but the household is also (or solely) producing for local, national or international markets. Agricultural commercialisation can be seen to be a crucial part of the broader development process, forming the basis for agricultural growth. Marketing problems have been identified as the most binding institutional constraints to increasing agricultural productivity, whereas connecting small, isolated communities into larger markets - at local, national and global levels - stimulates demand for farmers’ output. This, in turn, should generate demand for inputs (World Bank, 2008; World Bank, 2002: 32).

Markets and the private sector are important for pro-poor livelihood development and poverty reduction. Most poor people depend for their livelihoods on being involved in “a range of markets as private agents or as employees” (Dorward et al, 2003: 320). For smallholder farmers in rural areas of sub-Saharan Africa, crop sales are an important part of livelihoods strategies, as a source of cash income and for productivity, especially when there are weaknesses in many other rural markets – for example credit, insurance and, in some cases, labour. Micro-level studies of farmers who had recently made the transition from subsistence or semi-subsistence to commercialised farming find that commercialisation of agriculture benefits the poor because it directly generates employment and increases agricultural labour productivity. There are direct income benefits, both for ‘commercialising’ households and for agricultural labourers, with

important implications for the way higher average incomes are distributed within a community among various economic and social groups (von Braun and Kennedy, 1994).

There are also links between commercialisation, nutrition, and poverty. Market participation ('commercialisation') that impacts favourably on farm household income can improve welfare, food security and nutrition levels, especially nutritional status of children in the household (see studies contained in the volume by von Braun and Kennedy ed. 1994; von Braun, 1995). Another nutrition-focused literature argues that commercialisation has in fact a detrimental impact on the welfare of poor people. For example, two Integrated Rural Development Programme (IRDP) nutrition studies in the mid-1980s in Northern Province, Zambia, found that while the more commercialised farmers tended to grow more maize and millet both for household consumption and for sale, the net impact on household nutrition was negative (Table 1.1).

Table 1.1 Nutritional Status of Children under five years of age by farmer category⁶

Category of farmer	Number of 90kg bags of maize sold to provincial marketing union	% adequate nutrition	% mild malnutrition
Subsistence	0	70	26
Emerging	1-30	52	41
Commercial	30+	50	44

Source: IRDP (Serenje, Mpika, Chinsali Districts) Nutritional Impact Study (March 1985 and June 1985) – cited in Moore and Vaughan, 1994: 180.

However, reviews consider studies such as these to be flawed and one cannot generalise in asserting that commercialisation is bad for poor people in terms of household food consumption and nutrition (see Longhurst, 1988; von Braun and Kennedy, 1986). Many such studies are based on very small and potentially biased samples, and conclusions reached may not apply to all groups in all places and at all times. There is also a tendency to make simplistic comparisons of nutritional status with and without cash crops, without taking into account that incomes are not a sufficient condition of nutritional adequacy. Other things matter, for example: policy bias against poor households, especially credit and extension; risk and other barriers to adopting productivity-enhancing technology; input supply policy especially seed and fertiliser for subsistence crops; the importance of morbidity and especially health and sanitation.

⁶ Data are for a sample of 205 households with a total of 166 children aged between 6 months and 5 years (Moore and Vaughan, 1994: 180).

Markets play a crucial role in livelihoods development and poverty reduction of smallholder farmers. In poor, remote areas, markets are key in linking the local economy to wider regional, national and global economies. Inter-market linkages are important and there are often mutually reinforcing relationships when accessing markets for different things. For example, access to financial markets can increase (terms of) access for the poor to assets in other basic markets, and vice-versa (DFID/OPM, 2000). However, market access can be a problem for smallholder farmers. In remote rural areas markets are often absent or do not function well and are associated with high transactions costs and increased risk. This is discussed in more detail in chapter 5.

Smallholder farming households face many constraints that hamper market participation. First, while markets can be a highly effective mechanism in exchange, co-ordination and resource allocation, they often fail. Poor people often identify problems with markets as a major impediment to successful livelihoods activities, and these problems may relate to both the absence *and* the effects of markets (Dorward et al, 2003:320). For example, when food markets are absent households need to be self-sufficient in food, which in turn constrains their ability to reallocate land and labour to 'cash' crops. This may be exacerbated by margins between a high buying price and a low selling price (Sadoulet and de Janvry, 1995). Where the effects of markets are concerned, the increase in competition that comes with greater participation in markets can be a double-edged sword as prices are driven down and some producers are unable to compete. Productivity gains need to be high enough to offset income lost through lower prices.

Second, not only may the risks and costs of participating in markets be too high, but also there may be social or economic barriers to entry. Poor people and others, for example particular ethnic groups, may be unable to gain access or are otherwise excluded. Markets can discriminate against poor people because they lack economic assets, political or institutional rights or the right social connections. Regulations and social norms, for example via gender proscribed roles and other rules governing property rights or labour market participation, may be biased against poor people and supportive of more powerful players (see OECD, 2006; IFAD, 2001).

Given there are potential barriers to entry, what influences households' decisions to participate as sellers in the market for agricultural products? The direct benefits of markets – that is, profit-making and thus potentially higher incomes - create incentives to participate with direct benefits outweighing costs. Participation depends on ability to overcome costs of participating caused by thin or missing markets. This is often achieved through informal channels and institutional arrangements such as contacts and networks, enabling people to access markets and to potentially achieve better terms in markets. Social context and relationships therefore are key, and pose significant challenges for conceptualising and modelling economic relationships. The concept of social capital in economics goes some way towards this, and the thesis explores this literature in detail in Chapter 2.

There are also underlying processes of development and change in agriculture. As economies grow, farming systems are transformed from 'informal' subsistence consumption and production to more formal, exchange-based systems. Economics tells us that this process engenders rising opportunity costs of family labour because of increased off-farm opportunities, and increased market demand for food and other agricultural products triggered by urbanisation and/or trade liberalisation. This is closely linked to processes of diversification, from staple food production and subsistence agriculture into diversified market-oriented production systems, due to "rapid technological change in agricultural production, improved rural infrastructure, and diversification in food demand patterns" (Pingali and Rosegrant, 1995; p172). Although the research considers output marketing only in the short term, not agrarian transformation, these processes are relevant. The snapshot of the three villages presented here is embedded within these longer-term processes and the bigger picture of an increasingly more globalised, dynamic and changing agri-food system.

1.3 Social Institutions, Economic Decisions and Livelihood Strategies in Rural sub-Saharan Africa

The role of social institutions in people's economic decisions and livelihood strategies in rural sub-Saharan Africa has been well documented in the social sciences. These are often crucial in mediating access to resources and economic opportunities where economies are under-developed, characterised by cash and credit constraints, and the majority of people are poor with few or no assets and limited access to markets. Early

anthropological studies document the way social networks, relations and norms can be central to decisions around participation in labour exchange, markets and migration (see Richards, 1939; Watson, 1958; Long, 1968; Harries-Jones and Chiwale, 1962). There is also clear theoretical support for the role played by repeated social interactions in reducing transactions costs, solving free-rider problems and curbing opportunistic behaviour. Even criticisms of the social capital literature have not questioned the importance of the social context per se, but rather its treatment within the social sciences (see Durlauf, 2002a, 2002b; Fine, 2001).

The many ways that people in African societies enhance their standing in networks thus improving access to resources and assets are documented in various studies, see for example Glazier's 1985 study of the Mbeere or Parkin's classic study of Giriama community near the coast of Kenya (1972). More recent institutional analyses find labour exchange and markets to be heavily embedded in social structure and, historically, networks, social institutions and social relations have been key to control over resources and accessing labour and goods markets in rural Africa (Berry, 1986; 1989; 1993):

“People's ability to generate livelihoods or increase their assets depends on their access to productive resources and their ability to control and use resources effectively. Access depends, in turn, on participation in a variety of social institutions, as well as on material wealth and market transactions...Since pre-colonial times, Africans have gained access to land, labour, and capital for agricultural production both through exchange and through membership and status in various social units. Rights to occupy, hunt, administer or cultivate land, for example, were often contingent on membership or status in a compound, descent group or community”.

(Berry 1989: 41)

Well documented, too, is the role social networks play in allowing migrant workers to access opportunities and resources and as a source of social support, through kin or other contacts, in both urban and rural areas (De Haan et al, 2002; Berry, 1989). The decision to migrate, where to migrate and how to cope in a new location can be influenced greatly by the ethnic, kinship and friendship networks in which people are involved, both at the origin location and destination, via network information effects (Dasgupta, 2000; Lucas, 1997. See also Iversen et al, 2009; and Iversen, 2006 on South Asia).

A comparative study of four societies (Ghana, Nigeria, Kenya, Zambia) shows that status or influence within a network depend on “frequent demonstrations of one’s ability to command a following”. This makes it easier for people to mobilise labour through the network, and make claims on fixed assets including land (Berry, 1993: 147). Actions to enhance social standing, and thus the ability to call on elders, include investing in ceremonies, community organisations and human capital. Contributing to community groups and projects (for example cooperatives, village maintenance groups, religious associations, and self help groups), investing in the careers of dependents and advancing those of kin and clients not only expand someone’s following but also demonstrates commitment to the network, and further strengthens prestige (Berry, 1985:82).

In Ghana and Nigeria, prestige may be enhanced by building a house in one’s ancestral village, used either for rental capital or as a form of ‘symbolic capital’, always ready for dependents and guests, and a sign of commitment to kin and community thus giving the owner the right to make claims on the loyalty and resources of relatives and kin (Berry, 1985: 78, 181-182; Okali, 1983: 174). Investing in public goods, such as community services and infrastructure, can also create opportunities to access resources outside the immediate locality.

In rural Zambia, existing social relations and social constructions of the meanings of work and obligation provide the context for mobilising labour and negotiating employment terms. For example, Gatter’s (1990) anthropological study of a village economy in Luapula Province in Northern Zambia finds kinship to be the foundation for economic life, alongside a range of formal institutions. However, social underpinnings are not confined to kin. Group membership also provides access to other people’s labour – especially important in farming communities where seasonality and the scale of land preparation mean labour-intensive tasks need to be carried out in a relatively short time period. Participating in work groups and group labour can also provide access to information. A study of Mambwe people, also in Zambia, describes the way women join non-reciprocal work groups on the farms of influential men in order to network with other women to exchange trading and market information (Pottier, 1988: 127). However, while there can be positive benefits from engaging in agricultural work groups, in some cases they can serve to reflect and reinforce inequalities between richer

and poorer farmers. This is particularly the case in locations where most farmers tend to grow the same crops. Those with relatively low status and influence and thus less command over labour, i.e. 'worse' social networks, can find labouring for others conflicts with work in their own fields, that the timing of their own cultivation suffers and, as indicated in one study of North-Eastern Zambia, "their share of rising agricultural proceeds lagged behind their contributions to increased output" (Berry, 1993: 155).

Network Fluidity and Commercialisation

Social networks and network boundaries, by their nature, are fluid, changing and negotiable. People can increase their social memberships over time, or limit the size of their network for example by closing 'borders' of networks. Migration and internal movement also impact on network size and composition. Historically, in Northern Zambia when households and villages split up inhabitants would scatter among new settlements rather than moving as a group (Richards, 1939). The Zambian "village is not a permanent social entity but rather an institution through which a large and varied company of people pass at different speeds" (Kay, 1966:33 *cited in* Moore and Vaughan, 1994). Networks also tend to be organised along different lines, for example: kin and descent groups; location of origin; religious affiliation; occupation; education level; shared political interests. People switch their focus from one network to another as circumstances demand and may also react to a deterioration of their position in one social network by joining another, creating new networks or multiplying their memberships.

Agrarian transformation, commercialisation of rural economic activity, and investment in education, have altered fundamentally the nature of rural social networks in Africa, particularly in transforming the meaning of seniority. Headmen in North-Eastern Zambia, for example, found their 'followings' becoming less stable because of increased availability of alternative status symbols that act as a means by which people can acquire a following and exert influence, for example education, church affiliation and imported goods (Kay, 1967 *cited in* Moore and Vaughan, 1994).

The way smallholder farmers in sub-Saharan Africa acquire and use productive resources have also altered over time, affected by commercialisation processes,

population growth, and social, economic and political factors. For example: “in many rural areas commercialization has been accompanied by relative dispossession of women with respect to agricultural capital” (Berry, 1993: 163). The role of social networks as channels of access to resources (land and labour) has also altered. These processes combined have brought about changes in agricultural production, though implications for patterns of agricultural change, notably in access to labour and thus agrarian class formation, vary by location:

“...while access to land, and fixed capital goods remained closely linked to membership in social networks, access to labor has become increasingly individualized. Because land, labor and capital are combined differently in different processes of production and exchange, the effects of farmers’ investments in social networks have varied from one locality to another”.

(Berry, 1993: 166).

Increased marketisation of agricultural production would suggest a shift towards hired labour through more formalised, market arrangements away from ‘network’ labour of family, kin and other network partners, bound by social norms such as reciprocity and obligation. However, empirical evidence suggests there has been no clear, definitive transformation of family into market labour. Berry’s institutional analysis found in Ghana and Nigeria a shift to hired farm labour in the process of commercialisation of smallholder farming, and “modest upward mobility among farmers and farm workers”. By contrast, while the processes of resource access through social channels operated in similar ways to the West African cases, with farmers continuing to invest in networks for resource access, in Central Kenya and North-Eastern Zambia agricultural employment historically appeared to continue to be differentiated by class and gender, remaining less ‘commercialised’ in the market sense (Berry, 1993:180).

The next section locates the thesis research in the context of Zambia’s economic development and agriculture sector development, drawing on historical literature on commercialisation in Zambian smallholder farming.

1.4 The Study Area

This section describes the geographical and policy context for the study, including: economic development, the character and extent of poverty, and agricultural policy and development in Zambia, before focusing on the background to smallholder agriculture in Northern Province incorporating both economic and sociological perspectives.

1.4.1 Economic Development and Agricultural Commercialisation in Zambia

Econometric analyses in the research use data collected during the 2002/2003 growing season, a time when Zambia was experiencing renewed economic growth following a period of economic reform, low growth and increasing poverty levels throughout the 1990s. This section sets out the historical background to the study period, focusing on economic policy, poverty profile and agricultural development.

Economic policy and economic development

Rich in mineral resources, Zambia was relatively prosperous at independence in 1964 by virtue of copper income. The post-colonial government centrally managed the economy, focusing on nationalisation, subsidising major industries including agriculture, with import substitution and infrastructure investment sustaining the economy. With a thriving economy, growth prospects were good: over the first 8 years following independence GDP grew by an average of 4.2 per cent (World Bank, 2007).

However, global crises impacted sharply on Zambia's copper-dependent economy. The 1973 oil price shock, followed by a slump in copper prices in 1975 led to aggregate economic decline. The Zambian economy was beset by high and variable inflation; chronic budget deficits; falling export earnings; deteriorating balance of payments; and a constant decline in GDP growth per capita. GDP per capita fell by 15 per cent between 1976 and 1979 and external debt per capita grew by almost 60 per cent between 1975 and 1979 compared with a 27 per cent increase over the period 1970-1974 (UNDP, 2003; World Bank, 2007). The unstable macroeconomic environment was compounded by the state's increasingly chaotic management of the economy and Zambia's role in regional opposition to apartheid in South Africa (Wood and Shula, 1987:300-6).

By the mid-1980s the international financial institutions pushed Zambia to restructure the economy under a structural adjustment programme (SAP). However, reforms were only partially implemented. By 1990, Zambia's income per capita had fallen to almost a third of its level at independence. Following elections and a new government led by the Movement for Multiparty Democracy (MMD) in 1991, reform efforts were renewed

and implemented more systematically and extensively, focusing on privatisation and deregulation in a bid to stabilise the economy and foster economic growth. Reforms included: privatisation and parastatal reform; removal of exchange controls; trade liberalisation; liberalisation of the banking sector; decontrol of interest rates; maize and fertiliser marketing liberalisation and decontrol of agricultural prices.

However, reforms did not halt, much less reverse, Zambia's decline as expected. The 1990s overall were characterised by poor economic performance and high levels of poverty. Over the period 1990-1999, the average annual growth rate was only 0.4 per cent, compared to the sub-Saharan African average of 2 per cent (Figure 1). GDP per capita declined by an average of 2 per cent per year between 1990 and 1999 period (Figure 2), and population growth was high, averaging 2.4 per cent per annum between 1990 and 2000. High and unsustainable levels of external debt meant that by the mid-1990s Zambia was one of the most highly indebted countries in the world at US\$720 per capita in 1996 compared to the Sub-Saharan African average of US\$385 per capita (World Bank, 2007). Poor economic performance was further compounded by the effects of major droughts on the agricultural sector in the 1991/92 and 1995/96 growing seasons and poor rainfall during the 2000/01 and 2001/02 seasons (UNDP, 2003).

Figure 1.1: GDP growth (%) – 1961-2010

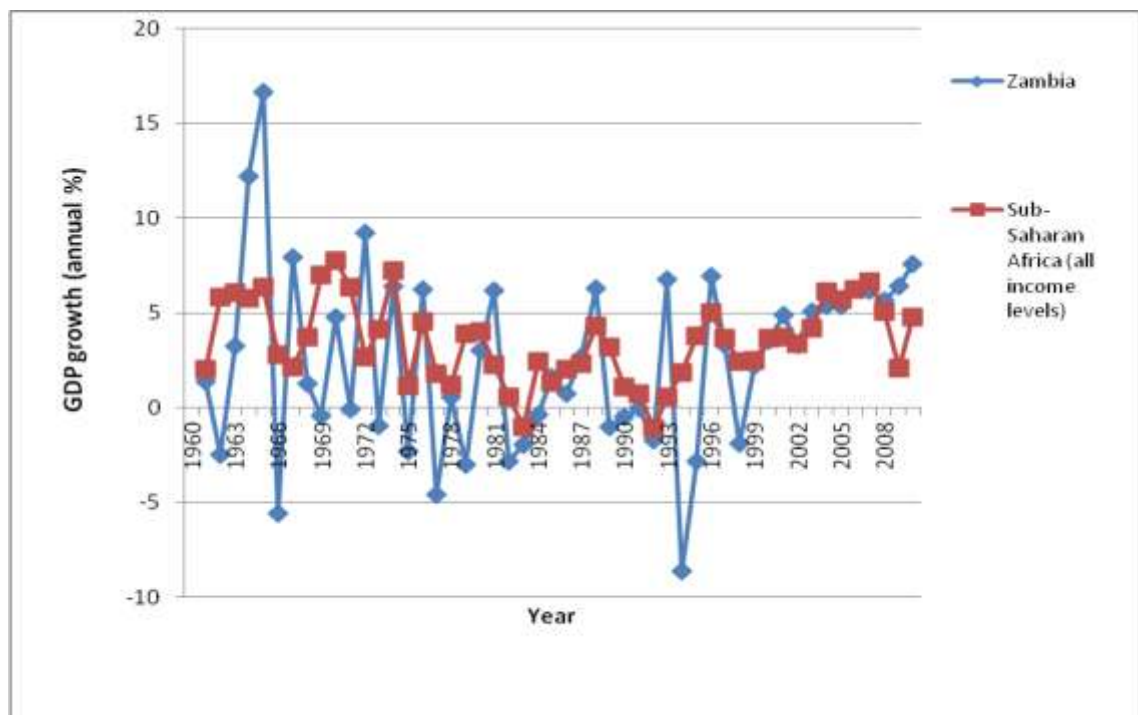
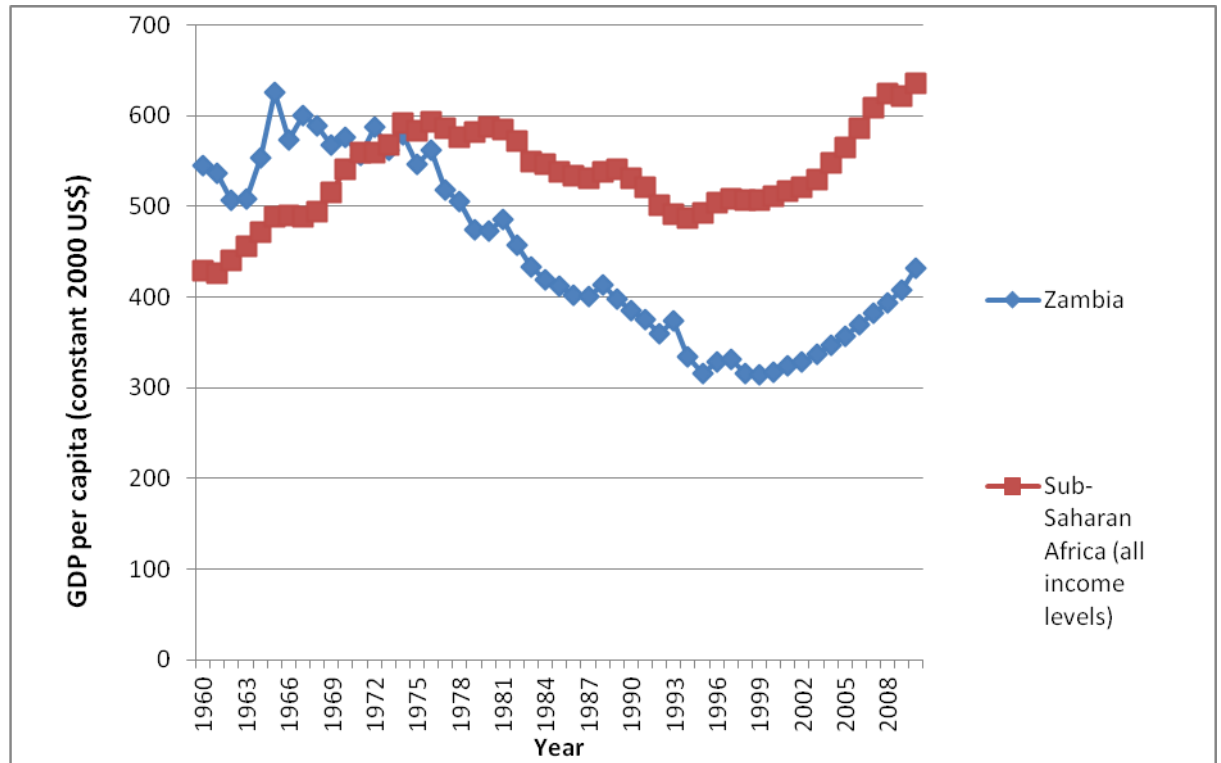


Figure 1.2: GDP per capita – 1965-2010

Zambian smallholder agriculture

During the period of the empirical study, agriculture was the mainstay of Zambia's economy, providing 70 per cent of total employment, 23 per cent of total GDP and 25 per cent of total exports (World Bank, 2007; Govereh et al, 2006). The latest available sectoral employment figures (2004) put agricultural employment at 85 per cent, and industry and services at 6 per cent and 9 per cent respectively⁷. As well as providing the most important source of income and employment for the rural population there are also important linkages to other sectors of the economy and urban populations through multiplier effects and food prices. There is thus a strong connection between agricultural development, rural growth and poverty reduction. Indeed, agricultural growth is the main conduit for poverty reduction and increasing rural incomes in Zambia.

Subsistence agriculture is Zambia's biggest private sector employer. With a population of just over 11 million people in 2002 (11.1 million, World Bank, 2007), 2 million

⁷ CIA World Fact Book <https://www.cia.gov/library/publications/the-world-factbook/geos/za.html>; Last accessed 12th September 2011.

people were estimated to be subsistence farmers. Table 1.2 below sets out a typology of agricultural producers in Zambia.

Table 1.2 Typology of Agricultural producers in Zambia

	Approximate Number of Producers	Approximate Farm Size	Technology, Cultivation Practice	Market Orientation	Location	Major Constraints
Small-scale Farmers	800,000 households	< 5ha (with majority cultivating 2 or less ha of rain-fed land)	Hand-hoe, minimal inputs, household labour	Staple foods, primarily home consumption	Entire country	Remoteness, seasonal labour constraints, lack of input and output markets
Emergent Farmers	50,000 households	5 – 20 ha	Oxen, hybrid seed and fertiliser, few with irrigation, mostly household labour	Staple foods and cash crops, primarily market orientation	Mostly line-of-rail (Central, Lusaka, Southern `provinces) some Eastern and Western provinces	Seasonal labour constraints, lack of credit, weak market information
Large –scale Commercial Farmers	700 farms	50 – 150 ha	Tractors, hybrid seed, fertiliser, some irrigation, modern management, hired labour	Maize and cash crops	Mostly Central, Lusaka and Southern provinces	High cost of credit, indebtedness
Large Corporate Operations	10 farms	100 ha and more	High mechanization, irrigation, modern management, hired labour	Maize, cash crops, vertical integration	Mostly Central, Lusaka and Southern provinces	Uncertain policy environment

Reproduced from Alwang and Siegel (2005): Source: World Bank (2003, p.66 – 67), Francis, et al., (1997, p.13).

Using nationally representative data from the PHS and associated supplementary surveys, Zulu et al, (2007) examine crop production and marketing patterns of Zambian small and medium-scale farmers, including regional patterns and comparisons between poor and non poor producers, for the 1999/00 and 2002/03 growing seasons – coinciding with the period of empirical study of the thesis. For the sample as a whole, crop production made up 72.5 per cent of household income during 2002/2003, with the remainder comprising animal products (5.1 per cent) and off-farm income (21.7 per cent). Over one-third of crop production (35.1 per cent) was in cereal (maize, sorghum, millet, rice), and 17.7 per cent roots and tubers. On the whole, farmers grow a diverse variety of crops:

“While maize remains the dominant crop in production, income from crop sales are considerably more diversified. In particular, there appears to be a great rise in smallholder revenue from the sale of fresh fruits, vegetables, and non-food cash crops as compared to maize. In both the 2000/01 and 2003/04 marketing years, horticultural crop sales were roughly equivalent to the value of maize sales nationwide. Sales of animals and animal products are also shown to account for a substantial portion of sales revenue in the smallholder farm sector, with these accounting for about 50% to 75% as much sales revenue as that generated from maize sales”.

(Zulu et al, 2007: 22)

The next sub-section sets out an agricultural policy timeline for Zambia, over the period 1964 to 2003, before discussing the impact policies have had on agriculture sector performance in relation to smallholder farmers and also rural poverty.

Agricultural policy and agricultural development

Following independence in 1964, Zambia’s agricultural policy can be divided into three different periods: from independence in 1964 to the 1974 collapse of copper; 1974 to structural adjustment in the early 1990s; 1990s to early 2000. This subsection sets out the main policy developments to provide the policy context behind the agricultural sector situation during the study period.

i) Independence to the collapse of copper prices (1964-1974)

Following independence in 1964, the United National Independence Party (UNIP) government, under President Kenneth Kaunda, focused agricultural policy on promoting and expanding maize production, aimed at improving smallholder farmer welfare. Rural resettlement plans were revived and new rural institutions were created. The main policies were fertiliser subsidies, pan-territorial and pan-seasonal pricing and government-led maize-buying stations to guarantee marketing services for smallholders nationally. The buying stations were run first by the National Agricultural Marketing Board (NAMBOARD) and later the Zambian Cooperative Federation (ZCF). Marketing boards and cooperative societies were largely inefficient, mainly due to mismanagement although this was exacerbated by the sheer costliness of providing services to remote areas, which led to difficulties in recovering their costs. However, expanding marketing infrastructure in rural areas combined with fertiliser subsidies did lead to a massive increase in maize production during the 1970s and 1980s (see Govereh et al, 2008; Howard and Mungoma, 1996).

By the 1970s, macroeconomic instability and the collapse of the world price of copper (in 1974) meant falling employment and incomes outside agriculture. Agriculture proved to be more resilient, with commercialisation held up as being the main buffer.

ii) 1974 to Structural Adjustment

During the 1970s and 1980s, agricultural policy continued to focus on commercialising the agricultural economy, particularly maize cash-cropping. Input subsidies and pan territorial pricing continued as part of the government drive towards food self-sufficiency. Responding to problems of unemployment and high food prices in urban areas, the 1980s saw a return to rural areas and rural livelihoods of former migrants. However, the high cost of inputs meant that returns to farming were generally low and therefore only a viable livelihood option for those with an additional income source. During this period and continuing to the 1990s, smallholder production was dominated by maize, at 70 per cent of total cropped area (Zulu et al, 2000) and just over three-quarters of the total value of smallholder production (76 per cent), compared with just 10 per cent of total smallholder crop income coming from cassava. All other crops combined made up the remaining 14 per cent (Govere et al, 2008). However, the system was not sustainable and maize unable to be collected by marketing boards was left to rot, while farmers were increasingly paid with promissory notes rather than cash.

iii) Structural Adjustment to 2002/2003

Macroeconomic structural reforms under a structural adjustment programme began in 1991 under Chiluba's government and included: market and price liberalisation; selected sector reforms; interest and exchange rate policies; banking sector reforms; privatisation of state and parastatal enterprises; trade liberalisation; and stabilisation measures in financial and currency sectors. Reforms throughout the decade in maize marketing and trade were controversial: maize prices were deregulated and private trade introduced. However, reforms were partial as government continued to play a limited role in agricultural markets, particularly in relation to ensuring food security during drought years through maize imports (Govere et al, 2008). Subsequent agricultural policy reform from 1996-2001 fell under the umbrella of Agricultural Sector Investment Programme (ASIP), launched in 1995 and implemented in 1996, with the shift to sector-wide approaches (SWAs) and donor and government harmonisation

of policy and integrated sector investment, with the aim of engendering growth in agriculture based on decentralisation. A component of ASIP was the Rural Investment Fund (RIF), which operated through farmers' groups and provided financial support in the form of capital, basic agricultural infrastructure and a matching grants scheme, to smallholder farmers in disadvantaged areas. Funding for small-scale community infrastructure has also been provided since 2000 through the Zambia Social Investment Fund (ZAMSIF), a World Bank supported assistance programme, aimed at alleviating poverty and emphasising community participation.

iv) Agriculture in the early '00s

Partial reforms in agricultural support - mainly, maize production subsidies and pan-territorial pricing, in the early 90s, led to a decline in area cropped to maize and fertiliser use, and consequently production declined. ASIP saw an increase in areas of crops cultivated, and a diversification away from maize as smallholder producers (and consumers) switched to more 'traditional' crops – food staples and export crops - including cassava, sweet potatoes groundnuts, tobacco and cotton, due to substitution effects (Table 1.3). In some regions, notably the north, where rainfall is higher, smallholders have also branched out into horticultural produce as important income sources.

Table 1.3 Growth Rates (per cent per annum) for Selected Crops Produced by Smallholder Farmers, Zambia, 1991 – 2004

Crop	Area growth rate % per annum	Yield growth rate % per annum	Production growth rate % per annum	Sales growth rate % per annum
Maize	0.2	0.5	0.6	-1.8
Sorghum	-0.3	0.7	0.4	2.4
Millet	0.0	0.5	0.6	1.9
Cassava	1.6	1.7	3.3	5.2
Groundnuts	0.9	3.0	4.0	3.9
Cotton	3.6	1.7	5.3	5.4
Soybean	2.9	1.8	4.8	3.5
Sunflower	-0.5	1.4	0.9	-1.5
Sweet potatoes	4.6	2.0	6.6	6.6
Mixed beans	1.8	-1.3	0.6	1.0

Source: Govereh et al., (2006), computed from raw PHS data, Central Statistical Office, Lusaka.

ASIP performed poorly in agricultural production as well as poverty and hunger reduction according to the 1998 mid-term review and the agricultural sector performance analysis. Apparent declines in maize production were certainly policy-driven, given the reduction in government support. However, production of other crops

increased, and smallholder farmers have benefited from the availability of improved varieties in cassava, groundnuts, and sweet potato (Govere et al, 2008).

Despite a general pessimism in policy discourse that the sector had been in decline, empirical evidence from Post-Harvest Surveys (PHS) of small and medium scale farmers conducted by Zambia's Central Statistical Office (CSO) suggests improvements in the agriculture sector and a decline in rural poverty overall (see Zulu et al, 2000; Jayne et al, 2007), despite inherent policy biases against smallholder agriculture and poor response to reform of the economy as a whole.

“At the start of the liberalization process in 1991, 88% of rural households were estimated to be under the poverty line. Following the major drought of 1991/92, the rural poverty rate increased to 92% in 1993. However, since this point, rural poverty appears to have declined markedly, to 83% in the late 1990s, and to 74% by 2003. Estimates of “extreme poverty” in rural areas have also declined ... over the past decade.

...It is likely that the reduction in poverty over the past decade has been driven by the combination of growth of increasingly important food crops, such as cassava, sweet potatoes, groundnuts (and most likely, domestically consumed horticultural crops) as well as the export-led growth in cotton and tobacco, which have helped to buoy rural incomes despite the decline in maize production and the well-documented negative shocks affecting rural livelihoods mentioned earlier.”

(Jayne et al, 2007: v)

Nevertheless, poverty figures are still unacceptably high, especially in rural areas, and aggregate figures are likely to belie what is happening at regional and sub-regional level, and in particular might not be capturing differences between different ‘types’ of household: “The households not selling maize, which make up roughly 75% of the total number of smallholder farms in Zambia, are largely subsistence oriented farmers, selling very small surpluses of other crops, have relatively small farm sizes, are generally further from markets and roads, have relatively little off-farm and livestock-related sources of incomes, and therefore have very low total incomes” (Zulu et al, 2007: 16). These farmers, like those in the research, will also vary in degree of commercialisation, and this will be in relation to sales of food crops such as cassava, sweet potato, beans.

The next section sets out the impact of Zambia's economic policy and performance on the socio-economic status of its citizens.

Character and extent of poverty

Poverty and inequality in Zambia are serious and widespread. Economic mismanagement, the external shocks/crises of 1970s, and inappropriate policies had far-reaching effects into the 1990s, leading directly to and further compounded by debt problems. The poorest people bore the brunt of the impact of policies instituted under the structural adjustment programme. Price rises, especially for food as subsidies were eliminated, meant a greater proportion of people's income was taken up by food expenditure. This was coupled with unemployment, especially in urban areas, a decline in real wages and a worsening of rural terms of trade as the government attempted to keep maize prices low to cushion the impact on the urban population. This was subsidised by foreign borrowing. Further, high interest rates meant financial products were out of reach of the poorest people.

Data from the 2002-03 Living Conditions Monitoring Survey suggest that more than half of Zambians were not meeting their basic needs, with more than a third 'severely poor'. Poverty is more prevalent in rural areas, where it is deeper and more severe especially in remote provinces and districts, compared with urban areas. Rural poor people are mainly engaged in semi-subsistence farming using family labour, and the poorest households tend to be female-headed with high dependency ratios spending a larger proportion of their household budget on food. Land constraints play a large role in rural poverty in Zambia. Despite its relatively low population density there is a shortage of viable agricultural land with access to markets (Jayne et al, 2008). In all, just 14 per cent of all arable land in Zambia is currently cultivated, the remainder being too remote without adequate infrastructure to make it commercially viable (Chizyuka et al, 2006). Table 1.4 sets out selected socio-economic indicators for Zambia and the Sub-Saharan African average.

Table 1.4 Selected Indicators 2002

Indicator	Zambia	Sub-Saharan Africa average
Poverty headcount ratio <\$1/day (per cent)	76	44
poverty headcount ratio <\$2 a day (per cent)	94	75
HIV prevalence 15-49 age group* (per cent)	17	7
Rural population (per cent)	64	65
Rural population density, people per sq km of arable land	136	361
Population growth (annual per cent)	1.8	2
Rural population growth (per cent)	1.4	1
GNI per capita, PPP (current US\$)	801	1698
Life expectancy at birth, total (years)	37	46
Mortality rate, infant (per 1,000 live births)	102	103
Mortality rate, under 5 (per 1,000)	182	173

*2003. All other data 2002. Source: World Bank (2007) *World Development Indicators*.

The following section describes Northern Province, Zambia in relation to smallholder farmer livelihoods, and the role of social institutions in village economic life, both historically and currently.

1.4.2 Northern Province

Zambia's Northern Province was chosen as the focus of the study, under the DFID-ESCOR-funded project of rural labour markets, because of its rich anthropological tradition as well as historical status as a labour reserve for Zambia's once-thriving copper industry. Rural and remote, the provincial capital Kasama is 850km from Lusaka (Figure 1.3).

Figure 1.3 Map of Zambia: Provinces



Source: http://en.wikipedia.org/wiki/Provinces_of_Zambia; Last accessed 13th September 2011.

The dominant ethnic group is Bemba, traditionally a matrilineal tribe. Smallholder farming is the principal occupation of the majority of the Northern Province population, with some large-scale coffee and sugar plantations located along the line of rail. The prevalence of Tsetse fly constrains the rearing of livestock and while rainfall is relatively high (averaging 1,250 mm per annum), soils are acidic and generally poor, acting as a constraint on agricultural production. This also determines choice of cultivation system, which for smallholders is dominated by the traditional *citemene* system – slash and burn agriculture (see Stromgaard, 1985, for a detailed study of this system). Land is cleared of trees and branches burned to produce ash, which is subsequently applied to the soil prior to cultivation to counteract acidity, enhance nutrients in the soil and improve soil quality. This is a shifting cultivation system. Fields are generally cultivated for 3 to 5 years, at which point new gardens are established. Land preparation is labour intensive and tasks tend to be organised along gender lines; men cut the trees, while women collect, pile and burn the branches. The tasks are seasonal with tree-cutting in June-July and branch-burning just prior to the onset of the rainy season in October. Some cultivation is also carried out on *ibala*: gardens or fields fed by fertiliser, where available.

Households grow a range of cereal crops, tubers, fruits and vegetables and cropping patterns vary across the province depending on agro-ecological conditions and available markets and infrastructure. As elsewhere in the country, since the 1990s there have been moves away from a reliance on maize production to a more diversified crop base, in response to the scaling back of government support to maize production, and cassava features prominently for food security and as a relatively low-risk, low-input crop well-suited to the environment. According to the PHS survey data analysed by Zulu et al (2007), across Zambia as a whole, “the value of cassava production is about 40% to 70% the value of maize production. There was an increase in production of 71% between 1992 and 1998 in Northern Province alone. The bulk of this cassava is grown in the northern, more rainfall abundant part of the country. The increase can be attributed to advances in productivity through the introduction of early maturing, pest resistant varieties. In addition to this, the withdrawal of price supports for maize may have led farmers to diversify their energies to a crop that is suited to the agroecological conditions in the northern part of the country, which is cassava” (Zulu et al, 2007: 22).

Non-farming activities are also important for household livelihoods, including hunting, fishing, basket and mat-making, beer-brewing and gathering wild foods, because of the seasonality of farming and for food security. Households gather wild food from the bush such as caterpillars, mushrooms, chikanda and wild fruit, for household consumption but also for barter or to sell for cash. Food (often animal-proteins such as fish and meat) and beer are also used to access labour, especially for mobilising work groups to carry out heavy tasks such as tree-cutting for *citemene*. This ‘food-for-work’ (*ukupula*) for neighbours and kin has a strong reciprocal element.

Social relations and access to resources

Audrey Richards’ classic study of the Bemba Tribe in Northern Zambia observed: “In Bemba society economic transactions are based on a system of personal relationships – ties of kinship or political status” (1939: 226). Consequently, her account of production and consumption systems of the villages in her study was also bound up in accounts of kinship and political systems, going beyond descriptions of livelihoods activities but also considering the social relations of these activities and arguing that they cannot be separated. It is the (social) relations that enable production. As Moore and Vaughan (1994) summarise in their re-study: “these social relations, it seemed, could not be seen in any way as secondary to what people did; rather they were part of the very substance of life, imbricated in every meal and in every day of agricultural labor...The power and persuasiveness of ‘*Land, Labour and Diet*’ lies, in part, in her demonstration of the inseparability of the material and symbolic worlds and of their mutually constituted nature.” (Moore and Vaughan, 1994: xi-xii). The strength of influence of social structure on economic activities in Northern Province and neighbouring Luapula Province villages was also emphasised in Stromgaard’s (1985) study (see also Poewe, 1978; 1979).

This is still true today. Recent work on migration and remittances finds that in Zambia social norms related to networks govern what would normally be considered to be an ‘economic’ action, leading to migration patterns and remittance behaviours that contrast with findings from migration studies that demonstrate how transfers from migrants fuel rural development in sending communities (Cliggett, 2005). In the Zambian context remittances are made in the form of ‘small gifts and behaviours’, for example providing

practical assistance on visits to the sending community: “Rather than provide significant support to relatives in sending communities, Zambian migrants invest in social networks over time through ‘gift-remitting’. These ‘gift remittances’ facilitate options to return to home communities, or to maintain mutually beneficial social ties for both migrants and relatives in home villages” (Cliggett, 2005:35). This is described by the author as gift-remitting in the Maussian sense: “as part of a system of mutual exchange, creating and maintaining social relationships” (Cliggett, 2005; 39)⁸.

For rural households, networks are particularly important for accessing non-household labour during peak times in the agricultural calendar. This means investing in relationships, in both existing and new networks, but successful households also attract kin and this means household structure and composition can be complex. Moore and Vaughan’s study of maize growing smallholders finds that the more commercial households tend to employ relatively more non-household labour compared with their less commercially successful counterparts (Moore and Vaughan, 1994: 193; 222; 226). This is not necessarily confined to kin networks, and reflects the tendency for village structures to differ widely across the province, with kinship ties only one of many motivations for a household’s choice of settlement. This means historically “the structure of Bemba villages was multiplex and strategic, and such villages often contained households whose kinship ties to the dominant matriline were those of affinity and/or allegiance” (Moore and Vaughan, 1994: 195). Various factors have influenced village structure, and hence institutional structures such as networks, in Northern Province, namely: migration into and out of the region, particularly in relation to the province’s traditional role as a ‘labour reserve’ for the Copperbelt mines; colonial and postcolonial settlement policies; the need to be near roads and other infrastructure and services in order to make a living, especially from cash crops, due to the remoteness of the region; and the strategic importance (both political and economic) of mobilising a following.

In emphasising the role of social and cultural norms, social organisation, and social networks in accessing resources in the context of smallholder commercialisation, the preceding discussion highlights the need to go beyond approaches formalised in terms

⁸ See Marcel Mauss (1923-1924/ 1950) “The Gift: Form and Reason for Exchange in Archaic

of individual optimisation problems. It points to the importance of incorporating into analyses a conceptualisation of institutional structures as having history and continuity, rather than viewing these as entities that emerge purely to fulfil economic functions. The next chapter (Chapter 2) reviews literature in economics and economic sociology that attempts to achieve this.

.....

The remainder of the thesis is structured as follows: Chapter 2 presents a review of the social capital, social networks and New Institutional Economics (transactions costs) literature, the latter focused on smallholder agricultural commercialisation. This is followed by a discussion of fieldwork methods in Chapter 3. Chapter 4 describes social networks measures constructed from the data and analyses the social networks in relation to the characteristics of individuals in the three study sites. Chapter 5 discusses social networks hypotheses and sets out an initial exploration of the social networks data in relation to outcomes at the household level using bivariate and multivariate analyses. Chapter 6 presents empirical results of an econometric model of household crop marketing. Chapter 7 concludes.

Chapter Two

Modelling the Social Context: Social Capital and Social Networks

Economic transactions embedded in social relations are prevalent throughout sub-Saharan Africa, with much exchange taking place on reciprocal or kinship bases. This certainly typifies market activity of individuals and communities living in the study areas, as discussed in Chapter 1. Conceptualising ‘economic’ exchanges taking place within the social realm is challenging. So too is methodology: characterising, capturing, modelling and analysing these phenomena. Social interactions, institutions and relations go beyond a narrow, binary way of thinking in terms of possessing or not possessing a certain attribute. Rather, people are embedded within their social sphere and operate on economic, political, cultural and social levels in relation to others who are also socially embedded, in their sense of ‘self’ and in their actions. This chapter explores social science responses to these conceptual and methodological problems, drawing on literatures that cut across both economics and sociology.

Neoclassical economics has long been criticised for its ‘undersocialised concept of man’ (Granovetter, 1985). Two parallel but related strands of the literature relevant to the thesis go some way towards incorporating social concerns into economic approaches to development: social networks and social capital. Social Network Analysis has tended, until recently, to remain firmly rooted in sociology (sociometry) while social capital as a concept or ‘metaphor’ has emerged across the social sciences from its early roots in sociology and anthropology. In this chapter, the way social relations have been incorporated into economics is examined, concentrating first on attempts to define and conceptualise socially embedded resources through the metaphor of social capital (Section 2.1). Second, criticisms of the social capital literature are set out, leading to a discussion of the social networks approach (Sections 2.2 and 2.3). The chapter then considers how these concepts have been put into practice through econometric studies focusing on social networks and social network approaches to social capital (Section 2.4). Specific measures and terminology used in social networks approaches, and how these potentially take us further in analysis by offering more depth than the social capital approach dominant in economics, are discussed in Section 2.5. Against this background, the chapter concludes with a discussion of conceptual frameworks and

theoretical approaches to markets and networks (Section 2.6) before discussing the literature modelling smallholder market participation from a New Institutional Economics perspective, taking a transaction costs approach (Section 2.7).

Empirical studies that have looked at the role of social networks and/or social capital in economic activity, such as trade, market access etc have focused on a narrow set of social networks measures, that essentially just proxy for size of network. There is another branch of the literature, however, for example work on migration and labour and job search that attempts to capture a richer set of information about the characteristics of the networks beyond how large these are. This will provide useful directions for the fieldwork. In all of this there are conceptual issues, in terms of how a network is defined, and also particularly around endogeneity. This chapter highlights some of the attempts to resolve these problems.

2.1 Defining and Conceptualising Socially Embedded Resources: Social Capital

Economic approaches to development have absorbed social concerns by incorporating the concept of social capital. Described as “one of the most powerful and popular metaphors in current social science research” (Durlauf and Fafchamps, 2005: 1641), there has been an explosion of work in this area since the late 1980s. It was first considered in an economic sense in economic sociology, and since the earlier work of Bourdieu (1986), identifying its value, and Coleman (1988), setting out a conceptual framework, it has proliferated, with common ground in sociology, economics, anthropology and other social sciences⁹.

Social capital is manifested so broadly that in economics alone the term has encompassed norms, friendships, networks, trust, civic virtue, community spirit, and has been used to explain diverse phenomena as how people vote, national economic performance and health¹⁰. A recent review states: “Literally hundreds of papers have

⁹ Earlier still, Durkheim linked the value of being connected in community with mental health outcomes (1897). Hanifan (1916), credited with being the earliest proponent of social capital, highlighted community participation as important in improving school performance. For comprehensive reviews of social capital see Durlauf and Fafchamps (2005); Lin and Erickson (eds) (2008); Woolcock and Narayan (2000).

¹⁰ Perhaps the most widely cited work on social capital has been Robert Putnam’s (1995) study of social capital in the USA “Bowling Alone”, about the decline of participation in voluntary

appeared throughout the social science literature arguing that social capital matters in understanding individual and group differences and further that successful public policy design needs to account for the effects of policy on social capital formation” (Durlauf and Fafchamps, 2005: 1641).

What is social capital? A number of definitions abound in the literature, reflecting different disciplinary influences in conceptualising and operationalising it:

“Social capital is defined by its function. It is not a single entity but a variety of different entities, with two elements in common: they all consist of some aspect of social structures, and they facilitate certain actions of actors – whether persons or corporate actors – within the structure. Like other forms of capital, social capital is productive, making possible the achievement of certain ends that in its absence would not be possible” (Coleman, 1988: S98).

“Social capital refers to the norms and networks that enable collective action” (The World Bank).¹¹

“...the rules, norms, obligations, reciprocity and trust embedded in social relations, social structures and society’s institutional arrangements which enable members to achieve their individual and community objectives” (Narayan, 1997: 50).

“... social capital may be defined operationally as resources embedded in social networks and accessed and used by actors for actions. Thus, the concept has two important components: (1) it represents resources embedded in social relations rather than individuals and (2) access and use of such resources reside with actors” (Lin, 2001: 24-25).

“The basic idea of social capital is that one’s family, friends and associates constitute an important asset, one that can be called upon in a crisis, enjoyed for its own sake, and/or leveraged for material gain” (Woolcock, 2001: 12).

Some definitions focus on what it is, its sources, while others focus on what it does: its outcomes and consequences. It is also referred to variously as an attribute of individuals or of communities. It is useful therefore to consider social capital as a multidimensional, overarching term, rather than a singular concept in itself: “a code word used to federate disparate but interrelated research interests and to facilitate the cross-fertilization of ideas across disciplinary boundaries” (Durlauf and Fafchamps 2005: 1642). This has

organisations in the USA since the 1960s. While influential in that it attracted much attention to the links between social relations and economic outcomes, it has more recently been criticised for lack of rigour in terms of being “a piece of social science scholarship (Durlauf 2002b: 272). For work on trust and economic performance see Fukuyama (1995); On health, see Szreter and Woolcock (2002); voting outcomes see Bevelander and Pendakur (2007).

¹¹ See The World Bank’s Social Capital web pages: www.worldbank.org.

implications for how to ‘operationalise’ social capital, as measurement will depend very much on the element(s) of social capital considered relevant in a given situation or setting. The thesis will examine ways social capital has been operationalised in the literature in section 2.4.

2.2 Criticisms of the Social Capital Literature

Given its apparent meteoric rise to prominence across the social sciences, criticisms are also many and varied, both of social capital studies and of the term itself. At the most basic level its tendency to cover multiple levels and dimensions of social phenomena has led to charges of imprecise definition and ‘conceptual vagueness’. Some detractors question whether social capital can even be called ‘capital’ at all, suggesting it is a metaphorical term rather than one that is ‘conceptually rigorous’ with the necessary characteristics to be able to analyse it like other forms of economic capital (see Arrow, 2000; Solow, 2000).

Some studies, most notably in the earlier literature on social capital, used a circular reasoning to argue for the presence or absence of social capital, presenting “evidence of a beneficial group effect as evidence of social capital itself, and consequently to conclude that social capital is good” (Durlauf and Fafchamps 2005: 1644-1645). This inherent tautology makes measurement problematic and statements of positive social capital outcomes are not falsifiable. However, it is now accepted that social effects are not exclusively positive and social capital can produce social ‘bads’ as well as social ‘goods’ (Castiglione, 2005; Durlauf and Fafchamps, 2005; Warren, 2005).

Economic studies of social capital have also been criticised for putting forward ‘unconvincing arguments’ regarding the consequences of social capital and a poor or misleading use of evidence, including overstating empirical implications (Durlauf, 2000). There is some confusion regarding the causes and effects of social capital, with a tendency towards mixing disparate ideas, especially “functional and causal conceptions of social capital” (Durlauf, 2002: F460) and conflating very different notions of individual motivation.

The failure of many studies to distinguish between correlation and causality, even to establish a clear causal relationship at all, is the most serious criticism of the social

capital literature in economics, in terms of its ability to predict and make inferences. This has been pinpointed as a largely overlooked problem in social capital studies focusing on civil liberties and political freedoms (e.g. work by Grier and Tullock, 1989; Scully, 1988; Kormendi and Meguire, 1985) as well as in cross-sectional analyses. Some studies take correlations and statistical significance to imply causality when this may not in fact be the case. Even though studies usually point to significant and positive correlations between social capital and positive outcomes, the direction of causality is questionable. This has been put down to the “absence of any well-delineated theory about the relationship between underlying social norms which embody social capital and the various activities which are alleged to signal its presence or absence” (Durlauf, 2000: 262).

Establishing causal links between social capital and welfare outcomes requires explicit ways of measuring social capital, so one can actually demonstrate it has an impact on outcomes of individuals or groups. Some suggest data must be able to describe and explain endogenous social capital formation to identify social capital effects. Further, if studies do not use comparable observations, and regressors are missing from a specified model there will be unobserved heterogeneity in the residuals, and they are no longer exchangeable because the distribution of a given error will depend on the distribution of the included and omitted variables (Durlauf, 2002: 464). Exchangeability violations are discussed in more detail in Chapter 6.

However, explicit modelling of the creation of social capital, often through proxies such as group formation, may not be possible. At the very least, suitable instrumental variables are necessary, itself one of the biggest challenges in empirical work; but the problem is more general than arguing that social capital is an endogenous variable. The groups in which individuals are organised are often endogenous and there will be various forms of self-selection that empirical work needs to take into account.

A related challenge is distinguishing between individual and aggregate efficiency effects: when external effects of social capital are captured by those outside the network, individual returns will underestimate social returns to social capital. Further, is it even possible to uncover the role of social capital in the presence of other types of social effects? To identify the effect of social capital, other possible institutional

solutions need to be adequately controlled for, and many studies do not address this sufficiently (Durlauf and Fafchamps, 2005).¹² Issues related to model specification - endogeneity, exchangeability, identification, self-selection - will be discussed in greater detail in Chapter 6 alongside estimation of the empirical model.

Disputes over definitions can be thought of as intrinsic to academic debate, and not in themselves reasons to negate the credibility of social capital studies. Further, there may even be a consensus emerging across the social sciences towards a definition of social capital with a concrete, empirical base: “Social Capital refers to the norms and networks that facilitate collective action” (Woolcock, 2001; p13).

Technical criticisms, however, are weightier. They raise questions about whether it is even possible to uncover social capital effects from the kinds of data to which economists usually have access. Improving ‘tangibility’ suggests social capital needs to be more tightly defined and conceptualised with data collected specially that measures social capital along clearly defined indicators that do not confound the social capital ‘resource’ with its outcomes. There is a role here for using mixed methods and/or interdisciplinary approaches: “the credibility of the social capital literature will be augmented when non-statistical evidence is better used to motivate assumptions and suggest appropriate ways for formulating hypotheses” (Durlauf and Fafchamps, 2004: 61).

2.3 Social Networks and Socially Embedded Resources

That economic relations are embedded in the social sphere is not the same as saying people possess social assets taking the form of capital in an economic sense. However, features of people’s social networks – the people they know and the links between them - and the ability to use these networks can make a difference to the way they make decisions, to their access to different capitals and resources. It is important to remember too that networks and resources are not identical; rather, networks are necessary to

¹² In a 2002 review Durlauf examines critically three leading empirical studies claiming to provide evidence to support the role of social capital in socio-economic outcomes - Narayan and Pritchett (1999), Knack and Keefer (1997) and Furstenberg and Hughes (1995) - to ascertain whether the empirical evidence so far does in fact achieve the objective of supporting a “significant explanatory role for social capital”. He finds identification problems in all three papers.

access and use embedded resources. Moreover, “variations in networks or network features may increase or decrease the likelihood of having a certain quantity or quality of resources embedded. Thus, network features should be seen as important and necessary antecedents exogenous to social capital” (Lin, 2008: 58). Different network conditions, for example density (the proportion of ties between people in a network relative to the total possible number of ties) or openness, can provide access to different kinds of resources. These in turn generate different returns. For example, a denser network may mean greater likelihood of sharing very similar resources, whereas more open or sparse networks might mean better access to better or more varied resources or information (Burt, 2001).

History of social networks research

Social networks research dates back to the 1920s and ‘30s, with the development of ‘sociograms’ and the field of sociometry. It very quickly became interdisciplinary. Other branches of the social sciences – sociology, social psychology, anthropology – as well as statistics, mathematics and, later, computer sciences, found the concept of networks useful and developed new network methods based on both empirical and theoretical motivations. Mathematical foundations of network methods were established, encompassing graph theory, statistical and probability theory and algebraic models (Wasserman and Faust, 1994). Social networks have been used by anthropologists as a way of understanding behaviour in ‘complex societies’ (Mitchell, 1974 and 1969). Developments in economic sociology¹³ and economic anthropology have led more recently to concepts being applied in economic studies.

Formal network analysis uses insights from social anthropology and sociometry, and a set of analytical tools have been developed to cope with the complexity of networks and multi-stranded relationships, allowing one to go beyond ‘cultural’ explanations to structural and relational features underpinning exchange behaviour (see Woolcock, 2001). Network concepts and tools are useful for capturing the way exchange takes place in rural economies, with the advantage of being able to be used in many conceptual frames of reference (Homans, 1974).

¹³ See work by Granovetter (1985, 1992), Bourdieu (1977), Blau (1964), Homans (1974) on social networks in economic sociology.

Social networks and social capital

The networks view is compatible with definitions of social capital focusing on the value inherent in social networks, viewing social capital as ‘resources embedded in one’s social networks, resources that can be accessed or mobilized through ties in the networks’ (Lin, 2001a: chapter 2, also page 51). Studies have distinguished between vertical (bridging) and horizontal (bonding) associations between people, often called ‘bonding’ or ‘bridging’ social capital. So, while social capital has been placed in ‘a family of capital theories’, it has a ‘network-based conceptual origin’ (Lin, 2008: 50). Indeed, the role of networks in facilitating exchange has been deemed ‘one of the most compelling empirical findings in the social capital literature’ (Durlauf and Fafchamps, 2004: 61).

Capturing the social sphere using social network analysis could go some way towards addressing many of the criticisms of social capital by providing tighter, more clearly defined and coherent conceptualisation with analytical tools grounded in established sociological roots, allowing more rigorous measurement of social dimensions of economic life. A networks approach seems to fit best what the thesis is attempting to achieve in terms of identifying the potential role played by social relations in enabling smallholder ‘peasant’ farmers to make a living out of farming. Further, social networks analysis is compatible conceptually with a New Institutional Economics Framework, on which the empirical models in the thesis rest. Chapter 4 examines in detail social network analysis, setting out social networks concepts, hypotheses, measurement and indicators used in the research, Chapter 5 analyses social networks in relation to transactions costs, and Chapter 6 sets out an empirical model grounded in New Institutional Economics’ transactions costs approach. The next section explores empirical studies of social networks in the economics literature, including social capital studies focusing on networks though placed under the banner of social capital.

2.4 Operationalising social capital and social networks

Although economists have recently begun to incorporate networks approaches to deal with the social context, there has been relatively little formal modelling of social networks and social capital in the economics literature. Economic approaches have

varied but tend to fall along two main lines: those incorporating variables constructed using network concepts into standard economic analyses such as regression analysis, and those using a model of the network itself. The thesis draws on both strands of the empirical literature.¹⁴

Empirical econometric studies cover a wide field, spanning: game theoretic models based on prisoner dilemma games (Routledge and Von Amsberg, 2003; Annen, 2003); general equilibrium growth models exploring moral hazard and monitoring with social capital incorporated in the form trust and trustworthiness (Zak and Knack 2001); models of cooperation in corporate cultures, again with social capital conceptualised as ‘trust’ (Rob and Zemesky 2002)¹⁵; and a small but prominent literature in development economics based on micro-level studies of households and communities, which will be discussed below.

Having made some of the more considered and technical criticisms of key econometric social capital studies, Durlauf and Fafchamps, in their recent review in the *Handbook of Economic Growth*, state: “In our judgment the role of social factors in individual and group outcomes is of fundamental importance in most of the contexts in which social capital has been studied” (Durlauf and Fafchamps, 2005:1642). Indeed, though there is no consensus on the effects of social capital or social networks on welfare outcomes, many micro-level studies in developing economies find ‘well-connected’ people are more likely to be better off along key welfare indicators such as housing, health and employment than less well-connected people (see Woolcock, 2001; Woolcock and Narayan, 2000; Narayan and Pritchett, 1999).

As well as studies incorporating social networks and explicitly taking a network approach, the thesis also considers those falling under the social capital rubric using network variables - mainly studies of rural households in developing countries focusing on specific phenomena related to community institutional life: membership of village-level groups and associations; group characteristics; and various ‘community-level’

¹⁴ A small body of theoretical work in economics sets out models of network formation. For example, a theoretical analysis of network formation by Bala and Goyal (2000) formulates the process of network formation as a non-cooperative game. A theory of investment and exchange in a network, where the network is a group of buyers, sellers and the pattern of links connecting them, has been developed in Kranton and Minehart (2001). For a review see Jackson (2007).

¹⁵ A wide ranging discussion of the relationship between Social Capital and formal modelling is given in Dasgupta (2003).

measures - as proxies for social capital and therefore potentially relevant to the analysis here. The studies provide some empirical evidence on the value of social networks in the rural context and useful insight into potential network measures that elucidate the role of social dimensions of people's lives in economic outcomes.

2.4.1 Econometric studies of Social Networks in Economic Life

Links between social networks and development have been explored in a range of contexts in individual, household and community level studies in developing countries. Studies typically describing and measuring people's social networks tend to focus primarily on the role of participation in organisations and networks in determining economic outcomes, and on the way bilateral relationships facilitate cooperation, investment and exchange between actors through benefits such as information sharing (see for example: Fafchamps and Lund, 2003; Fafchamps and Minten, 2001, 2002; Isham 2002; Narayan and Pritchett, 1999; Barr 2002). This section reviews empirical work on social networks and: risk-sharing networks; trading and other (socio) economic outcomes; technology and information diffusion; and migration, job search and employment.

The Formation of Risk-Sharing Networks

There is a burgeoning literature applying network approaches to modelling mutual insurance networks in developing country analysis. This mainly focuses on identifying the way reciprocal exchange networks are formed between "self-selected" individuals, outlining the social connections underpinning these networks and identifying who might be excluded and why (Platteau, 1991; Fafchamps, 1992; Fafchamps and Lund, 2003; Fafchamps and Gubert 2002; Barr, 2002; Goldstein, deJanvry and Sadoulet, 2005; De Weerdt, 2002; De Weerdt and Dercon, 2006). Overall, studies find risk is more likely to be shared among small tightly-knit clusters or networks, often, though not necessarily, organised along kin or geographical lines. Links between people also tend to be multi-faceted.

Using data from rural Ghana, Goldstein et al (2005) examine the role of social relations, intra-household as well as community-level, in asking for and obtaining help when faced with shocks: 'social capital' variables used include the probability of knowing any person in the community, years respondent/family have lived in the village; number of

fostering episodes; number of organisations respondent belongs to. The research finds correlates of asking for and receiving help include kinship and membership of religious organisations. This is backed up by other research. Network formation is also explained by kinship, geography, and degree of exposure to risk by Murgai et al (2002) in a study using data from Pakistan. This is interesting because it raises the possibility that social networks are endogenous. If social networks are formed because of say geography and risk, those are the same things that might affect how commercialised a household is. Data from a village in rural Tanzania finds religious affiliation and wealth, as well as kinship and geography, strongly determine risk-sharing network formation (De Weerd, 2004). In a study of risk-sharing networks in rural Philippines, Fafchamps and Lund (2003) find risk-sharing, in the form of gifts and loans, takes place within networks of friends and neighbours rather than at village level, usually for consumption purposes. Network variables used in their analysis include dummy variables for networks of friends, relatives etc. Network characteristics are captured by: Number of network members; Number who own rice fields; Number with craft skills. It is clear from examining the literature that a large focus is on 'numbers' or size measures. The thesis research attempts to take the consideration of networks further by measuring and analysing other network characteristics such as heterogeneity and diversity.

Socio-Economic Outcomes

Effects of social capital on per capita household expenditure as a proxy for household incomes have been modelled for households in rural Tanzania (Narayan and Pritchett, 1999). Social Capital is defined as "the quantity and quality of associational life and the related social norms" (Narayan and Pritchett, 1999:2). Measured for both households and villages, indices are constructed based on membership in groups, characteristics of the groups, household values and attitudes. Trust variables - degree of trust in strangers, kin, village chairmen, district officials, central government - captured using household survey data, are used as instruments for endogenous social capital variables. The study finds village level social capital to dominate individual level social capital, raising per capita household incomes.

One study of households in Kwazulu-Natal in South Africa using panel data constructs an index of individual membership in groups to proxy for social capital, with mixed results in terms of the effects of social capital on household welfare outcomes, in this

case per capita expenditures (Maluccio, et al, 2000). The component variables of the index are: number of groups, gender heterogeneity of group membership, and group. Community social capital levels are computed as aggregates of individual indices. The research finds both individual and community social capital measures are significantly associated with expenditure in 1998 but not 1993. A related study using the same dataset examines the impact of social capital on child height-for-age in the context of risk-sharing and shocks, measuring social capital as number of associations in the community and interaction of family income with community income (Carter and Maluccio, 2003). The research finds social capital measured along these dimensions helps households to mitigate more easily the effects of negative individual specific economic shocks.

A positive impact of social capital on per capita household expenditure is also found for rural households in Indonesia (Grootaert, 2000). Social Capital here is measured across a number of variables: the number of memberships in associations; diversity of associational memberships; an index of participation in group decision-making; measures of cash and time contributions to associations; a measure of orientation towards community. The social capital index is statistically significant, with the most important variables the number of memberships, 'internal homogeneity' of associations and the level of participation in decision-making.

While these studies argue for a positive relationship between associational life and economic and welfare outcomes, others argue that the relationship is more subtle than the variables of choice allow for (Krishna, 2001; Varughese and Ostrom, 2001). In a study of villages in Rajasthan, India Krishna (2001) finds the relationship between social capital and welfare outcomes is sensitive to what he terms 'capable agency' - notions of effective governance. This includes strong leadership in organisations and frequent interactions between villagers and clients. He argues that group measures often used to proxy for social capital will be associated with better social outcomes only when capable agency is present. Group homogeneity measures, often used to proxy for social capital based on the hypothesis that these denote stronger social ties, were found not to predict well levels of collective action in a study of forest users in Nepal (Varughese and Ostrom, 2001). The authors found successful group activity depends on more than

social ties per se, and institutional features such as decision-making structures play a cohesive role when members are heterogeneous.

The effect of social networks on trader profitability (through lowering transactions costs) has also been modelled for agricultural traders in Madagascar, marketing food staples such as cassava, rice, potatoes and beans (Fafchamps and Minten, 2002). Models of imperfect competition and monitoring provide the theoretical motivation for the empirical analysis. Social networks are characterised as “social network capital” with variables capturing: the number of close relatives in agricultural trade; the number of non-family traders known by respondents; and the number of friends and family members who are able to support the business financially in difficult times. The research takes account of potential endogeneity of social networks with a rich set of instruments, mainly personal background variables beyond the control of the respondent or based on past activity, including: age and age squared, various indicators of place of birth, religion, number of brothers and sisters, number of children, profession, education, and business experience of parents, and history of informal lending and borrowing. Results identify three distinct dimensions of social networks that affect productivity measured as value added and total sales: i) relationships with other traders; ii) relationships with potential lenders – both of which increase productivity; and iii) family relationships which, by contrast, appear to reduce it.

Technology and information diffusion

Whether and how networks facilitate the exchange and use of knowledge – termed information externalities in the economics literature - allowing network actors to innovate, has been explored in a number of studies of farmers in developing countries.

Social capital was found to be a statistically significant predictor of household adoption of improved fertiliser in rural Tanzania, with some regional differences (Ishan, 2002). Here, social capital was measured as: ethnic homogeneity (measured at the village level) of organisations in which households are members; level of household participation in organisation decision-making; and the extent to which leaders of village organisations have different livelihoods compared with village members. In terms of social networks these constitute measures of network diversity and thus access to novel information and resources. Research on high-yielding varieties (HYVs) of wheat and

rice in India during the Green Revolution using household-level panel data, found that both farmers' own experience and their neighbours' experience with HYVs significantly increased the profitability of these varieties (Foster and Rosenzweig, 1995).

Neighbours can be vital in sharing valuable information in farming communities. Studies specifically using social network analysis (SNA) include Conley and Udry's (2010) research in Ghana on social learning through networks, focusing on the adoption of new agricultural technologies. Networks of 450 individuals in four clusters of villages in eastern Ghana were mapped and used to demonstrate that when it came to social learning geographical proximity of network partners (being geographic neighbours) does determine communication between farmers, with social learning also occurring between 'information neighbours'. A descriptive study of the influence of social networks on access to information in rice-farming communities in northern Vietnam (Hoang et al., 2006) found agricultural information runs through informal channels, for example kin networks, neighbours and friends. The research found kinship networks to be vital in accessing information, with networks of neighbours playing a key role in disseminating technological innovation. However, the research does not include any causal analysis. Other research suggests that associations also play a role in information sharing and technology adoption. Group characteristics including cohesiveness and motivation of group members were found to be positively correlated with technology adoption in rural Ethiopia and Kenya, based on a study of smallholder networks within agro-forestry projects in four sites using SNA (Darr and Pretzsch, 2006).

Whether social learning leads *initial* decisions to adopt a new crop to be correlated with social networks is modelled by Bandiera and Rasul (2006) in the context of smallholder farmers in Northern Mozambique. Information on sunflower cultivation actually available to each farmer from their social network is defined and measured as the number of adopters among actors' self-reported network of family and friends i.e. people with whom farmers have strong social ties and are more likely to exchange information. Other measures of networks used are adopters in geographical proximity to farmers, and religion of network partners. The research estimates farmers' propensity to adopt as a function of the number of adopters among their family and friends and finds the relationship to be inverse-U shaped. This suggests social effects are positive when there are few adopters in the network, and negative when there are many. Results also

suggest that those farmers with better information about the new crop are less likely to be influenced by adoption choices of others in their network when deciding whether or not adopt. There is no correlation between the adoption decisions of those of different religions, and while there is a correlation between the decisions of those of the same religion, this is not as strong as family and friends.

More recently, Spielman et al (2008) have carried out social network analysis of smallholder innovation networks in rural Ethiopia. Networks were mapped in ten locations of all actors involved with innovation dissemination and measures such as network density coreness (the degree of closeness of each actor to the network core) and degree centrality (the number of ties that the actor has relative to the total number of ties in the network as a whole) of actors were calculated. While the research does not model explicitly the links between various dimensions of actors' social networks and outcomes, results are interesting in terms of the role of institutions. They suggest that public sector plays a dominant role in providing information, inputs, and credit, though not in marketing and transmitting price information, in rural Ethiopia. Private sector on the other hand plays only a peripheral role across all these dimensions of the smallholder farmer innovation network.

Job search, labour exchange and migration

Research into the role played by social networks in labour markets is confined largely to developed country cases, although studies on networks and migration do include developing countries. The social networks and labour market literature encompasses job search and immigration (Granovetter, 1973; 1974; Dasgupta, 2003), self-employment decisions (Allen, 2000, using data from Wisconsin, USA), employee referrals (see Montgomery, 1991, for a succinct review), the role of social structure in determining labour market outcomes (Wahba and Zenou, 2005; Montgomery, 1991; 1992), social networks and earnings (Mortensen and Vishwanath, 1994) and migration (Massey, 1988; Light et al, 1993; Munshi, 2003; Iversen, 2006; Iversen et al, 2009). Work encompasses both the 'strength of weak ties' highlighting the value of networks of acquaintances/ weak ties in job search in strengthening the information base, as well as the value of strong ties in the context of migration.

The formation of labour-sharing arrangements in rural Ethiopia is explored using a networks approach bringing empirical and theoretical work together (Krishnan and Sciubba, 2009). The study stands out in the way it examines the effects of not only the number of links (network size) but also key features of network structure and composition, which the authors term 'architecture'. Using data from a 1994 household survey of rural households across 15 villages in rural Ethiopia, the research tests a stylised model of network formation for sharing labour, first investigating how network architectures affect network structure, then effects on outcomes - total value of harvest in the main season - estimating a standard Cobb-Douglas production function with regression analysis.

Network variables include: *heterogeneity of networks* based on a range of characteristics of network partners including age and sex of household head; household labour quantity and quality; and wealth characteristics, including land and livestock holdings; *network symmetry*: network architectures are defined as symmetric if the number of links of an actor's partners is approximately equal to the number of links the actor has; *clustering and inter-connectedness* - the ratio of actual links connecting each household to the maximum possible number of links. The authors address potential endogeneity of networks by including as instruments variables that may affect network formation but unlikely to have any bearing on productivity. The instruments are related to how households might be embedded in their village and their relative role and include: number of close blood relatives living in the village; whether the household head was born in the village; the average number of years of residence of the head of household in the village. Locational fixed effects are captured by a household's neighbourhood within their village. The research finds that both number of links and architecture play a critical role in determining impact of social networks on output. Focusing narrowly on links alone downplays the impact of labour sharing on output (Krishnan and Sciubba, 2009).

There are strong links between networks and migration, and research finds migrants tend to choose destinations where they have friends, relatives or other networks (Massey, 1990; Lucas, 1997; Winters et al, 2001). New migrants tend to be assisted by people who have migrated earlier, for example previous migration experience and extensive community migration has been shown to increase the probability of migrating

to the US – termed cumulative causation theory (Massey and Garcia Espana, 1987; Myrdal, 1957). Further, other work on Mexico-US migration finds the development of strong community networks override household characteristics in constraining ability to migrate. So, at least until migration is established in a community, kinship and cultural ties and social networks are key elements of the migration decision (Winters et al, 2001). This suggests that the ‘urban job lottery’ is non-uniform and not anonymous or random as Harris-Todaro analysis would suggest. Instead, this can be modelled as strategic behaviour where actors cooperate in the first game, and then reciprocate, leading to ever-decreasing costs of migration and increasing returns.

A dynamic theory of rural-urban migration is set out in Iversen (2006), underpinned and motivated by empirical observations in rural Karnataka. His model suggests that it is not only the rural-urban wage gap that influences migration, but also the strength of village-specific “social network multipliers”. The magnitude of the multiplier effect depends in part on migrants’ ability to take advantage of existing social networks: “The nature and size of these multipliers depend on patterns of social interaction and are sensitive to attributes such as the social fabric, caste composition, destination links and migration histories of rural source communities” (Iversen, 2006:7).

2.4.2 Criticisms of the social networks approach in economics

Social network concepts and analysis go some way towards addressing criticisms of the social capital literature’s attempts to address social concerns. It offers tighter definitions of elements of the social sphere in its focus on the network of individuals and institutions that actors/ households have access to, the links between them and the resources embedded within them. Social network analysis also provides tools to measure the nature and characteristics of links and of actors themselves (called ‘nodes’). This helps to address issues, such as omitted variables, that can lead to biases and other inferential problems when attempting to analyse causal links between the social sphere and outcomes. However, with the exception of the recent papers by Spielman et al (2008), and Krishnan and Sciubba (2009), network studies in economics, unlike the networks literature in sociology, tend not to use any information on the structure of the network, focusing instead on number of links or on group membership.

Thus, challenges remain.

- *Group effects*: How can social network effects be distinguished from other group effects such as information spill-over, or the presence of common features like underlying political or legal institutions? Like social capital, social networks can easily proxy for other ‘social determinants’. The thesis, in clustering observations by village in the empirical analysis, takes account of this. The thesis also does not rely on group level variables alone to represent social networks. This will be discussed in Chapter 4.
- *Missing data*: this is especially a problem when researchers attempt to represent a whole network based on data that is sampled or recorded from an individual (ego) basis, for example omitting network partners who are not part of the sample such as those located elsewhere. This ignores potentially valuable ‘weak ties’ or bridging networks. Studies also potentially miss information on the dependent variable by eliciting information on potential as opposed to real links, or by randomly sampling individual actors’ relationships as well as the actors themselves. Further, “understanding ... relationships, seems to require sampling approaches that cannot be focused on individuals alone” (Santos and Barrett, 2008:14)¹⁶. The thesis randomly samples households, then selects individual respondents from within households (usually the household head or their representative plus one other household member of the opposite sex), who are then asked about their complete network – both the nature of the link and attributes of the network partner.
- *Recall*: Related to the previous point, as with all empirical, survey-based research, respondents’ recall abilities also affect data quality and the extent to which there are missing data (see Brewer 2000). This is more of an issue in studies that attempt to map an entire network, constructing structural variables such as density or centrality. The thesis, however, concentrates on the networks of households sampled in the three study sites, constructing variables on the nature of the links between actors (or egos) and their network partners rather than on the networks themselves. This is a simple, ‘egocentric’ approach.

¹⁶ Santos and Barrett (2008) critique of social networks literature in economics focuses solely on network formation, which is not the focus of the thesis. However, methodological points regarding sampling are still pertinent.

Methodological issues are revisited in more detail in Chapter 3 on Research Methods and Chapter 4 on Social Network Analysis.

2.5 Social Network Analysis

This section describes briefly key concepts in social network analysis focusing on defining social networks, methods used for collecting information on social networks, and social networks measures.

2.5.1 Definition

Social networks are defined earlier in this chapter (Section 2.3) as the people one knows and the links between them. More formally, following the definition of Hanneman, “A social network is a set of actors (or points, or nodes, or agents) that may have relationships (or edges, or ties) with one another” (2000:18). Actors can be either individuals or organisation, and a tie between actors has both ‘content’, defined as the type of relation e.g. employer-employee, and can include information or resource flows, advice or friendship, or for organisations shared members, and ‘form’, defined as the strength of the relation e.g. frequency of interaction (Powell and Smith-Doerr, 1994). Any type of social relation can be mapped as a tie.

2.5.2 Methods

Social network analysis provides a set of analytical tools for studying patterns of structure made up of interpersonal relations with the aim of “uncovering the patterning of people’s interactions” (Freeman, International Network of Social Network Analysts website www.insna.org). The underlying formal theory, methods, and applications are organised in mathematical terms and are grounded in the systematic analysis of empirical data. Social network analysis employs standard data analytic techniques and applied statistics, which usually focus on observational units (cases) and their characteristics, and applying these techniques to data on ties among the units, known as structural variables. The relationship (including presence, intensity or strength of the relation) and not the individual is often the main unit of analysis. The key feature that distinguishes network theory is use of this structural or relational information to study or test theories. However, attributes of the actors may also be included and these

measurements on actors themselves make up network composition. The ‘layering’ of the data in this way results in complicated datasets so analysis needs to be relatively sophisticated, relying on graph theoretic, algebraic and/or statistical methods, some of which can allow for simultaneous analysis of structural and composition variables. Data analysis of whole networks (otherwise known as a ‘sociocentric’ approach) is concerned with measuring structural patterns in relationships and is generally carried out through the construction of mathematical algorithms and the use of specialist social network analysis computer packages such as UCINET or Krackplot¹⁷. However, taking an ‘egocentric’ or personal relationship approach, the focus is on individuals (ego) and the people they know (alters), that is the networks of relationships around individuals. This approach is rooted in the anthropological studies of the Manchester School (see work by Radcliffe-Brown and others).

2.5.3 Concepts and Measures

Social network variables fall into three main categories – structural, composition and affiliation. This section sketches out some of the main concepts in social network analysis. Definitions are discussed further in Chapter 4.¹⁸

Structural Variables

Structural variables describe the structure of the network. They relate to the shape or pattern of links in the network and describe the ties between the actors. Measures include: size of network; network density; geodesic distance; measures of centrality; power; cliques and subgroups.

Affiliation Variables

An affiliation network is specific type of network involving relations between a set of actors and a set of ‘events’ that the actors ‘belong’ to, such as participation in a particular organisation and can extend to informal social occasions. Affiliation variables give the subset of actors that belong to each ‘event’.

¹⁷ The software is readily available from the websites of social network academics. Much of it is free of charge. See www.analytictech.com.

¹⁸ Three key texts setting out social network concepts and analytical techniques are Wasserman and Faust (1994), Scott (1991) and Hanneman (2000).

Composition/Attribute variables

Composition or attribute variables refer to the data on individual actors' attitudes, opinions and behaviour. They encompass characteristics such as age, sex, income, education etc. that are measured as values of particular variables. Attribute data can be converted into relational data.

The thesis takes an egocentric approach to network analysis. In the study sites, personal social networks, operating through individual links and through membership of groups and institutions can serve myriad functions encompassing social support, emotional and material aid, companionship, information, social control, behavioural models and access to resources. This analysis concentrates on the latter - that is, *instrumental* networks: those that effect the transfer of material goods and services between people, including consideration of networks for social support and information, bearing in mind that different functions in networks may not be neatly distinguished. Using a social networks approach in the analysis enables the research to capture the role that social relations play in economic participation in the three communities. Links with individuals and groups that mediate access to: i) labour; ii) markets; iii) social support; and iv) information, are considered, concentrating on first-order, direct ties between respondents and their network partners.

The next section describes different approaches and conceptual frameworks for analysing networks in relation to markets.

2.6 Conceptual Frameworks: Theoretical Approaches to Markets and Networks

The motivation for the thesis research is founded in part on the apparent tension between the way neoclassical economics has tended to view the economy and markets, and subsequently model economic and market behaviour, and how markets are conceptualised and analysed in other social sciences. The discussion in Chapter 1 of smallholder commercialisation, couched in the role of social and cultural norms in accessing resources, highlights further the need to go beyond approaches formalised in terms of individual optimisation problems. It points to the importance of incorporating into analyses a conceptualisation of institutional structures as having history and continuity, going beyond viewing these as entities that emerge purely to fulfil economic

functions. Focusing on economics, economic sociology and economic anthropology, this section briefly sets out different approaches in conceptualising networks as institutions, ways of integrating sociological views of market exchange with economic analysis, and why integrating social networks variables can make a difference.

Neoclassical economics' view of markets is based on arms-length exchange between 'anonymous' buyers and sellers. This is anchored in a formal understanding of economics as a rational choice between the alternative uses of scarce resources, focusing around optimising behaviour of individuals as rational actors who, in pursuing self-interest, maximise some target function, for example utility or consumption, under constraints. This view of economic behaviour, based on weighing up costs and benefits before making decisions, suggests little room for considering social or cultural influences on the choices people make and the actions they take – generally a central concern in sociological and anthropological approaches.

The division between the disciplines, however, is not clear-cut, to which the *formalist* versus *substantivist* debate in economic anthropology attests. The formalist view of the economy accords with neoclassical economics. The substantivist stance, formulated by K Polanyi (1944) opposes this viewpoint, assuming neither rational decision-making by economic actors nor scarcity of resources. It focuses instead on the way economies are embedded in society and culture, and people formulate livelihoods strategies in interaction with social and natural environments as well as material conditions, and not necessarily by maximising utility. Polanyi's work focused on pre-industrial societies with economies embedded in religious, political and social institutions. In this context, prices are assumed not to be set by supply and demand but by tradition (reciprocity) and political authorities (redistribution). This is held up in direct contrast to 'modern' societies.¹⁹

In this regard the substantivist approach also presents a false dichotomy between the economic and social spheres. Some of the earliest work on the importance of social interactions in modern economics is Veblen's (1934) *Theory of the Leisure Class*, which first describes the way people's behaviour conveys signals about their wealth –

¹⁹ See also Marshall Sahlins' (1972) *Stone Age Economics*.

i.e. the principle of conspicuous consumption. Recent approaches focus on: incorporating economic models into other social sciences; incorporating insights from other social sciences into conventional economic models; and network approaches.

Rational Choice Sociology extends neoclassical economic models to topics traditionally the domain of sociologists (see Coleman, Homans, Becker, and others). The approach argues that all social action is rationally motivated and ‘instrumental’, based on assumptions of the orthodox neoclassical view of rationality. In sociology, the basic framework rests on Homans’ exchange theory (1961), with underlying principles drawn from behaviourist psychology, though later work converges with microeconomics. Based on formal mathematical models, analysis incorporates the way changes in social environment (culture, norms, and social structure) influence people’s choices and behaviours and, conversely, how interaction between individuals determines the social environment itself. Key works include Becker’s (1976) *The Economic Approach to Human Behavior* and Coleman’s (1990) *Foundations of Social Theory*. Analysing systematically effects of prices on market behaviour where social interactions are important, Becker and Murphy (2000) include social forces in models of utility maximisation and equilibrium in group behaviour. They find the influence of social forces to be ‘pervasive’, with mutual interaction between social forces and market behaviour, which they call “social markets”. The approach is mainly criticised for its traditional foundations of rational choice analysis and the way it still takes an economic approach to behaviour, in particular in its assumption that existing economic institutions are efficient solutions to certain problems in the market. Further, its individualistic focus means collective action, social norms and social structure do not form part of analyses.

Recognising the economy as part of a bigger social structure, socio-economics also argues for a strong role for society and social relations (including cultural values and norms and politics and institutions) in governing the economy, and mutually reinforcing relationships between economy and society (for example Etzioni, Akerlof, and others). Thus a much broader perspective than neoclassical economics is needed to solve economic problems. Like rational choice sociology, proponents of socio-economics tend to integrate findings from other social sciences such as psychology, sociology,

anthropology directly into economic models – however these models do not necessarily rest on rational choice assumptions.

Rooted in the neoclassical economics of the Chicago school, New Institutional Economics and Transaction Cost Economics (see Williamson, and others) use economic tools to examine social phenomena, complementing a sociological approach in viewing institutions as social constructions of reality but not necessarily taking the most efficient forms. NIE departs from the mainstream in its key behavioural assumption of *bounded rationality* that lies at the heart of transaction cost economics. Bounded rationality posits that, rather than being optimising agents, people are "intendedly rational, but only limitedly so" (Simon, 1961: xxiv), optimising only locally at best. In practice, people cannot make perfectly rational decisions because they are limited by their cognitive ability in processing information and solving complex problems (Williamson, 1985). While bounded rationality does not represent a wholly 'socialised' view of economic actors, in implicitly recognising impacts of wider social context it does go some way towards acknowledging limitations of rationality (De Bruin and Dupuis, 1999). NIE approaches are discussed in more detail in 2.7 below.

The importance of social relations in establishing economic organisation and action are explored directly in the New Economic Sociology (see Granovetter, Harrison C White, among others). Economic action is considered to be embedded in networks:

"Economic action is socially situated and cannot be explained by reference to individual motives alone. It is embedded in ongoing networks of personal relationships rather than being carried out by atomised actors. By *network* we mean a regular set of contacts or similar social connections among individuals or groups. An action by a member of a network is *embedded*, because it is expressed in interaction with other people."

(Granovetter and Swedberg 1992: 9).

Is there a separation between exchange carried out in a personalised way in a network and that carried out at arm's length, mediated by competitively determined prices in a market? Granovetter argues that "complex economic networks do not arise exclusively from technological or economic factors but also have social and institutional foundations that structure ownership, control and exchange relationships in the economy" (Rauch and Hamilton, 2001: 14; see also Swedberg, 1991; Granovetter and Swedberg, 1992). Networks are not confined to purely socially defined relationships, and can be seen as the 'intersection and interaction of personalized exchange with arms-

length exchange' (Rauch and Hamilton 2001: 1; chapter 1 in Rauch and Casella, 2001). People build their social networks through interaction in markets, employment and work and 'economic calculation of a long term gain' (Friedland and Robertson, 1990).

In terms of networks approaches, studies from this perspective find levels of embeddedness in both industrial and pre-industrial economies. The approach fits well with the realities of life in less-developed rural economies, such as rural sub-Saharan Africa. Here, as discussed in relation to smallholder agriculture in rural Zambia, exchange of goods, labour and other economic activities tend not to operate through impersonal market transactions as assumed in neo-classical economics. Personal ties permeate productive relationships, and gift exchange is especially important in allocating resources and providing social insurance among rural African households (Fafchamps, 2004; See also Platteau, 1991; Dercon and Krishnan, 2000; Fafchamps and Quisumbing, 2002; 2003). That is not to say that markets themselves are unimportant. Synthesising a broad range of studies of market institutions in twelve countries in sub-Saharan Africa based on empirical data and economic models, Fafchamps (2004) states: "Markets play a paramount role in Africa, arguably more so than in developed countries. The reason is the relative absence of large hierarchies and the weakness of those that are present...controlling for differences in the domain of gift exchange, markets play a much more important role than in developed economies" (Fafchamps, 2004: 9). This relative absence of large hierarchies in both government and corporate sectors, and consequently weaknesses in contract enforcement, is also put forward in trying to explain the dependence on social networks in markets in sub-Saharan Africa.

2.7 Modelling Markets and Networks

This section expands the discussion of New Institutional Economics to consider transactions costs approaches to modelling market participation in the context of smallholder farmer commercialisation, and explores the role of social interactions and informal institutions in responding to transactions costs and missing markets.

2.7.1 Smallholder crop market participation and transactions costs

There can be poverty reduction benefits of enabling smallholder farmers, like those involved in this study, to develop their farming businesses and become more commercialised. One small but prominent seam of work examines smallholder farmer commercialisation within a transaction costs framework. This framework falls under the umbrella of New Institutional Economics (NIE), which uses the tools of economic theory to examine more social phenomena that normally fall outside the standard conception of 'economics' (see Coase, 1937; Williamson, 1975, 2000; North, 1990). NIE makes a distinction between *organisations* – firms, governments, groups and societies - and *institutions* – the rules of the game, comprising both formal laws and informal social norms governing behaviour at the individual and community level including economic and social interactions. NIE approaches social and economic analysis from the stance that institutions matter.

Transaction costs are the costs of resources used to create, maintain, use and change institutions and organisations, for example costs of participating in the market process (Furubotn and Richter, 1997:40). At their simplest they can be defined as "costs other than price incurred in trading goods and services" (Swedberg, 1990:115), costs such as those involved in searching for information, for example finding the best price or best candidate for a job. They can also be thought of as "the governance of contractual arrangements" (Williamson, 1979). They can be observable (explicit) costs or unobservable (implicit) costs. In the context of production, they are the costs of market exchange borne by the producer associated with, among other things: information about products, inputs and prices; bargaining; making, monitoring and enforcing contracts; screening costs; and transfer costs such as storage and transport.

While rooted in the neoclassical economics of the Chicago school, NIE departs from the mainstream in important respects. One such departure is the key behavioural assumption of bounded rationality that lies at the heart of transaction cost economics. Bounded rationality posits that, rather than being optimising agents, people are "intendedly rational, but only limitedly so" (Simon, 1961: xxiv), optimising only locally at best. In practice, people cannot make perfectly rational decisions because they are limited by their cognitive ability in processing information and solving complex

problems (Williamson, 1985). This contrasts with the assumptions of the orthodox neoclassical view of rationality, focusing around optimising behaviour of individuals as rational actors who, in pursuing self-interest, maximise some objective function, for example utility or consumption, under constraints. While bounded rationality does not represent a wholly ‘socialised’ view of economic actors, in implicitly recognising impacts of wider social context it does go some way towards acknowledging limitations of rationality (De Bruin and Dupuis, 1999). This makes the framework an obvious choice for exploring social-embeddedness of economic transactions.

Empirically, that many households fail to participate in crop markets can be explained by transactions costs (see Hirschliefer 1973; de Janvry et al., 1991, among others).²⁰ Transaction costs can increase costs to a household of observing market information on which to base transaction decisions. They tend to reduce the net benefits of exchange because they affect returns to activities and potentially prices, which in turn influence traded output. They also shape the risk attitudes of farmers (Zaibet and Dunn, 1998), and can limit the participation of small-scale farmers in the market economy (Delgado, 1997)²¹.

Marketing constraints can increase transaction costs associated with market participation. High marketing costs in rural areas in particular are related to the physical aspects of transactions: transport, marketing, packaging, and handling. More remote smallholder farmers tend to face high transport costs due to distance from farm to market. Further, their transaction costs are also higher because of low volumes. This is exacerbated by poor or non-existent infrastructure and high marketing margins (due to monopoly power of traders/middle-men). Policies can also directly affect these costs, for example, those regulating traders or sectoral policies such as taxation of spare parts for trucks (Delgado, 1995). Incomplete and asymmetric information means that information is costly to acquire and use, poor information in turn increases risk and results in high search costs and costs of monitoring contracts (Zaibet and Dunn, 1998: 833). Work by Pingali et al (2005) couches the transactions costs challenge faced by

²⁰ For theoretical foundations of transaction costs see also Sadoulet and de Janvry (1995); Delgado (1991).

²¹ Using a non-separable household model to simulate the effects of transactions costs on aggregate supply and demand and marketed surplus in generic African households, Minot

smallholder farmers in terms of a broad, agri-systems view of agricultural markets. The authors argue that a highly integrated, modern agri-food system with its demands in terms of standards and contractual requirements, means potential market participants (smallholder farmers) are facing “a new set of transaction costs that emerge from dealing with a food system characterized by different rules, regulations, and players” (Pingali et al, 2005: 2).

Combined with weak information flows and a weak institutional environment, marketing constraints can lead to smallholder farmers being locked in a vicious cycle: increased risks and transaction costs ultimately depress the level of economic activity because of investment and market failure, further raising transaction costs and risk of transaction failure, especially where markets are thin and risk and vulnerability are high – an ‘underdevelopment trap’ (Dorward et al., 2004; Dorward, 1999). Market failures due to transaction costs can result in alternative institutional arrangements, for example sharecropping and interlocking markets (see Sadoulet et al., 1997; Bardhan, 1980; among others). Reciprocal (and barter) exchange is a widespread means of obtaining goods and services where markets are missing - but these arrangements can persist even when markets are present and assumed to be more efficient (Kranton, 1996).²² A transactions costs framework thus allows an examination of the kinds of institutions that minimise transactions costs in production and exchange.

Empirical evidence on the role of transaction costs in making one household more commercially-oriented over another suggest conceptual and measurement difficulties. For example, if transaction costs are so high that the transaction does not take place then there is in effect no transaction cost, so how can it be measured? Do transaction costs of observed transactions differ from those that prohibit transactions? Other challenges include how to proxy for unobservable or unobserved transactions costs. (see Staal et al 1997; Dorward, 1999). Developing country studies that have modelled smallholder crop market participation using household models within a transaction costs framework include: Goetz (1992) on coarse grain markets in Senegal, building on Strauss (1984) on the agricultural household market surplus decision in West Africa; Key et al (2000) on

(1999) finds that transaction costs reduce marketed surplus as well as supply and demand elasticities.

²² See Delgado, (1999); Binswanger and Rosenzweig, (1986); Timmer, (1997) for further discussion of alternative institutional arrangements.

Mexican corn producers; Heltberg and Tarp (2002) on agricultural supply response in Mozambique; and Bellemare and Barrett (2006) on livestock markets in Kenya and Ethiopia. Also in Ethiopia, Gabre-Madhin (2001) models the effect of transaction costs on choice of grain brokers. Farm household specialisation and diversification decisions - food crops versus cash crops – in the presence of transaction costs are considered in Omamo (1998a). In a separate paper, Omamo (1998b) explores transport costs in smallholder choice of crops (cotton versus maize or sorghum) using a nonseparable household model on Kenyan smallholder data.²³

Using farm household data from South-Eastern Senegal (142 households across 12 villages), Goetz (1992), models the discrete decision whether or not to participate in coarse grain markets separately from the decision of how much to buy or sell. The latter is modelled as being conditional on participation using a selection model. This builds on Strauss' (1984) model of market surplus decisions by farming households in West Africa. In the Goetz model, transactions costs are proxied by: information (a regional dummy to reflect differential costs in gathering information), poor access (market distance), cart ownership and two interaction terms of the information variable with access and cart ownership respectively. Also included are variables theoretically expected to affect degree of participation, including prices and household demographic characteristics. While the estimation considers and corrects for selectivity bias, other forms of potential endogeneity in the regressors are not addressed. Results suggest that better information increases the probability that a farmer will participate in markets. Degree of participation is associated with access to technology.

Goetz' model has been built on by Key et al (2000), and Heltberg and Tarp (2002), who make a distinction between fixed and variable (or proportional) transactions costs (FTCs and VTCs respectively).²⁴ As discussed later in Chapter 5, this allows identification of the first stage equation in a two-stage sample selection model by excluding fixed transaction costs from the 'outcome' equation of the decision of how much to sell (or

²³ There is also a cluster of work focusing on milk marketing, particularly smallholder dairying in Ethiopia and Kenya and the role of cooperatives in reducing transactions costs by Staal et al (1997); Holloway et al (2000; 2004).

²⁴ By contrast, Bellemare and Barrett (2006), model crop market participation using determinants termed "fixed and variable costs of market participation" rather than explicitly using a transactions costs framework. The motivation behind their work is to examine whether

buy), because they are thought to represent fixed costs of access to markets and market information and thus invariant to the quantity of the good traded: “Once in the market supply is not affected by the FTCs because only the marginal return to production affects production decisions” (Key et al, 2000: 249). The different authors acknowledge on the whole that both fixed and variable transactions costs are unobservable but can be explained as a “function of observable exogenous characteristics...that affect these costs when selling and buying respectively” (Key et al., 2000: 247). Both FTCs and VTCs are therefore proxied by factors that explain these costs. All regressors are assumed to be exogenous. Table 2.1 summarises the way relevant studies of smallholder farmer commercialisation have measured or proxied for transaction costs.

Table 2.1 Transaction Cost Measures in Studies of Smallholder Commercialisation

Author/ Year	Title	Dependent Variable (commercialisation outcome)	Commercialisation Determinants
Heltberg and Tarp (2002)	Agricultural Supply Response and Poverty in Mozambique	Sells/ does not sell; amount marketed, aggregate value a) total sales b) food crops c) cash crops	<i>Fixed Transaction Costs (selection only):</i> Ownership of radio/TV/phone Max education level of HH Head District population density <i>Variable Transaction Costs (selection and outcome):</i> Prices of maize and groundnuts Farm size per worker Dependency ratio Log number of trees Traction ownership dummy Log age of hh head Any member with a paid job Log mean maize yield in province Dummy for risky area Regional dummies Transport dummies Log distance from railway station Log distance from provincial capital
Goetz (1992)	A Selectivity Model of Household Food Marketing Behaviour in sub-Saharan Africa	Sells/ does not sell; amount sold (production minus consumption)	<i>Fixed Transaction Costs (selection only):</i> Information Poor access Cart ownership Information interacted with access Information interacted with cart ownership <i>Variable Transaction Costs (selection, outcome)</i> Prices of coarse grains and rice Equipment owner Transformation technology Ethnicity Number of persons in HH Dependency ratio Age of household head

the market participation decision is joint or sequential. The thesis returns to this issue in Chapter 5.

Author/ Year	Title	Dependent Variable (commercialisation outcome)	Commercialisation Determinants
Bellemare and Barrett (2006)	An ordered tobit model of market participation. Evidence from Kenya and Ethiopia	Volumes bought and sold: Net seller (positive net sales) cf net buyer (negative net sales) cf autarkic (net sales zero)	Gender and age and age-sq of household head Household size Dependency ratio Assets Land Income Herd size % female Tropical livestock units males, females tropical livestock units Log of average prices large and small livestock Animal births
Key, Sadoulet and de Janvry (2000)	Transactions Costs and Agricultural Household Supply Response	Volume bought and sold of corn	<i>Variables appear as Proportional Transaction Costs in the production equation and Fixed Transaction Costs in the threshold equation:</i> Crop transport costs Distance to/from market Sells to/ buys from official source Sells to customer/ buys from grower Owns truck Local member of agric or transport organisation <i>Consumption shifters (in production and threshold equations):</i> Household calorie demand Predicted HH income <i>Production shifters (in production and threshold equations):</i> Price HH Head age over 55 Crop and pasture land (ha) Uses local HYV Uses chemical pesticides or nat/ chem fertilisers Level of mechanisation Access to formal credit Access to common property Livestock Assets index
Heltberg (2001 mimeo)	Commercialisation and specialisation in Mozambican Agriculture	Value of: Food crop sales Cash crop sales Total crop sales	<i>Fixed Transaction Costs (selection equation):</i> Literacy of household head Maximum education District population densities <i>Variable Transaction Costs (selection and outcome):</i> Transport and market access: dummies for: transport; village sells to shop; village sells to trader Household characteristics (selection and outcome): Family size HHH age Amount of land cultivated (ha) Stock of animals owned Number of trees on farm Number of different agricultural tools owned Area characteristics (selection and outcome): Market prices Median yield of maize Flood/ drought dummy; Regions

Drawing on the work summarised in Table 2.1, proxies for Fixed Transactions Costs could include:

- Education/ literacy (as a proxy for ability to interpret and use information)
- Ownership of radio/TV/phone (as a proxy for access to/availability of information)
- Ownership of cart, bicycle, other transport
- Poor market access (for example road condition, proximity of market)
- Local membership of agricultural organisation
- Regional dummy (to capture production risk at local level)
- Whether producer sells to shops, traders, other intermediaries or directly to consumers (to proxy for search and screening costs)

Proxies for Variable Transactions Costs include:

- Assets: Farm size; equipment; stock of animals owned; number of tools owned
- Prices
- Household dependency ratio
- Traction ownership
- Transport availability
- Whether or not any household member has a paid job
- Location (for example household distance to railway or road)
- Input usage; access to and use of technology

There are potentially overlaps between fixed and variable transactions costs, for example, ownership of a cart, bicycle, or other transport could also proxy for transportation costs, especially where farmers transport crops themselves. These costs are variable transaction costs because transportation costs depend on volumes transported – farmers usually pay a fixed amount per sack to traders.

Modelling Mexican household corn market supply, Key et al. (2000) find that farmers selling to an official source and those located in a region with high participation in farmer organisations have lower proportional transactions cost and higher output. Significant and positive determinants of crop supply include higher “local” usage of High Yielding Varieties (HYVs) and mechanisation (as opposed to household usage), access to formal credit and prices. Results suggest that lowering transaction costs by promoting marketing organisations and improving crop transportation would increase

both market participation and participants' production volumes. In terms of methodology, Key et al.'s analysis also suggests that the household decision whether or not to participate in markets should be modelled separately from level of participation (volume of crop sales) as some households will decide not to produce for markets given fixed costs of market participation.

Focusing only on the supply side using a Heckman model, Heltberg and Tarp (2002) model farmers' crop marketing decisions in rural Mozambique. Results of their estimation for all crops (both food and cash) combined suggest that transport ownership, as a proxy for variable transaction costs, is significant and positively related to both the decision to participate and value of sales. However, the authors do acknowledge that the relationship between transport and participation is likely to be endogenous – causality could run in either direction. VTCs are also proxied by distance to railway which is also significant and negative as would be expected, as is the other proxy for VTCs, distance to provincial capital. This is likely to be because the closer one is to the provincial capital the greater the opportunity for and likelihood of pursuing non-farm activities and the less a household needs to rely on farming. Fixed Transaction Costs, thought only to affect the decision whether or not to sell, and not sales volume, are proxied by ownership of radio, TV or telephone (information access – significant and positive, though potentially endogenous) maximum education level of household head (ability to process information – not significant) and district population density (density of information and marketing networks – not significant). The mixed results mean that overall the authors do not find these variables to be convincing.

To summarise, transactions costs typically arise in smallholder farming in sub-Saharan Africa because of imperfections in markets combined with differential household access to information, assets and networks. When these costs prove to be so high as to preclude market participation, markets are said to be 'missing'. The discussion thus far suggests that to participate effectively in markets, smallholder farmers need to be able to overcome constraints due to market imperfections and/or keep transaction costs to a minimum. The next section explores the role of social interactions and informal institutions in responding to transactions costs and missing markets.

2.7.2 Social Networks and Commercialisation

From the preceding discussion it is clear that markets in rural sub-Saharan Africa are not necessarily competitive and do not solely consist of “arms-length exchange mediated by prices” (Rauch and Hamilton, 2001: 2). Instead, “a complex blend of informal and formal institutions” underlies market activities (World Bank, 2002: 179; North, 1995). Much exchange is personalised, operating through networks, and these links between people or organisations can be central to decisions around participation in markets. Importantly, they often prove to be crucial in mediating access to resources and economic opportunities for people operating in economies characterised by cash and credit constraints and where very many people are poor with few or no assets and limited access to markets. These informal institutions, based on social norms and networks, are, for some, central in making market transactions possible, especially for poor people without formal alternatives, in that they potentially lower transaction costs that act as barriers to participation.

Returning to the transaction costs framework, two distinct though related ways of looking at transactions costs in economic analysis can be identified. One posits that an institution’s role is to reduce transaction costs (Williamson, 1985); the other that institutions substitute for missing markets when there is imperfect information (Akerlof, 1970; Stiglitz on sharecropping/ rural organisation: 1974; 1986; 1989). So, conceptually, there are two reasons why institutions, which include social networks, are important in market exchange. There is clear theoretical support for the role played by repeated social interactions in reducing transactions costs, solving free-rider problems and curbing opportunistic behaviour. Social networks potentially offer a buffer against risk in markets and provide direct assistance in the form of access to transport, storage, information and potentially better terms of trade through relationships with actors in marketing chains. In the absence of formal markets, barter and cash exchange through social networks can provide direct access to goods and services.

Social networks can also potentially have negative economic consequences. Obligations towards family, kin and friends may lead to decisions that are not rational in an economic sense, leading to low-level equilibrium traps. The tendency for more successful households to support the households of family and kin who are less

successful, for example through fostering, while potentially increasing access to labour for the fostering household can also increase dependency ratios, household expenditures etc, thus having on balance a negative effect for the household in an economic sense. During fieldwork in the study sites one such ‘successful’ household certainly described just this situation: “You would expect us to be top in this community but we’re not, we’re more in the middle because we have extra people in our household now, and other people in the community expect me to help them”. Other potential negative externalities of social networks/ social capital highlighted in the literature include: fostering crime, behaviour detrimental to health, educational underachievement; social exclusion; barriers to social mobility (see also: Portes, 1996; Fine, 1999; Carroll and Stanfield, 2003; Torpe, 2003; diFalco and Bülte, 2011).

There are many examples in sub-Saharan Africa of the role networks play in sharing market information and lowering riskiness of transactions. Studies explicitly bringing social capital into market decision-making include decisions to use grain brokers by traders in Ethiopia (Gabre-Madhin, 2001) and social capital as a contract enforcement mechanism for traders in Malagasy flea market (Fafchamps and Minten, 2001). Anthropological studies of communities in the same geographical area and of the same ethnic grouping dominating the study sites in this research describe the importance of informal institutions in market exchange. Audrey Richards’ (1939) classic study of the Bemba Tribe in Northern Zambia observes: “In Bemba society economic transactions are based on a system of personal relationships – ties of kinship or political status” (1939: 226). The social relations of production and consumption behaviour of Bemba people in Northern Province have been ‘revisited’ by Moore and Vaughan (1994), and finds similar patterns. During fieldwork in the survey communities, the significance of social interaction and group participation in exchange relationships was evident. People’s lives and economic decisions are embedded in relationships that can be mutual or one-way. Activities such as trading, job search and the hiring of individuals and group labour operate through many formal and informal channels within the communities. These include women’s groups, resource maintenance groups, co-operatives and the Church.

The research hypothesises that one way social networks may relate to commercialisation is in helping to reduce transaction costs, *both* fixed and variable,

facilitating market access for some farmers. This could be through providing important sources of information, or through increasing trust thereby easing transactions and lowering supervision costs. Social networks are analysed in relation to proxies for transactions costs in Section 5.5.2 of Chapter 5. Further, do the characteristics of an individual or household's social networks help to explain whether and to what extent smallholder farmers participate in crop markets? While it is not difficult to establish a relationship or correlations between market participation and social networks characteristics of smallholder farmers, causal channels are less straightforward. Important, too, are the roles social networks potentially play in economic life in the absence of more "formal" institutions that would usually provide information and services. When considered in terms of transaction costs, market failure can be seen to be household- rather than commodity- specific (de Janvry et al, 1991: 140). The tendency towards a high degree of homogeneity between households in close geographical proximity raises questions, which the research attempts to address, of how some households in a particular location can access (better) markets than their neighbours.

To conclude, this chapter has summarised attempts in economics to characterise the social domain in relation to socio-economic outcomes and address what Granovetter has termed the 'undersocialised concept of man'. The literature on social capital represents the bulk of these endeavours and the concept is ubiquitous across the social sciences. However, its ubiquity also contributes to one of its major failings: lack of conceptual clarity, manifested in the way research often confounds what social capital is with what it does. This creates problems for robust analytical work. Social networks analysis, and studies that conceptualise social capital within a networks approach, may offer an alternative.

A social network analytical approach allows features of the institutional environment, not just 'capital' aspects of social networks, to be captured. In this way the network is viewed not only as something an actor uses but also something they experience. Social network analysis provides a set of tools for measuring features of social networks and constructing social networks variables to allow analysis of specific social components. This enables examination of interactions, transactions and exchanges themselves, as well as the attributes of the different actors and the relations between them. In this way

one can uncover how network structures can both provide opportunities and act as sources of constraints. A transactions costs approach to modelling commercialisation offers a conceptual framework that allows incorporation of the social context into the analysis.

The research will define the social context in terms of social networks of individuals, and the institutional context within which these networks are embedded. Anthropological and qualitative data are used to provide insights into individual's choices, which goes some way towards being able to set up a 'causal notion of the social context' (Durlauf and Fafchamps, 2005).

.....

The next chapter describes the rationale for the household survey approach, design of the survey instruments and details of the field research methods.

Chapter Three

Data Collection Methods

The fieldwork for the thesis research took place over three periods totalling four months between March 2002 and May 2003, including a preliminary field visit. The study sites are three villages in Northern Province, Zambia. Individual, household, group and community level data were collected, including social networks data, using a combination of methods and instruments. It was an ambitious exercise, requiring considerable resources and time. However, interpretation of the quantitative social networks data analysis would not be possible without the deeper, contextual information provided by qualitative research. My role in the fieldwork was in research design, supervising the household survey, and supervising and carrying out qualitative research with the assistance of a translator.

This chapter describes methods and methodological approaches used in the research design and data collection, first setting out how the study sites were selected and why (Section 3.1) and ethical considerations in undertaking such a study (Section 3.2). Section 3.3 explains the rationale behind taking a mixed methods (quantitative and qualitative) approach to the research and in particular in collecting the data. Section 3.4 describes the design of the household survey including special modules to collect social networks data, and qualitative instruments. Data collection procedures are explained in Section 3.5, followed by a section on managing a complicated, multi-layered dataset, including data entry and data cleaning (Section 3.6). The chapter concludes with methodological challenges (Section 3.7).

3.1 Study Site Selection

Three sites were selected during a preliminary visit to Northern Province during March 2002. Table 3.1 gives a summary of the survey sites. To capture diversity between sites, two villages²⁵ were chosen purposively according to differing degrees of market integration, proxied by distance from graded road: one with ‘good’ and one with ‘poor’

²⁵ ‘Village’ is defined as the population under an individual headman.

market access. The third site was a resettlement scheme, with intermediate market access.

Table 3.1 Survey sites

Study Site	Estimated Population	Total No of Households	Distance from tarmac (km)	Distance from Kasama (km)
Kabila	1600	200	80	87
Ngulula	670	100	1	25
Lufubu	480	65	200	200

Source: Fieldwork, Northern Province, Zambia, 2002-2003

The main disadvantage of sampling in this way is lower precision of sampling estimates, the sample is less representative of the population as a whole compared with random sampling. However, it was never the intention to obtain a sample representative of Zambia as a whole, rather to focus on ensuring the sample was representative of the population of smallholder farmers. Households were randomly sampled within villages in the case of Ngulula and Kabila, and full enumeration was attempted in Lufubu. The sample of households can thus be said to be representative of the village, or type of village ('good', 'intermediate', 'poor' market access) from which they are drawn. The village was chosen as the 'bounded' population because most labour-related transactions and social interactions are carried out at this level and as an administrative unit is an important focal point for social and political institutions. Confining the sample area to a single province meant that ethnicity was held constant to a very large extent. Bemba is the predominant tribal group in all three sites. It was also necessary from a cost point of view given the low population densities and large distances involved in working in a country the size of Zambia.

Ngulula village, close to the provincial capital Kasama, was chosen as the site with good market access. It is located close to a graded road and noted for its relatively high degree of integration in the market. Of the range of villages visited, Philip Village, situated 300km from Kasama either side of the tarmac road and comprising 180 households, was also considered as a potential study site with good market access. However, in this case proximity to the road is not an indication of market access. While Phillip Village is situated close to the tarmac it is distant from any markets and the population faces similar logistical and transport problems to communities situated off the main roads. Kabila village was selected as the site with little or no market access. It is located some 80km from a graded road and from 'large' population centres and could

only be accessed by a 30km feeder road. Bulunda village, 60km from Kasama with two hundred households, was also considered as a possibility for the study site with little market access but compared with Kabila village it was decided that Bulunda was not remote enough and part of the village was reached by feeder road of average quality at the time of the visit.

A resettlement scheme was selected as the third study site, with intermediate market access. It was also chosen for its interesting social set-up, with the implication that it was likely to be different from a traditional village in the way social networks are characterised because people were more likely to originate from different areas and less likely to be living in close proximity to kin. It could be hypothesised that within-community kinship ties would be much weaker here compared with so-called ‘typical’ villages, but also interesting is whether or not the same types of social and power relations and interactions are established once removed from the ‘traditional’ setting. Out of a total of eight Rural Reconstruction Centres in Northern Province, two schemes were visited: Lufubu Resettlement Scheme and Lukulu South Resettlement Scheme. The latter was 28km from Kasama with 180 households in all. Many of the plots were owned by people working in the civil service and based in Kasama and farmed by full-time managers. The inaccessibility of schools and other amenities meant that it was difficult for families to settle here so plot managers tended not to have families. It was decided that Lufubu Resettlement Scheme would be more suitable as a study site – it was smaller and appeared to have more characteristics of a ‘village’. Lufubu has restricted access to markets for produce and is situated 40km away from the nearest town, and relative to Kabila and Ngulula has intermediate market access.

3.2 Ethical Considerations

There are many stages in the fieldwork process where one is in danger of making ‘ethical’ transgressions. First, when entering the field there is a possibility of misrepresenting oneself and the research and generally misleading people in the study communities, whether intentionally or unintentionally. While working in the field, exploitation of respondents is very possible – engaging people in interviews that take an inordinate amount of time without giving them the opportunity to decline or coercing

people into taking part when really they do not want to. Finally, when exiting field it is possible to leave in ‘bad grace’ – without thanking the communities or keeping them generally informed, for example what the next stages of the research are or what the research team is going to do with their personal information. Linked to this is the overarching issue of maintaining confidentiality. People will generally have agreed to participate in the research and provide personal information on a basis of trust that this information will not be used to their detriment.

The research was considered to be ‘low risk’. This includes: level of vulnerability of participants; the potential for inducing stress of any kind in participants or generating other negative consequences²⁶. A strict protocol was also followed during both rounds of the fieldwork. On entering the communities meetings were convened with community members with a policy of inclusiveness – everybody in the community was invited to attend and participate. The intentions of the research were not concealed, research affiliations were declared and care was taken not to make promises that could or would not be kept. The research process was explained very carefully, community members were assured that they were under no obligation whatsoever to participate and questions were invited from community members that the research team endeavoured to respond to. During the fieldwork itself, prior to interviews for both the household survey and the qualitative group and individual interviews the objectives of the research were reiterated to ensure that participants were fully informed, and they were given the opportunity once again to ask questions and/or decline to participate. Demand on participants’ time was kept as low as possible and reimbursement was provided at the end of lengthy interviews in the form of bags of salt or bars of soap – goods that people need and value but often find difficult to access through lack of cash or goods to exchange them or inability to get to a shop. Respondents chose the time and location of interviews. There was no coercion on the part of the interviewer and the individuals’ right not to participate was respected, although there were no refusals to take part. The research was conducted with the assurance of absolute confidentiality and data are anonymised in the analysis and presentation of results

²⁶ See http://www.sussex.ac.uk/res/documents/application_form_for_ethical_review.doc for the Sussex Ethics Committee’s seven-point checklist for risk-level of research projects.

3.3 The Need for a Mixed-Method Approach

This section sets out in more detail the choice of fieldwork methods, focusing on the qualitative and quantitative traditions, highlighting their philosophical foundations and the advantages and drawbacks of each. Combined methods approaches are discussed, and the rationale behind the choice of methodological framework for the thesis research.

3.3.1 *Quantitative and Qualitative Traditions*

Traditionally, social science research methods have been divided into two approaches, quantitative and qualitative, often seen as being diametrically opposed to each other. A quantitative approach, as defined by Carvalho and White (1997) is ‘one that typically uses random sample surveys and structured interviews to collect the data – mainly quantifiable data – and analyzes it (*sic*) using statistical techniques’ (1997: 1). An explicit conceptual framework and hypotheses are usually constructed and key variables identified at the outset of the research. The data collected are then used to test these predetermined hypotheses. The quantitative approach is associated with the logical positivist paradigm.

By contrast, a qualitative approach ‘use(s) purposive sampling and semi-structured or interactive interviews to collect the data – mainly, data relating to people’s judgements, attitudes, preferences, priorities, and/or perceptions about a subject’, (Carvalho and White, 1997: 1). The data are then analysed using social or anthropological research techniques. This definition of qualitative data can be widened to encompass textual and visual data generated by interviews and observations and derived from documents, records and other secondary sources. The focus of qualitative data collection methods is on understanding a situation holistically and data tend to be generated through intensive (often repeated) encounters with a small number of people in their natural setting. Many different research traditions have utilised and developed qualitative approaches, giving rise to substantial heterogeneity in collection methods and analysis and consequently no single, ‘right’ way to do qualitative research. Qualitative research is by its nature mixed-method in focus and has been described as ‘bricolage’ – ‘a pieced-together, close-knit set of practices that provide solutions to a problem in a concrete situation’ (Denzin and Lincoln, 1998:3). Qualitative methods are associated with interpretivist and

constructivist traditions that emerged as counter-movements to positivism. The main characteristics of each approach are summarised in Table 3.2.

Table 3.2 Main features of qualitative and quantitative approaches

Characteristics	Quantitative Approach	Qualitative Approach
Theoretical Paradigms and Perspectives	Logical Positivism	Phenomenological/ Post-positivist/ postmodern/ Constructivism/ Interpretivism/ Naturalism
Methods of data collection	Structured, formal, pre-designed questionnaire	Open-ended, semi-structured, participant-observation, ethnography
Perspective	Objective	Subjective
Sampling	Probability	Purposive
Sampling error	Less sampling error, prone to more non-sampling error	More sampling error, tends to reduce non-sampling error
Sampling size	Large	Small
Geographic coverage	Wide	Narrow
Analysis	Deductive/ hypothetico	Inductive
Statistical Analysis	Important part of approach	Plays little or no part. Simultaneous use of several different sources and methods (triangulation) to validate findings. Some systematic content analysis and gradual aggregation of data based on themes emerging from research

Source: Builds on Denzin and Lincoln (1998); Carvalho and White (1997).

Describing differences between qualitative and quantitative methods according to types of data and the way they are collected, however, risks reinforcing a false dichotomy. First, both paradigms set out to collect data for a specified purpose. Further, data collected through surveys are not necessarily quantifiable and surveys have been used to elicit information traditionally thought of as being of a ‘qualitative’ nature (see: Sharp et al., 2003, study of destitution in Ethiopia for an example; Chambers 2003a for a review of studies that incorporate ‘qualitative’ questions into ‘quantitative’ surveys). Similarly, qualitative methods can be used to collect quantifiable data, for example it is customary to code qualitative data for themes, which may then either be interpreted qualitatively or translated into quantitative data and analysed statistically, although this may not be desirable. One study using participatory methods including participatory mapping has generated statistics (population estimates for Malawi, estimates of ‘very food insecure’ people in a population and the proportion of people who should be targeted) from a representative sample that are generalisable (Barahona and Levy, 2003). Study design incorporated statistical principles and Participatory Rural Appraisal (PRA)-type tools were adapted to allow standardisation and comparability of data produced across sites.

3.3.2 Tensions between quantitative and qualitative approaches

Potential conflicts lie in the basic philosophical underpinnings of each approach. At the philosophical level there are fundamental differences based on the ‘assumptions each method typically makes about the nature of reality’ (Christiaensen, 2003: 114), and in the tension between regarding people as passive objects of study versus taking a phenomenological approach (see Merleau-Ponty, 1962). The logical positivist paradigm assumes there exists one ‘single, external reality’, with the implication that it makes no difference who makes observation or how it is made. To capture this single reality as closely as possible researchers need ‘to increase the likelihood of achieving unbiased, objective answers to research questions’ (Christiaensen, 2003: 114), hence the reliance on statistical principles in study design and structure - for representativeness and generalisability of results for the population under study -and on standardisation and quantification in data collection - to solve problems of bias and variability in interviewer-interviewee interaction.

Interpretivist and constructivist traditions differ from the positivist paradigm in many ways²⁷ but in particular they do not support the premise of a single, objective ‘reality’. The starting point is the ‘recognition of a multitude of realities and the belief that objectivity and a value-free science are simply impossible’ (Christiaensen, 2003). Instead, understanding social phenomena is based on social actors’ own perspectives and the important ‘reality’ (‘realities’) is that (are those) perceived by the social actor. Semi-or unstructured, exploratory data-collection methods are used to involve multiple stakeholders to uncover multiple perspectives and meaning to understand contextually the topic of interest, and reality is constructed by the values and experiences of the individuals involved in the research. In the constructivist tradition the analyst goes one step further and ‘seeks to bring about change and empowerment of the stakeholders in the process’ (Christiaensen, 2003: 115).

The philosophical divide is used as an argument against combining approaches, based on the belief that the two methods are founded on incompatible assumptions and therefore cannot logically be used together (see Lincoln and Guba, 1985, and Cook and Reichardt, 1979, for extensive reviews of the paradigms debate). In practice, while the

choice of methods tends to be based on the researcher's own philosophy and assumptions, as well as the research problem, only at the extremes will the two approaches be 'incompatible', and can be summed up as 'dichotomies of objectivity versus subjectivity, fixed versus emergent categories, outsider versus insider perspectives, facts versus values, explanation versus understanding and single versus multiple realities (Christiaensen 2003, citing House, 1994).

3.3.3 Combined methods – transcending the tensions

While most discussions of the philosophies behind different methods reinforce the dichotomy between quantitative/positivist 'one reality' versus the multiple realities revealed through taking a 'qualitative' approach, I would argue that it is not necessary or even desirable for a survey to be designed and conducted strictly in the 'positivist' tradition to retain all the qualities of a quantitative approach. In survey design, more contextual/ qualitative material can inform the content and wording of questions, which can also be designed to allow for greater flexibility in responses than clearly-defined pre-coded categories. In incorporating an 'other' code to record responses that do not fall within those specified by the pre-coded system, with space to note the exact response for coding later, the questionnaire is able to capture elements not previously considered by the researcher, allowing for surprises. This goes some way towards capturing multiple realities and does not necessarily commit the research to the worldview set out in the pre-coded questionnaire led by the conceptual framework.

The degree to which a researcher is confined to imposing one world view through use of a pre-coded survey depends on flexibility in the field and also what has informed the research/ conceptual framework. Survey questionnaires can be and often are designed based on contextual information gathered from a variety of sources including those based on knowledge garnered through qualitative routes. However, the questions themselves are still predetermined with limited scope to change them in the field. In this respect surveys do impose a view of the world from the outset in that the set of questions is specified by framework chosen and the wording of questions is fixed from the start according to pre-identified variables necessary for testing hypotheses and modelling relationships. A flexible approach in the field can allow for modification

²⁷ See Guba and Lincoln (1989) for comparison of dominant research paradigms.

across the board if, from experience, there prove to be better ways of posing questions and eliciting responses. That said, there is little scope for allowing the research objects or variables to define the direction of the research. Questions are generally formulated with hypotheses in mind. If, in the course of fieldwork, a hypothesis proves to be untenable or if relevant issues arise that are not accounted for in the survey, then questions posed will not gather data on phenomena of potential interest, even though the researcher is aware of them.

Nevertheless, the ‘qualitative-quantitative’ distinction is an important one, in that it leads to the question of whether or not a researcher’s paradigmatic views drive or predetermine the set of research methods chosen. This is partly so – as well as differing underlying assumption about the nature of reality, each paradigm makes implicit assumptions about the best way to reveal that reality (or realities) and ‘formulate knowledge about it’ (Christiaensen, 2003: 114). Further:

‘... in using a particular methodological family to explore the subject of interest one also tends to imply certain ontological assumptions about that reality. In this sense the methodological debate is as much about the kind of reality that we want to discover as it is about methodology. This does not imply that both methods cannot be combined, but rather that the manner and the extent to which we use and combine both methods will affect the realities we discover.’

(Christiaensen, 2003: 115)

Equally, it has been argued that all research paradigms are valid and none are inherently linked to methods that have been labelled ‘qualitative’ or ‘quantitative’, a view defended by those who believe that epistemological issues should be separate from those of method (see Reichardt and Cook, 1979; Bryman, 1984). While paradigms are ‘important theoretical constructs for illuminating fundamental assumptions about the nature of reality’ (Patton, 1990: 39), at the practical level it may be more useful to consider the different approaches as research strategies rather than paradigms and use the method that best answers the research questions. Favouring ‘situational responsiveness’ over an allegiance to one paradigm or another means that ‘there aren’t just two paradigm-dictated choices’ but different methods appropriate for different situations (Patton, 1990: 39). In taking this view, it is clear that qualitative and quantitative methods are not necessarily antithetical to each other, nor are they competing, irreconcilable ways of approaching research and, as discussed later in this

chapter, there is considerable scope for using a combination of techniques in designing studies, data collection and analysis. It is hoped that this research will demonstrate this.

3.3.4 Integrating quantitative and qualitative approaches

The ‘desirability and usefulness’ (Hentschel, 2003: 20) of combining approaches has been acknowledged across the social sciences (see Carvalho and White, 1997; White, 2002; Kanbur ed., 2003; Jayaratne and Stewart, 1991) and the limits of taking a purely qualitative or purely quantitative approach have been long recognised – e.g. World Bank poverty measurement and analysis. Depending on the questions being asked the two approaches can be combined to make ‘powerful and relevant analyses’. The complementarities and tensions between the approaches have been highlighted in a collection of writings by participants of the workshop “Qualitative and Quantitative Poverty Appraisal: Complementarities, Tensions and the Way Forward” (Q-Squared) held at Cornell University in March 2001 (Kanbur, ed., 2003). How can combining quantitative and qualitative methods improve the quality of research, retaining the strengths of the two approaches and strengthening each further by utilising the results and insights of the other?

At the analysis stage, triangulation can be used to increase the internal validity of findings. This means using several data collection methods to assess whether any given finding is ‘authentic’. The weaknesses and limitations of respective methods may be offset by the strengths of others, thus exploiting the advantages of each approach while vitiating potential drawbacks. Table 3.3 sets out the relative strengths and weaknesses inherent in qualitative and quantitative approaches.

Table 3.3 Strengths and weaknesses of quantitative and qualitative methods

Quantitative	
Strengths	Weaknesses
<ul style="list-style-type: none"> • Sampling allows aggregation of the data; able to generalise, test hypotheses. Representativeness of the data. • Breadth – usually able to collect data from many respondents with respect to well-defined questions. Strives to be unbiased, reliable and rational Data can be used in models to simulate effects of policy options 	<ul style="list-style-type: none"> • Difficult to capture multiple realities, approach lacks depth and richness (Patton 1990; Lincoln and Guba 1985). • Relative lack of bias is arguable: there are biases, tend to be hidden – starting with the underlying assumptions. Questionnaire makes assumptions at the outset Biases can also arise in selection of indices and later in interpretation of data. • Sampling and non-sampling errors • Can miss what is not easily quantifiable • Failure to capture Intra-household dynamic
Qualitative	
Strengths	Weaknesses
<ul style="list-style-type: none"> • Data more detailed, descriptive and highly contextual. Richer definitions, data cover wider dimensions and some analysis is carried out by the respondents themselves. • Multiple interpretations of reality = refrains from a priori assumptions • Allows important dimensions of a topic to emerge during data collection or analysis (does not preclude hypothesis testing, or qualitative analysis in deductive framework). Easier to incorporate 'surprises' into the data gathering process • Explains causal processes • Accuracy and depth of information. 	<ul style="list-style-type: none"> • Unrepresentative participation and tendency for dominant groups and individuals only to participate, missing out those more marginalised etc, agenda-framing • Selection of sites too few or too atypical – inability to generalise beyond the research area/ study sites²⁸ • Can be difficult to aggregate data and make systematic comparisons • Inaccuracies and biases inherent in respondents telling researchers what they think they want to hear. • Success in data-collection dependent to some degree on the personal attributes and skills of the researcher. Risk that the researcher influences (partially predetermines) the responses through pre-selection of questions and through being actively engaged in the interviewing process.

Source: Builds on Chambers (2003b); Carvalho and White (1997).

That each method has its own strengths, which can counteract weaknesses in other methods, also gives an indication of how the methods could be integrated. Qualitative methods may counter or go some way towards overcoming the 'pitfalls of simplistic econometric approaches in uncovering causality' (Kanbur, 2003: 11). At a more prosaic level, knowledge garnered during qualitative research can be used to test semantics and concepts, ensuring clarity in questions and terms used in the survey. Data collection and

²⁸ There is a philosophically grounded set of arguments for generalisability and reliability in qualitative methods and techniques for handling qualitative data and hence methods of arriving at the 'truth claims' in them. At the heart of the rigour of qualitative methods is 'self-critical,

analysis from household surveys may be improved using qualitative methods based on the ability to probe in the field and uncover surprises – issues may arise during interviews that may not have been taken into account at the survey design stage. Analysis can then be reoriented based on observations in the field. This certainly was the case in the thesis research: the discovery of local nicknaming conventions during qualitative fieldwork in round 1, which had not been taken into account in the design of the name generator instruments for capturing the social networks data, meant that the subsequent survey in round 2 could capture nicknames thus allowing accurate coding of alters and ego in the social networks dataset. This is discussed later in the chapter.

Information provided by qualitative methods may also generate more ‘accurate’ hypotheses based on realities that can subsequently be tested using a quantitative approach, thus shaping the survey’s conceptual framework. Two sources of data from the same population facilitate cross-checking and replication. In particular contextual (qualitative) information may help to explain counterintuitive, inconclusive or puzzling survey findings. If the two sets of results differ markedly this could indicate methodological or data quality problems in one or the other.

The decision to combine quantitative and qualitative methods of data-collection and analysis in the DPhil research was driven by pragmatism, whereby the choice of method is based on technical rather than epistemological criteria (Patton, 1990). In addition, social network data are typically gathered using a variety of techniques ranging from survey questionnaires to interviews and observations. Methods were chosen that were appropriate for the given research questions, making sure there was a match between the kinds of realities revealed by the techniques chosen and what the research actually wanted to examine. It can be difficult to explore social processes using survey instruments where there is little scope for asking ‘why?’ or ‘how?’. As one of the strengths of qualitative research is its ability to explore social processes and the context in which they occur, it makes sense for a study of social processes or networks to utilise tools devised precisely for that purpose.

epistemological awareness...with reflection on how context, process, agency and interaction influence what is presented and what is perceived by the researcher’ (Chambers, 2003b).

The qualitative enquiry was intended to complement the quantitative and networks data generated by the survey, so that analysis would not rely solely on statistical correlations to support or reject hypotheses thrown up by the social science literature about the role that social networks play in livelihoods of poor people in rural areas of sub-Saharan Africa. Neither qualitative nor quantitative methods alone would be able to capture the complexities of social networks. Social network analysis is still in its infancy within economics and as such there is no systematic model to follow. Coupled with the dearth of secondary and historical data on social networks this meant it would be difficult to design a survey that would capture everything. Instruments needed to be more open-ended, involving respondents in describing and analysing their own actions and lives. The contextual knowledge provided by the more qualitative approaches allowed access to the meanings and motivations for particular courses of action. Furthermore, much social network and labour exchange activity is nuanced and not easy to capture with a questionnaire. Qualitative methods helped to reveal people's exchange behaviour in social networks in a way that a quantitative survey never could, or would find very difficult, and was better at revealing multiple and conflicting perspectives and realities. Other advantages were flexibility and greater insights with regards to causal processes. The decision to conduct a household survey was not questioned as there was a clear need to be able to make comparisons and generalisations, as well as conduct econometric and statistical analyses. Estimating exchange behaviour through social networks would not be possible with qualitative methods alone.

While the research does not go so far as to embrace an 'heuristic' approach to data collection and analysis, a phenomenological or *experiential* understanding could be said to underlie important decisions governing research methods, both in highlighting the importance of using methods that will capture people's experience of the world and acknowledging that there is no separate (objective) reality for people. Instead, focus resides not only on what people experience, but also *how* they experience the world, and methods are chosen to reveal commonalities in these human experiences. The phenomenological perspective can be summed up as follows: 'It is reflection which objectifies points of view or perspectives, whereas when I perceive, I belong, through my point of view, to the world as a whole, nor am I even aware of the limits of my visual field' (Merleau-Ponty, 1962: 329). However, this is not the place to elaborate on

particular philosophers, instead I want to draw out some more general implications of working within the phenomenological tradition.

The research is reliant on understanding social networks and exchange networks as they happen within the experiential situation. The experiential situation is portrayed by the ‘self-reported’ data. From the experiential examples offered I try and elucidate in greater detail the social matrices and networks that become apparent. Academic language and discussion is rooted in rationalism and the ill-fitting hubris of scientific methodology, therefore it is necessary to acknowledge that although I am writing of the experiential, I am doing it in a synthetic, analytical way. Now, this is unavoidable, but must be acknowledged, this is why Nietzsche claimed that his philosophy should be sung or danced rather than argued (*Birth of Tragedy*). Philosophers such as Nietzsche, Heidegger and Wittgenstein suggest that we must already understand to some degree before we can take on the strands of argument. The spirit of the implication of self and others in continuity with the social matrices (containing the networks) seen in the experiential situation, must in a way be grasped before any analytical discussion of said networks can make sense. Even so, it is impossible to get everything. The analytic will never contain all of the experiential.

The crucial role played by institutional and social contexts in underpinning both social networks and exchange behaviour and other aspects of economic life, suggested a mixed-methods approach to data collection using a blend of quantitative (survey) methods and more qualitative (contextual) techniques. The need for a panoply of methods is reinforced by the need for different types of data encompassing attributes, interactions, motivations and other contextual information.

The next section discusses design of the data-collection instruments chosen, related to evidence necessary to address the overarching questions of the DPhil research on the role of social networks in economic life.

3.4 Design of Data Collection Instruments

Data collection instruments were designed following the preliminary field visit. Relevant literature in economics, sociology, anthropology and history, and other

secondary sources of information on Northern Province in Zambia were consulted to facilitate contextual accuracy. Important questions considered from the outset when designing the instruments were: At what time of year would the survey take place and would it be necessary to conduct more than one round? What variables would be necessary to answer the research questions? Would data be collected and analysis conducted at the household or individual level, or both? What data would be needed at the community level?

First, in terms of timing, the seasonal nature of agriculture suggested multiple survey rounds, taking place at different points in the agricultural year. In Zambia there are two distinct climatic seasons: the ‘rainy’ season, running from November to April, and the ‘dry’ season from March to October. Fieldwork was divided into two rounds, with Round 1 scheduled to take place in October, towards the end of the dry season, and Round 2 during March and April, towards the end of the rainy season. To define either season as a ‘lean’ or ‘peak’ season depends on the criteria employed as each season will be lean in some respects but not in others, for example the availability of some types of food over others or workload in different activities. Importantly, whether or not an individual or household experiences a season as lean depends on socio-economic circumstances. This proviso notwithstanding, for many subsistence farmers in Northern Province the period from October to March is generally thought of as being lean in most traditional foods, but there is a peak in labour demand for land preparation, planting and weeding and ‘early’ harvests of green vegetables and beans. By contrast, most traditional food crops are plentiful from April to October, but farm labour demands are less overall; although most harvesting takes place during this period, it also incorporates a ‘slack’ period during June and July in terms of labour requirements in agriculture. This is when people have time to pursue leisure activities, for example playing football and netball, or other non-farm activities such as making reed mats.

In order to address the research questions, it was clear that data were needed that would reflect not only respondents’ attributes, characteristics such as age, education, livelihoods activities and land cultivated, but also interactions between people, both ‘social’ and ‘productive’ exchange behaviour. Early economic studies related to social networks focused on social processes as forms of capital and included variables to capture endowments of social capital constructed from survey data that were not

collected specifically for this purpose, so there was a reliance on proxies. More recent surveys have attempted to pinpoint people's social networks more directly through specially designed survey modules. However, surveys do not ask about specific network partners and tend to concentrate on social support networks in very broad terms. The World Bank/ Republic of Uganda Global Social Capital Survey (1998) focuses on membership of groups and attendance at church, as well as 'subjective well-being', 'violence and crime', 'political engagement'. There are some questions on informal interactions with others, but because they are very general and hypothetical and do not ask with whom do people actually interact and for what purpose, and actually gather names, the data are not suitable for a network analysis.

While affiliations are an important element of social networks, and instruments such as the Social Capital Assessment Tool of The World Bank are useful for gathering information on groups and community characteristics, 'social capital' does not necessarily manifest itself in social institutions such as attending church, which is the implication of the design of the early World Bank surveys (see for example Narayan and Pritchett, 1997, on Tanzania, with its focus on quality of associational life). One contribution of the DPhil research is the design of a survey module to gather data on social networks, recognising that social capital does not necessarily manifest itself in church or group affiliation, or in community-wide measures such as levels of crime and violence, but rests in transactions and interactions between people, their own personal networks and contacts and the way they operate and interact.

An important part of the household survey was a module on labour exchange behaviour, designed for the research project alongside which the DPhil fieldwork was conducted. The labour exchange survey module also provided important social network information used in the thesis research. Previous surveys, including the Zambian LCMS, have tended not to contain questions that would adequately capture a true picture of labour market activity in rural areas²⁹, and this has contributed to the opinion of many economists that labour markets in rural areas of developing countries and particularly sub-Saharan Africa, are thin or non-existent. In fact labour exchange is

²⁹ See Termine P (2001) 'What can be learned from the LCMS Instrument on Rural Labour Markets? A Case Study of Zambia. Mimeo, Institute of Development Studies, Brighton, UK. Background paper to Labour markets project.

widespread and dynamic among small-scale, subsistence farmers in rural sub-Saharan Africa and empirical evidence supports this (See Moore and Vaughan, 1994; den Ouden, 1995; Vaughan, 1998; McNetting, 1993; Binns, 1992). This dynamism can be captured partly with a survey so long as questions cover such features as reciprocal labour and activities that people may not consider to be ‘working for others’ because they do not receive a ‘wage’ or other explicit payment for it. Indeed, as there tends to be a stigma attached to carrying out farm wage work for neighbours, direct questions on wage employment are least likely to elicit accurate responses. Other important features necessary for understanding labour exchange behaviour include different types of remuneration for various tasks, why certain modes of payment are preferred over others by different people at different times and how labour exchange behaves during particular times of the year or season and in relation to the life cycle/ life stage of workers. Much labour exchange takes place through network links.

It made sense for some data to be collected at the household level, such as number of household members, assets held, type of abode and number of migrant workers. Other data were collected at the individual level – labour exchange, land cultivated, crops cultivated and other livelihoods activities – features not necessarily the same across the household. It was decided that social networks data would also be collected at the individual level as this, by its nature, is highly personalised information. The number of network partners, their locations and positions in the community, types of social relation and exchange between network partners can vary widely depending on gender or age, for example, although individuals can invariably benefit from the network relations of their family and other household members. Especially prominent are gendered differences in labour exchange through social networks, and other intra-household differences that could only be captured using individual-level data from multiple household members.

The following sections describe instrument design, keeping the quantitative-qualitative distinction for ease of exposition.

3.4.1 Household survey instrument design

The household survey was designed following the preliminary field visit. A World Bank manual on designing household survey questionnaires (Grosch and Glewwe, 1999) was a useful starting point, as well as household surveys conducted recently in Zambia including the LCMS (1998), agricultural census and post-harvest surveys (CSO Lusaka, Zambia; 1996-1997 and 1997-1998). Survey design also drew on previous ‘social capital’ surveys in both developing and developed countries and lessons learned from social network surveys³⁰³¹. This preliminary field visit and the knowledge and experience of in-country research partners provided valuable contextual information to ensure relevance of survey content. An initial draft of the questionnaire was piloted in a Bemba village close to Lusaka in June 2002. This incorporated training of the survey team who would be involved in the fieldwork proper. There was discussion and feedback involving the entire team following the pilot and the survey was consequently revised based on lessons learned and recommendations of the survey team. An enumerators’ instruction manual was produced to accompany the questionnaire.

Tracking devices were built into the survey in the form of detailed community, household and individual identifier codes, in order to locate respondents during the planned second round survey. This was especially important given the fluctuating nature of household form – a tendency for frequent divorce and remarriage, along with migration, means that the household identified in October could be rather different from that in March. A community-level questionnaire was also designed to record information at the village-level. This included ‘governance’ details such as the name of the Chief presiding over the area and the provision of services to the village, for example the nearest health facilities and schools.

³⁰ See National Statistics Matrix of Social Capital Surveys 2001 for a list of UK surveys. The World Bank Social Capital website (www.worldbank.org) contains a number of surveys, including a social capital assessment tool.

³¹ For social support questions the research drew on the Office for National Statistics UK ‘Guide to Assessing People’s Perceptions of their neighbourhood and Community Involvement’. The International Social Surveys Programme (ISSP, 2001: www.issp.org) also included useful questions on social support and on membership of groups and associations. The World Bank Republic of Uganda Global Social Capital Survey and World Bank Social Capital Assessment Tool contain useful questions on social assistance and access to productive resources.

The household survey questionnaire comprised well-defined questions, structured around ten modules for Round 1, whereas the questionnaire for Round 2 was somewhat shorter as some questions did not need to be asked twice. Extra questions were also incorporated into the Round 2 survey, informed by data gathered during Round 1: for example, it became clear from interviews that tithes are paid by church members every month and are a regular household outgoing, however this had been overlooked in the design of the Round 1 questionnaire. Most questions were pre-coded although some, especially those relating to network partners, were open-ended. The recall period for many questions/ modules was the previous six months to the day of the survey, and the surveys were conducted six months apart, to capture seasonality.

The questionnaire was divided into three parts to allow ease of enumeration given the length of the questionnaire, providing a natural stopping point for enumerators if they felt the interview had taken long enough and there would be a benefit from continuing another time. Part A covered modules 1 to 4 and was designed to be administered once per household, with questions relating to the household as a whole. Household was defined as a group of people who live in the same compound, provide things for each other and often share meals.³² Parts B and C covered modules 5 to 10 and were addressed to individuals. Table 3.4 lists the questionnaire modules for each round of the survey.

Table 3.4 Household Survey Modules

	Round 1 Survey	Round 2 Survey	Social Networks Survey (Round 1)
Part A	1. Household Roster	1. Household Roster	1. Household Roster
	2. Current Migrants	2. Current migrants	2. Current migrants
	3. Household Assets	3. Household assets	3. Household assets and land cultivated
	4. Resettlement Module (for Lufubu only)
Part B	5. Labour	5. Labour	5. 'Mini' Labour Module
	6. Former Migrants	No Module 6	6. Former Migrants
Part C	7. Social Networks	No module 7	7. Social Networks
	8. Livelihoods	8. Livelihoods	
	9. Land	9. Land	
	10. Means of Exchange	10. Means of Exchange	

Source: Labour Markets Survey, Northern Province, Zambia 2002-2003

³² Definition of Household, after consultation and discussion with Zambian research partners, is based on that used in the 1998 LCMS as a group of persons who normally cook, eat and live together.

Where possible, the household head and one other household member of the opposite sex were interviewed. Most often this was the male household head and his wife or wives. Where the household head was female the second respondent was often the eldest son or other male household member. The survey was long, which constrained the number of open-ended questions that could be included but the research team did follow up interesting and relevant issues with respondents (see below) and enumerators were encouraged to note interesting extra information that could be important for interpreting and contextualising the survey data later, at the analysis stage.

Social Networks Module.

Network interviews were conducted as part of the household survey during Round 1 of the fieldwork. A person-based data collection strategy was employed within the household survey questionnaire and from this a set of indicators, each referring to different aspects of social networks, was constructed. The social networks module of the household survey is in Appendix A. These data are complemented by qualitative information gathered using a range of techniques drawing on emerging work in social capital measurement, based on Lin's Network Theory of social capital (see van der Gaag and Snijders, 2003a; van der Gaag and Snijders, 2005; Lin and Dumin, 1986). Different measurement instruments were combined within the household survey questionnaire and comprised³³:

- **Name Generator:** maps an ego-centred social network and is widely used in social network analysis to generate a respondent's single core network. The data were collected in a specially designed social networks module. Respondents were asked to name all those people, neighbours, kin, co-workers, friends, with whom they interact from any sphere of their life from outside their own household, and their names recorded in a response matrix. Further questions were posed concerning attributes of the network partner (sex, age), the nature of the relationship between network partner and respondent and multiple role relationships. These data form the 'interaction' network of the respondent.

³³ For a more detailed explanation of the different types of instruments see Van der Gaag and Snijders (2003b).

- ***Resource Generator***: captures access to a fixed list of specific social and economic resources relating to several spheres of life. For example, in the labour markets module, the respondent was asked to supply names of employees or employers in terms of those they worked for or employed to work for them from outside their own household, before going on to answer questions regarding the nature of the employment arrangement. In the migration module respondents also gave names of household members who were working away from the household as migrant workers. Similarly, respondents were asked in a separate subsection of the social networks module to name all those they turned to under various scenarios for assistance and/or information. This section allowed sources of support to encompass institutions and organisation, as well as individuals, with room for multiple responses. Again, follow-up questions covered age, sex, relationship to respondent and multiple role relationships.
- ***“Position Generator” (Prestige networks)***: respondents were asked whether they have kin or friends in an ‘influential’ position, the name and position of the ‘influential’ person, focusing on the presence of social resources rather than the relationships per se, although the data do provide additional information on network extensiveness and diversity.

Once the survey module had been completed respondents were either shown the lists, or they were read out by the enumerator, and asked if anybody was missing. Research in methodology for social capital measurement suggests that the three types of generator are each appropriate for generating different types of network information (Van der Gaag & Snijders, 2003b). Potential drawbacks of each of the three generators are discussed in Section 3.7.

Recording names and other details about people’s network partners potentially raises ethics issues. However, the data were completely anonymised following the construction of the social networks datasets prior to analysis.

Distinct sections of the module capture various dimensions of social relations:

- i) Non-household member interaction;

- ii) Prominent people and network elites;
- iii) Involvement in formal and informal groups and clubs;
- iv) Exclusion;
- v) Community mobilisation;
- vi) Social support: health, credit, farming, work, form-filling, marketing;
- vii) Information networks: livelihoods and health.

The data allow construction of variables such as access to social support networks, participation in social institutions, as well as measures to capture network content in terms of attributes of people the respondent interacts with, including frequency of interaction and relationship to respondent. As part of the sample survey, whereby we have information on the networks of the sampled respondents/ households and not of the entire network, the research takes an egocentric approach. This allows us to build up a picture of respondents in relation to their social roles, rather than uncovering the macrostructure of the entire network of the three communities. The ego-centred approach is discussed further later in this chapter and in Chapter 4.

Labour networks questions are embedded in the labour module of the questionnaire. Respondents were asked to name all employers and employees with basic attribute information. In the social support section of the social networks module a question on who the respondent would go to if they, or a member of their household, needed work will allow a degree of cross-checking of network information. Migration modules ('Current Migrants' and 'Former Migrants') contain questions about contacts. The instruments taken together generated data on the following networks that combine to make the overall social network of the respondent:

- i) Migration
- ii) Labour (as both employer and employee of individuals and groups)
- iii) Daily interaction
- iv) Social support
- v) Information
- vi) Affiliation to community-level groups and societies
- vii) Trading and access to markets
- viii) Prestige networks: friends and kin.

There was no constraint on the number of people respondents could nominate in response to different questions. This format was designed with minimising measurement error in mind, elaborated later in Section 3.7 on methodological issues and also in the econometric analysis in Chapter 6.

3.4.2 Qualitative instrument design

The qualitative research was designed to uncover complexities of social networks and labour exchange that could not be captured by the survey instrument alone, to allow a micro-level, in-depth study. Information was generated through case study methods adapting anthropological and participatory techniques. Although the fieldwork design, data collection and subsequent analyses draw on anthropological methods, the study makes no pretence as to ethnographic comprehensiveness. Neither is it ‘participatory’ in its truest sense in terms of involving in research design and empowering the study communities. Participatory behaviours were upheld regarding principles, approaches and methods and we had the consent and active participation of the communities concerned. Researchers were transparent about the purpose of the research and what people could and could not expect, and on the whole meetings were well-attended with participants appearing to enjoy the process. Nonetheless this was a research exercise and as such instruments were largely ‘extractive’.

The qualitative element of the fieldwork employed a basket of techniques, both visual and conversational, to reveal information along the main themes of the research: social networks, labour exchange, access to economic opportunities, market access and migration. They were designed drawing on ideas from Chambers (2002) and Rietbergen-McCracken and Narayan (1998), consultation with Zambian colleagues, attention to in-country media and other sources of information about the area under study, including information gleaned from the study sites during the fieldwork. Approaches were combined and tailored to what was considered to be both socially ‘appropriate’, while being able to supply information relevant to the research themes.

During Round 1 qualitative and contextual methods concentrated on semi-structured interviews with key informants and interviews with formal community-level groups.

Prior to the fieldwork a short ‘group’ questionnaire and interview guides were designed to guide discussion around perceptions of village life and social networks as well as community and group-level data to provide a basis for collecting affiliation networks data. This stage of the qualitative work also allowed for issues to come to light that might not have been considered previously, allowing them to be pursued further later by incorporating them into the design of instruments to be used in Round 2, thus demonstrating the iterative nature of fieldwork method design.

Round 2 of the qualitative fieldwork was designed around follow-up interviews with groups identified during Round 1, in addition to semi-structured interviews with informal groups in the communities using interview guides, seasonal activity and labour calendars produced using PRA methods and personal testimonies and oral life-histories of individuals with an aim to gather case study material for between ten and fifteen individuals in each of the study sites. Table 3.5 lists the qualitative methods used. This case-study-style approach was chosen for its ability to help causal mechanisms to unfold allowing plausible causal inferences to be made, and to elicit information on constraints affecting farmer behaviour. As with the focus groups discussions, interviews were semi-structured and designed to stimulate discussion on a range of issues related to social networks, livelihoods and poverty.

Table 3.5 Qualitative methods used

Individuals:
<ul style="list-style-type: none"> • Semi-structured interviews • Personal Testimony: Oral Life History combined with Daily Activity Chart • Seasonal activity and employment calendar (Labour calendar) – male and female • Personal testimony: Village histories • Structured direct observation • Transect walks
Groups:
<ul style="list-style-type: none"> • Personal Testimony: Oral Life History combined with Daily Activity Chart • Seasonal activity and employment calendar (Labour calendar) – male and female • Institutional Mapping • Determinants exercise, including preference ranking • Gender Division of Labour Exercise • Structured direct observation • Semi-structured interviews (with and without ‘group questionnaire’ interview guide)
Community:
<ul style="list-style-type: none"> • Village and natural resource mapping • Structured direct observation • Personal testimony: Village histories • Semi-structured interviews with ‘community questionnaire’ interview guide

Source: Fieldwork, Northern Province, Zambia, 2002-2003.

3.5 Data Collection

Data collection methods were ‘sequential’, although not strictly so, rather than a simultaneous mixing of methods, with the aim of creating an ongoing dialogue between the qualitative and quantitative components of the research. First, data collection instruments were designed using an iterative process. The process of collecting data in the field was also dynamic and evolving, using an eclectic set of tools and methods. The qualitative fieldwork and the household survey were conducted side-by-side. The demands of supervising the household survey and conducting qualitative field research simultaneously were great and there was a tendency for the household survey to take precedence over other activities, given the structured nature of the exercise and the concentration of survey team members on this element of the research plan, so Round 2 was designed to allow extra time for qualitative work. Again the household survey and qualitative fieldwork were conducted alongside each other, followed by a further period devoted solely to qualitative work, with a smaller team of three researchers remaining post-survey to continue contextual data collection.

During fieldwork there was continuous evaluation and discussion of the data collection processes, including challenges faced by enumerators when posing certain questions in the survey. This meant that many potential problems, such as difficulties with questions or sections of the survey, could be addressed and remedied on the ground. One particular problem arose with the section of the questionnaire on fish farming activities – questions related to household fish consumption proved to be difficult as people had no idea how much fish they had consumed and no way of estimating. Enumerators were told not to waste time probing too much on these questions, especially as by the time this section was reached late in the interview respondent-fatigue was evident. Data collection overall was systematic and care was taken to preserve the integrity of the data, in particular through i) avoidance of ‘leading’ questions (although the scripted questionnaire meant that this was less of an issue here than with the less structured qualitative research); ii) exact recording of responses; iii) systematically observing and recording the information gathering processes; and iv) a ‘working’ awareness of power relationships between different groups and individuals in the population and between the researchers and the researched.

3.5.1 The Household Survey

The research team comprised six enumerators, four male and two female, five of whom had been trained during the pilot survey. The survey was co-supervised by a civil servant from the Zambian Ministry of Labour who had participated in the preliminary field visit. Together with researchers from a Northern Province regional agricultural research station, his presence helped to smooth access to the communities. The survey communities on the whole treated the survey seriously, however the presence of a government official also meant that despite efforts made on entry to the survey sites to convey clearly the exact purpose and meaning of the research many of the respondents harboured concerns about possible ‘hidden’ implications of the work: would there be repercussions and would they be adversely affected? Uncertainty about the research appeared to play on existing insecurities; for Ngulula and Lufubu Resettlement Scheme this reflected concerns about security of land tenure. Consequently in the beginning there was some reticence on the part of respondents to open up to the research team, although on the surface they appeared to be cooperating. This mistrust abated as familiarity with the research team increased, and appeared to be absent from Round 2.

The same six enumerators took part in Round 2 of the survey with one additional female enumerator who had participated in the pilot survey. This freed up one of the original team to work full-time as a researcher and translator in the qualitative fieldwork. The Round 2 survey was co-supervised by an experienced field researcher and supervisor who worked part-time in the Ministry of Planning. The continuation of largely the same staff in the second round of the survey contributed to a loss of outsider status by the team and a more relaxed interaction with respondents as relationships were strengthened between researchers and community members.

Sampling

Full enumeration (a census) of Lufubu Resettlement Scheme was attempted in terms of household coverage, partly due to the relatively small size overall of the community (65 households in total) and partly because one important focus of the rural labour markets project was resettlement schemes. For Ngulula and Kabila villages a sample of 40 households each were drawn. The sample frames were the household registers provided

by the respective headman of each village, and households to be included in the survey were selected at random using random numbers lists generated in Excel. Before the sample was drawn the household lists were updated with the help of village committee members to add new households and remove those who were deceased or had moved away permanently since the registers were compiled. The lists were reasonably up-to-date thanks to a recent national census of population and housing (carried out by Zambia's Central Statistical Office in 2000). If a sampled household was not available for interview after three attempted visits by enumerators, the household was replaced using a random number table. Table 3.6 summarises the sampling frame used. Out of sixty-five households in Lufubu Resettlement Scheme, just sixty-two were enumerated. One person was missing due to sickness, one was absent and the remaining household was an oversight – the village committee omitted it from the register because the plot had only recently been occupied and the plot-holder had not yet moved permanently to the scheme.

Table 3.6 Sampling frame

Study Site	Total No. Households	Estimated Population
Kabila	200	666
Ngulula	100	600
Lufubu RS	65	480

Source: Labour Markets Survey, Northern Province, Zambia, 2002-2003

Survey Data Collection

The household survey was based on a pre-coded questionnaire in English translated into Bemba during the interviews. Enumerators were all native Bemba speakers and training included vocabulary to ensure uniformity of translation. Using native Bemba speakers facilitated access and acceptance. Households were allocated to enumerators in Round 1 and interviewed by the same enumerator in both rounds. Supervision involved attending interviews. During Round 2 respondents appeared to be considerably busier than the previous visit, especially those engaged in early beans harvests.

Each evening the team discussed progress made during the day at a debriefing session, tabling any issues and problems arising during the survey. They also provided feedback of a contextual nature that helped to inform the qualitative exercises. In addition, the completed questionnaires were checked at the end of each day as part of supervision

duties and corrections and clarifications made where possible. Sometimes this would require repeat visits to households. Data were also re-coded in the field when required.

Sample Attrition

Relocating households the second time round was relatively straightforward using the four-tier identifier code and household registers. In addition, given the relatively short interval between visits of just six months, enumerators remembered clearly households they had been assigned and the location of households by village section had been noted at the first visit. Table 3.7 details the number of people and households interviewed during the survey. Sample attrition for both households and individuals was 5.5%. Household coverage (percentage of total households) was 37, 20.5 and 91 per cent in Kabila, Ngulula and Lufubu respectively.

Table 3.7 Sample Attrition

Survey Site	Round 1 households enumerated	Round 1 individuals enumerated	Round 2 households enumerated	Round 2 individuals enumerated	% Total households: final sample
Kabila	41	71	37	64	20.5
Ngulula	43	55	41	53	37
Lufubu	61	105	59	102	91
TOTAL	145	231	137	219	..

Source: Labour Markets Survey, Northern Province, Zambia, 20003-2003.

3.5.2 Generating qualitative data

The qualitative data collection was conducted with the assistance of two Zambian research assistants, one male and one female. Interviews took an informal approach and refreshments were provided as a small ‘payment’ for participants’ time and effort. The emphasis in Round 1 was on focus group discussions and semi-structured interviews with formal and informal village groups and with key informants. Institutional background information was gathered from interviews with people in relevant local government departments such as the agricultural extension service, Ministry of Agriculture and the Government of Zambia Resettlement Office. In Round 2, activities were expanded to include in depth interviews with individuals in the study sites, encompassing oral life histories, personal testimonies and group exercises using Participatory Rural Appraisal techniques. These included: seasonal calendars; village mapping; institutional mapping; daily time use diagrams. During the exercises participants generated their own analysis of key elements of their livelihoods and

market behaviour. Interviews were conducted in Bemba and translated into English *in situ*, although I built up a working knowledge of the key vocabulary for crops grown and different types of agricultural and non-farm activity over time. This helped to identify key words in people's conversation and allowed follow-up questions to unravel points made and a more in-depth investigation. Some interviews were taped with the permission of the interviewee; direct quotations used in the thesis come from the taped interviews.

From the outset, an awareness of the following issues was instilled in the research team and steps taken to guard against inherent biases, inconsistencies and 'misleading' information that might arise as a consequence:

- i) *village/ hierarchical structure*: The views of 'those at the top' are likely to be easier to access and more apparent with a tendency for prominent community members such as village committees to dominate discussions, giving the impression that they are voicing generally held views, whereas others may be less willing, even in private, to give views that oppose those of village leaders (Casley and Lury, 1982). This can extend to a tendency for village elites and dominant community members to put forward their friends and relations for oral life histories and other research exercises, increasing self-selection and other biases.
- ii) *'insider/ outsider scale'*: There could be incidents of respondents telling researchers what they think they want to hear. A related issue is a tendency for people to put on their 'best face' to outsiders and giving only an 'official' view (Casley and Lury, 1982). Other issues: responses given based on a perception of what the respondent thinks they may be able to get from the research team or out of the research project; having to accept as 'authentic' what people tell us – the retelling of their understanding of own situations, derived from their 'stock of knowledge', balanced by an awareness of respondents engaging in 'leg-pulling' because they are not taking the research seriously.

Efforts were made to establish the points of view of less vocal groups in the community, such as poorer people, those who were more marginalised and those living further from village 'centre', and care taken not to avoid reliance on particular informants. This included transect walks, interviewing people encountered along the way. Enumerators also reported interesting interactions with respondents during the course of the survey

and identified individuals and households that they felt it would be important and valuable to talk to. Sampling is dealt with in more detail in the following sections.

Key to the success of the qualitative research was establishing relationships with people that would permit intrusion, making people feel comfortable with the research and research processes and building up trust. It was initially difficult to transcend the precedence set by the household survey and the presence of ministry people, which gave an 'official' air to the research and meant that people were initially reticent to give out highly personal information on their social networks and their labour activities. In some cases there was also a perception that the survey team were expressing favouritism towards those chosen for the household survey or for the oral life histories. The team endeavoured to make the process as transparent as possible and encouraged the people to ask as many questions as they liked, allowing for a two-way interaction between the researchers and the respondents.

Individual interviews

Individuals were interviewed for their life history as a starting point for a discussion about the main issues of the research, by researchers working in pairs. Interviews were based on open-ended questions. Conversation was steered around perceptions of village life to open up the arena for new issues to come to light as well as following research themes identified in the interview guide. The interviews began by asking people to tell the researchers about themselves, their lives, where they were born, what their parents did, where they went to school, taking their own story right up to the present-day. On the whole it followed a free-conversational format but researchers ensured issues concerning social networks and labour exchange were covered. Interviews took a considerable length of time to complete and it was not always possible to cover all issues in one sitting. On some occasions interviews were conducted over a number of sessions. If respondents did not raise issues contained in the interview guide then the interviewer would instigate a discussion, but only as a last resort.

Where couples were interviewed, the discussions with the husband and wife were conducted separately by two different interviewers with the male interviewed by the male researcher. Respondents drew a daily activity chart with guidance from the

interviewer, to show graphically how they spend their day giving activity patterns for the rainy season and the dry season, followed by a short discussion of what they had drawn. The discussion also included semi-structured interviewing to probe further on themes emerging from the life history. Table 3.8 provides a tally of numbers of interviews in each study site. All interviews were discussed in depth by the research team at the end of each day and sometimes participants would be revisited to clarify points made or to follow up on interesting leads.

Table 3.8 Qualitative Fieldwork – Total Number of Interviews

Study Site	Oral Life History	Personal Testimony	Semi-structured interviews	TOTAL
Kabila	17	9	17	33
Ngulula	20	15	10	45
Lufubu	15	8	14	37

The sample of individuals and groups for the qualitative fieldwork was taken from the bounded set of inhabitants of the three villages selected for the household survey. From this set, respondents were purposively sampled using the following categories to ensure sufficient diversity:

- Migrant worker
- Employer
- Employee
- Female-headed household
- Old person
- Young man
- Young woman
- Intra-household – 3 couples from different age groups (plus daily activity charts).
- Prominent person
- Marginalised person
- Village group leader
- Village group member
- Non-member of groups
- Section heads or village committee member

All people sampled fell into two or more categories so there was considerable overlap. Respondents were found using snowball sampling, through consulting the questionnaires from round 1, taking transect walks, ‘accidental’ sampling and common sense. At least one male and one female were selected in each category where possible, covering a wide age-range to ensure that the data portrayed the perspectives of a broad cross-section of the community, including representation of different socio-economic

levels, community and employment status and geographic/spatial locations. The latter was especially important in Kabila because the village was particularly large in that there were two hundred households covering a wide geographical area. The village comprised six different ‘sections’, some far from the village centre and relatively more isolated, and initial interviews suggested that labour market behaviour was markedly different in the further-flung sections compared to the village centre. Individual cases are not sampling units in a statistical sense. Table 3.9 details respondents (sex, age group and status) participating in oral life histories, conducted during the second round of fieldwork.

Table 3.9 People selected for Oral Life Histories – Round 2 of Fieldwork

Study Site	Oral life Histories	Category
Kabila	<ul style="list-style-type: none"> • 1 male, 1 female • 1 male • 1 female • 1 male, 1 female • 1 male • 1 male, 1 female • 1 male • 1 female • 1 female • 1 male • 1 female 	<ul style="list-style-type: none"> • Youngish/ middle-aged couple, male b1962 • Prominent person, Branch Chairman • FHH with migrant worker • Couple: employers and produce surplus • Section leader, elder • Older couple keeping orphans • Young man, newly married, living with in-laws • Married woman, well-connected • Young married woman • Young man, married, young children • Older woman, not member of any groups
Ngulula	<ul style="list-style-type: none"> • 1 male, 1 female • 1 female • 1 male • 1 male, 1 female • 1 male • 1 female • 1 female • 1 male • 1 male • 1 male • 1 male • 1 male, 1 female • 1 male • 1 female 	<ul style="list-style-type: none"> • Older couple, retired here, well-educated • FHH • Young Man • Middle-aged couple, prominent group member • Member of village committee • FHH, elderly, marginalised • FHH, ‘destitute’ • Young male, migrant, single, about to get married • Young male with big plans, grade 12 education, single • Older man, not a member of any groups • Married Man b 1970, 7 children • Young couple, no children yet • Headman • Married woman, originally from Kasama
Lufubu Resettle-ment Scheme	<ul style="list-style-type: none"> • 1 male, 1 female • 1 female • 1 female • 1 male • 1 female • 1 male • 1 female • 1 male, 1 female • 1 male • 1 male • 1 female 	<ul style="list-style-type: none"> • Middle-aged couple • FHH, ‘average’ • FHH, not destitute, well-connected, group leader, employer • Relatively rich, polygamist • Wife of above, managing land for migrant brother • Relatively rich, polygamist, group leader • Married woman, middle-aged, settled <1 year, poor and struggling • Older couple, migrant workers • 1975 settler on largest plot • New settler, employs people, married • Middle-aged woman, married, well-connected, group member/ leader

Group discussions and exercises

Qualitative materials from group discussions and Participatory Rural Appraisal (PRA)-style exercises were used to inform the research. Techniques used include:

- i) Village and natural resource mapping
- ii) Semi-structured interviews
- iii) Personal testimony: Village histories
- iv) Institutional Mapping
- v) Seasonal Activity and Labour Calendar
- vi) Gender Division of Labour Exercise
- vii) Determinants exercise, including preference ranking

The exercises were sequenced to schedule the relatively ‘simple’ ones first such as drawing the village map, and later covering more complicated issues, either because they involved a certain amount of analysis on the part of the participants or because the issues were of a more sensitive nature, such as gender division of labour. This was to ensure that the community/ participants were more comfortable with the tools and concepts employed to increase their effectiveness, and a greater degree of rapport between the facilitators and the community allowing for more open, frank discussions. Participatory mapping exercises provided visuals to facilitate discussion and in all cases the participatory exercise was accompanied by short semi-structured interviewing with the groups concerned as well as discussions of the exercise at hand. The combination of techniques allowed for cross-checking of issues and findings (‘triangulation’).

The determinants exercise was designed to explore and identify the relative importance placed on social networks for livelihoods improvement by respondents. Mixed focus groups were convened to discuss economic opportunities and successful livelihoods, focusing on what it means to ‘get on’ in the context of their lives. Once it was established what these concepts meant to the participants, the focus was switched to what was needed to access a particular economic opportunity identified by the group. The participants listed everything they thought they would need in order to ‘get on’ or to benefit from that opportunity. Respondents then voted for the ten most important. Each was written on a separate piece of paper and participants assessed the different items, according to which they considered to be the most important for getting on,

giving reasons for their choice. The items were presented in every possible combination of two until all possible combinations had been considered. A matrix was drawn up of responses, which were ranked and listed according to the number of times they were preferred over the other items (pair-wise ranking). The group discussed the results.

This exercise was challenging to both participants and facilitators. Potential biases arose during the exercise, related to differences and gaps in understanding that needed to be addressed as the discussion and exercise proceeded. These included: the tendency to choose a new item as opposed to the ‘preferred’ item when an element appeared in two consecutive pairs; confusion and difficulties in coming to consensus when respondents could see no apparent direct link between two items drawn in a pair for comparison; a propensity to relate the elements in a pair to each other rather than relating each of them to the ‘outcome’ i.e. better livelihoods in farming.

3.6 Data Entry and Data Cleaning

Data entry for both the household survey and the qualitative fieldwork took place in Zambia – either in the RuralNet offices in Lusaka in consultation with the survey enumerators, or while in the field (in the case of the qualitative data).

3.6.1 Survey Data Entry

For Round 1 of the survey, data entry was carried out by three data entry assistants at the RuralNet offices in Lusaka using SPSS. The data template and coding scheme had been designed in advance alongside questionnaire design. The data entry assistants had not been involved with the pilot survey data entry but were trained by the person who had entered the data from the pilot. Because of the length and complexity of the questionnaire there were some problems with inaccuracies in data entered from Round 1 that necessitated a return trip to the RuralNet offices in Lusaka to consult the hard copies of the questionnaire and re-enter much of the data. Lessons-learned from this experience meant that for Round 2 those involved with collecting data in the field also carried out data entry once the team had returned to Lusaka. This proved to be more

successful in terms of quality of the dataset because those who had collected the data were more accurate owing to a deeper understanding of how the questionnaire worked.

3.6.2 Qualitative Data Entry

Detailed, verbatim notes were taken during all of the interviews and PRA exercises conducted for the qualitative strand of the research, and all materials such as maps and calendars produced by participants were kept by the research team. All interview notes and transcripts were typed up directly following the interviews, on the same day where possible, but much of the time the team was staying in the villages themselves without access to electricity so this was not always possible. The recording and transcription of interviews was an extremely time-consuming task, but care was taken to process everything in-country to allow for discussion and clarification with research assistants.

3.6.3 Social Networks Data Entry and Dataset Construction

The network data from each of the survey modules was aggregated into a ‘Tiewise’ dataset containing characteristics of network partners and ties.³⁴ The respondent and each of their network partners have only one tie but this tie may have multiple role relationships. The Tiewise dataset is a hierarchical dataset; records for each network partner or ‘alter’ are nested within a respondent record. Network measures are derived from the Tiewise into a ‘Netwise’ dataset (constructed by various aggregations of the data). This presents the data at the level of the individual respondent (as an ego-centred network dataset), with measures on the respondents’ alters aggregated to provide the social networks variables. When identical network partners were mentioned by a respondent in response to different questions, they were coded systematically so that for each respondent the network partners have their own unique identifier code, allowing for an analysis of multiplex and multifunctional relationships. This was carried out manually – converting names (often given in multiple forms including nicknames) into unique identifier codes - and proved to be complicated and time intensive.

³⁴ See Müller C, B Wellman and A Marin (1999). “How to use SPSS to study ego-centred networks”. *Bulletin des Méthodologies Sociologiques* 64: 63-76.

3.6.4 Data Cleaning

Household survey data were cleaned several times following both rounds of the survey, checking for consistency in data, missing responses, and ‘do not know’ responses. A sample of ten per cent of the questionnaires was taken and data entry for these questionnaires was checked for accuracy. Cleaning qualitative data was a question of cross-checking notes between researchers to ensure accuracy of recording information and comparing and contrasting perceptions of the same accounts or events.

3.7. Methodological Concerns

This section elaborates potential and actual methodological pitfalls relevant to the field methods used in the DPhil research, considering first sampling errors, then non-sampling and measurement errors.

Sampling errors

In taking a case study approach to the research, there was in effect a two-stage sample design for the household survey. As discussed above, survey sites were purposively selected for their relative market access. In terms of selecting households within each site, for Kabila and Ngulula random samples was selected using an up-to-date sampling frame – the most recent (2000) national population census list of all households in each village. Resource constraints led to an approach of interviewing a chosen minimum number of households in each site – 40 per village. Each household had equal probability of being selected from its respective village list, by using a random number table. Lufubu Resettlement Scheme had 66 households in total, so in this location the survey attempted complete coverage and managed to interview 62 out of the 66 households. The remaining households in Lufubu were absent or unavailable due to sickness at the time of the first survey but were visited during round 2 and interviewed during qualitative fieldwork. Relative representativeness of each household to its share of population was ensured by weighting households in the total sample according to its proportion in the subsample, to take account of differential selection probabilities by community. The survey team was fortunate to experience complete cooperation and as

the team lived in the survey sites during the fieldwork, revisits were possible for clarifications or if respondents were not at home for the original visit.

One replacement was made during round 1 of the fieldwork in Kabila village using a replacement rule. However, when the survey team revisited the household in round 2, it could not be found. It transpired that someone who had already been interviewed in round 1, had been interviewed again as this replacement ‘household’ in round two, with a different enumerator to their original interview, giving false names and a false household roster. The deception only came to light when the qualitative team returned ten days later with the questionnaires, the original household having absented itself during the survey period in round 2. Both ‘households’ were omitted from the dataset. As households were recompensed for their time in participating in the survey (a payment in-kind of a small bag of salt) there was a clear incentive to attempt to beat the system in this way, and had the research been based on a one-off visit to villages rather than split into two rounds to capture seasonality, this incident would have slipped through the net. On consideration, it is fortunate there were not more incidents of this kind, but the level of familiarity and knowledge built up between the survey team and people living in the study sites meant they could be kept to a minimum.

One of the strengths of the survey data collected in the research is the depth of information available for each household, although to achieve this sample size was sacrificed. Nevertheless, the sample is still large enough to permit robust estimation for each population subgroup (study site). Because more than one person was interviewed in each household (one adult male and one adult female, where relevant), at the individual level the sample is larger still. It was decided, however, to analyse the data at household level. At the outset, while the study was being designed, it was believed that different information, for example on crop cultivation, would be given by males and females in households based on the prior that men and women in households cultivate different crops in different fields. This was not the case in practice, and respondents reported crop cultivation, consumption, sales and losses at the household level resulting in duplicate information.

Clustered survey design usually results in lowered precision of sampling estimates in terms of representativeness of population as a whole. Households within the same

cluster, in this case village, will be more similar to each other in terms of behaviour and characteristics compared with households in other clusters. This tends to be more pronounced in rural areas where households with shared geography also share agro-ecological conditions, face the same prices and tend to belong to the same ethnic groups (see Deaton, 1997). Samples drawn in this way thus risk lowered variance. The research is based on just 3 clusters (villages), and the nature of rural lives of smallholder farmers in regions such as Northern Province suggests relatively little variability between households within each cluster. However, one of the motivations for the research is to examine how factors such as social networks, which are highly personal and do tend to vary from individual to individual, make a difference to households who ostensibly face very similar conditions and who appear to exhibit very similar characteristics, so this is not considered to be a problem here, and on the contrary can be viewed as a desirable property of the sample. This is discussed further in Chapter 5. Further, neighbourhood effects are captured using a dummy variable for location. Given the purposive nature of sample site selection, this variable also captures differential market access.

Non-sampling errors

Both respondents and enumerators can find surveys tiring and this may lead to errors as issues can be glossed over, people may choose not to answer a particular question or enumerators may not fill in parts of the questionnaire. In terms of using quantitative methods to collect data on social processes it is important to bear in mind the trade-offs that take place between the convenience of being able to gather specific, comparable information across a larger number of respondents and potentially missing important nuances and explanations due to the inherent oversimplification of social processes in reducing the complexity of human interaction in this way. It is true that the questionnaire for the research was bulky and as a result interviewer and interviewee fatigue almost certainly set in. It may have been preferable from the point of view of both parties to spread the questionnaire over several visits to the respondent, although in some cases, geographical location restricting access to respondents would have precluded this in the time we had available. However, the routine of checking completed questionnaires at the end of each day meant that re-visits we carried out where necessary. Using a well-trained and experienced survey team also meant errors related to misunderstanding questions on the part of both enumerators and respondents

was kept to a minimum, though cannot be ruled out completely. Complementing the survey with qualitative fieldwork, including following up some survey respondents as case studies, meant that more probing of issues could take place than is usually possible with questionnaire-based household surveys. The time spent living in each survey site during the course of the fieldwork meant also that further follow-up was possible, as well as making it more likely for ‘untruths’ to be uncovered.

Conducting network interviews as part of the household survey heightened many of the problems outlined above. Given the demands of the survey even without the social networks module, enumerators were unable to probe to any great extent should the question/response demand it. This is where follow-up interviews proved to be invaluable in clarifying and expanding on people’s responses. The name generator instrument in particular, with its free-recall structure also posed further demands in data entry – which was carried out by the RuralNet team in Lusaka – with much room for error in recording names of network partners. However, the construction of the actual social networks dataset was carried out by myself, using the original questionnaires including clarifications on names and nicknames.

In relation to social networks data themselves, there are issues related to measurement validity, reliability, accuracy, and error related to the nature of these data (see: Wasserman and Faust, 1994: 56-59; Marsden, 1990). Measurement error is related to the differences between the true structure and the observed structure of the social network. This can be caused by choice of reporting period leading to possible reporting and recall errors by respondents in relation to exactly who they have interacted with over a certain time period and the details of this interaction, such as frequency, purpose etc. Such errors are inherent in self-report data such as social network data collected from people reporting their own interactions. Studies on informant accuracy examining observed interactions compared with what people actually report of their interactions have found that ‘on average, about half of what informants report is probably incorrect in some way’ (Bernard et al, 1984: 503). However, studies suggest this is only true of certain situations, for example recalling in great detail interactions at a party (see Freeman et al, 1987; Brewer, 2000). Overall, when principles of memory and cognition are taken into consideration when interpreting the studies, what people report of their interactions appears to be related to ‘long-range social structure’ rather than to particular

instances (Freeman et al, 1987. See also: Marin, 2004), and therefore closer to the ‘true’ structure of the network thus minimising measurement error. Measurement error was also minimised in the thesis research in the choice of a free recall rather than fixed choice design in the bulk of the social network data collection. Avoiding restricting the nomination process by asking people for the names of all the people they interact with rather than asking, for example, who are the three people they interact with most, minimises measurement error of the network itself and in measurement of network properties.

A question often posed of qualitative in comparison with quantitative research is whether or not the findings are representative, if not of the population then certainly of the group of people under study, rather than purely anecdotal? Putting aside Lipton’s (1992) assertion, of both economics (quantitative) and anthropological (qualitative) research, that until a researcher has studied a village they are not going to know whether or not sample is representative, also noting that “the idiosyncratic person...can well be key to the whole puzzle” (Lipton, 1992: 1543), it is important to bear in mind that the objective of qualitative work is not to be ‘representative’ in a statistical sense of being generalisable to a whole population, or in this case the population of Northern Province smallholder farmers. Rather, the qualitative work strives to provide in-depth information about *how* and *why*, while the household survey focuses on *what* and *how many*.

Issues of *validity* and *reliability* need also to be considered, especially in relation to the social networks data given its less ‘tangible’ nature compared with other attributes such as crops grown or level of education. *Validity* (or internal validity) refers to whether or not the data/ researchers actually measure the concepts they intend to measure. External validity refers to the ‘extent to which theoretical constructs and postulates generated or tested are applicable across groups’ (see LeCompte and Goetz, 1982: 43-53). Validity is linked to questions arising due to conflicting findings from different methods, although it is important to remember that allowing for this is one of the main motivations behind taking a combined methods approach to empirical research and triangulating between methods, and can be a desirable outcome of the research and an interesting finding in itself. However, one needs a relative degree of confidence regarding the internal coherence of a dataset and this is in part measured by the extent to which a number of

people agree about what they see and hear when they are all observing the same phenomenon in the same setting. In social networks analysis there is very little research on validity of some concepts, for example, measures of friendship.

In asking whether or not a measure of a variable or a concept is *reliable*, this means do repeated measurements give the same estimate of the variable? Key approaches to assessing the reliability of social network data include: test-retest comparison; comparing different question formats and reciprocity of sociometric choices (see Tracy et al 1990). However, in practice it can be problematic to use the test-retest approach as one cannot expect social phenomena to remain static over time, and in this respect research can be difficult to replicate (thus raising questions about external reliability). For qualitative data in general, and including social networks data, validity and reliability depend to a great extent on the 'skill, sensitivity and integrity of the researcher' (Patton, 1990). Essential elements of this include 'systematic and vigorous observation and skilful interviewing' so that findings are useful and credible, and this accords with the view in qualitative inquiry of the researcher as instrument (Patton, 1990).

To assess the degree of validity and reliability of the data collected in the research one can consider coverage and selection criteria. In terms of coverage, the household survey approached full enumeration in Lufubu (91 per cent of households), with lowest coverage in Ngulula at just over 20 per cent, and more than one-third of households enumerated (37 per cent) in Lufubu. The qualitative fieldwork broadened the reach of the research in including members of the population who had not been selected for the household survey. There was some overlap in participants of the survey and the qualitative fieldwork in these locations, though many households who had not been selected for the survey viewed the focus group discussion and group exercises as an opportunity for them to 'have their say'. Some households and individual household members were selected for revisits for follow-up discussions. Gender balance of participants of oral life histories/ individual interviews was good: of 44 interviews during round two, 23 were male and 21 were female, and care was taken to ensure gender balance in the composition of mixed focus group discussions. One exception to this was focus group discussions with the more prominent employers in each location – these tended to be male-dominated. Purposive sampling for in-depth interviews and

focus group discussions meant a broad range of age groups and ‘positions’ in society were covered.

The integration of methods in the research raised many practical issues, not least tensions caused by underestimating time needed for both the survey and qualitative components. The resource demands (including budget needed) of carrying out in-depth research in this manner, including the level of detail required in the household survey, constrained sample sizes, although in relation to population size coverage was good. Demands of the quantitative and qualitative elements of the fieldwork were often competing and in most cases the household survey took precedence. As this became apparent in Round 1 of the research, I was able to balance this out in the subsequent Round 2 of the research to allow more time post-household survey for a smaller team of researchers (myself and two translators) to remain in the study sites to carry out further research.

Challenges notwithstanding, taking a combined methods approach has helped to provide a more rounded picture of social and economic realities in the study sites, allowing the strengths of each approach to overcome the limitations of the other. This proved to be vital given the nature of the research in attempting to uncover the social-embeddedness of economic behaviour and relations.

.....

The next chapter uses these data to begin to analyse the local social environments and networks of small-holder farmers in the three study sites.

Chapter Four

Social Networks in Three Zambian Villages

This chapter analyses the social networks of survey respondents in smallholder farming households in the three study sites. Social networks variables are constructed using social network analysis techniques. The data are concerned with exchanges between the respondent (ego) and their network partners (alters), the ties between ego and alter, and attributes of alters. Information about the nature and extent of relationships provides insights into the social context, going beyond simpler measures such as group membership or network size. This enables us to consider the richness of people's connections and analyse the ways in which different types of networks may be important for accessing different resources for different people.

Section 4.1 presents a visualisation of the social networks of three interconnected respondents from one of the survey sites. This is followed by definitions of the social networks variables constructed from the dataset to capture different dimensions of social networks: network size; prestige networks; and network diversity (Section 4.2). Correlations between the variables are presented in Section 4.3. Associations between the different social networks variables are set out in Section 4.4 to demonstrate that different characteristics are associated with different network properties. This analysis is extended in Section 4.5 using multivariate analysis of the social networks variables in relation to respondents' individual and household-level characteristics. Section 4.6 concludes.

4.1 Egocentric Social Networks

The economic literature has for a long time examined the role of networks in economic life, more recently couched in terms of social capital, or social network capital. However, the overwhelming tendency has been to use group membership, and characteristics of group membership, and/or social network size to capture social network effects. Table 4.1 shows social capital and social network measures used in a selection of prominent econometric studies linking social capital and social network

capital to outcomes, drawing mainly on those in rural contexts. These studies are discussed in more detail in the literature review in Chapter 2. As the summary table demonstrates, these studies tend to capture social networks/capital using group membership and characteristics of group membership such as length of time in groups, and measures of network size to proxy for social networks/capital. The dearth of studies of social networks in relation to smallholder farmer commercialisation is reflected here. The most relevant studies to the thesis research focus variously on risk-sharing and insurance networks, technology diffusion, household welfare, agricultural traders, and not directly on smallholder farming outcomes. Barr's (2002) work on the functional diversity and spill-over effects of social capital, based on data on entrepreneurial networks in Ghanaian manufacturing, is included here as a contrast to the tendency to focus in rural studies on group affiliation and related group membership characteristics, and network size, as proxies for social networks.

Table 4.1 Social Networks/ Capital Measures Used in a Selection of Econometric Studies

Author(s)	Study	Network/ Social Capital Measures Used
Conley and Udry (2010)	Learning about a New Technology: Pineapple in Ghana	Farmer's information networks: who they know and talk to about farming
Goldstein et al (2005)	Inclusion and Exclusion in Mutual Insurance Networks in Southern Ghana	Probability of knowing any person in the community; years respondent/family have lived in the village; number of fostering episodes; number of organisations respondent belongs to
Fafchamps and Lund (2003)	Risk-sharing networks in rural Philippines	Dummy variables for networks of friends, relatives etc. Network characteristics captured by: Number of network members; Number who own rice fields; Number with craft skills.
Fafchamps and Minten (2002)	Returns to Social Network Capital Among Traders	Number of close relatives in agricultural trade; the number of non-family traders known by respondents; and the number of friends and family members who are able to support the business financially in difficult times
Barr (2002)	Functional Diversity and Spillover Effects of Social Capital (study of Ghanaian manufacturing)	Mean number of contacts; Mean diversity of contacts; Mean number of liaisons with an average contact per year; Mean number of contacts of entrepreneurs co-networkers; Percentage of contacts who would assist in a crisis
Maluccio, et al (2000)	Social capital and household welfare in South Africa	Index of individual membership in groups to proxy for social capital: three components: <i>density</i> - the number of group memberships per household; <i>performance</i> - the average reported performance of the most important groups in the household; and <i>participation</i> - the average reported frequency of meeting attendance for the most important groups in the household
Narayan and Pritchett (1999)	Cents and Sociability: Household Income and Social Capital in Rural Tanzania	Social capital indices based on membership in groups, characteristics of the groups, household values and attitudes. Trust variables - degree of trust in strangers, kin, village chairmen, district officials, central government

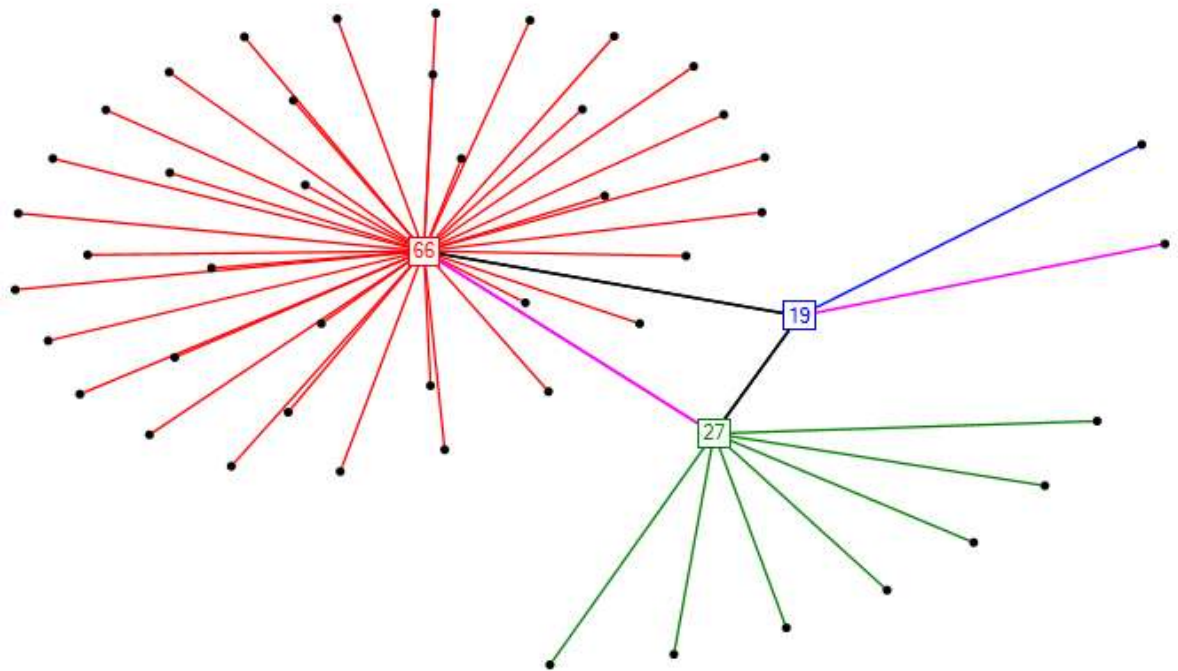
The research is interested in social networks and network relationships as the basis for access to resources contained within the network, and ultimately what this means for smallholder farmer outcomes. Drawing on anthropological and sociological studies of social institutions in rural Zambia, discussed in Chapter 1, the research recognises that different types of networks may be important for accessing different resources for different people. To capture the channels through which social networks may enable access to resources, the social networks survey module described in Chapter 3 collected egocentric network data related to exchanges between the respondent (ego) and their network partners (alters), the ties between ego and alter, and attributes of alters. This allows the construction of social networks measures going beyond size and affiliation variables generally used in studies, described above.

Egocentric network approaches are appropriate when hypotheses are posed at the individual level. An added advantage of an egocentric over a full-network approach is that it is less problematic statistically and econometrically. Standard statistical procedures including hypothesis and significance tests apply because the independence of observations assumptions are not violated. Analysis of egocentric data first requires summaries of network composition and of features of the individual ties to generate a set of variables that are attributes of the respondent. These social networks measures can be used as independent variables to predict outcomes. Measures of network properties include network size and network heterogeneity, as well as variables capturing features of individual ties including: frequency, duration, multiplexity and intensity. These are defined and discussed in more detail in Section 4.2 below. In the research we are concerned with only first-order, direct ties between respondents and their network partners ('alters'), and not ties between alters. Ego and each alter has only one tie, but this tie may have multiple role relationships.

First, consider what the social networks of the survey respondents look like. Figure 4.1 shows a network diagram or *sociogram* of the egocentric network of three of the respondents participating in the research.

Figure 4.1: Egocentric Networks of Three Interconnected Egos

Egocentric networks of three interconnected egos



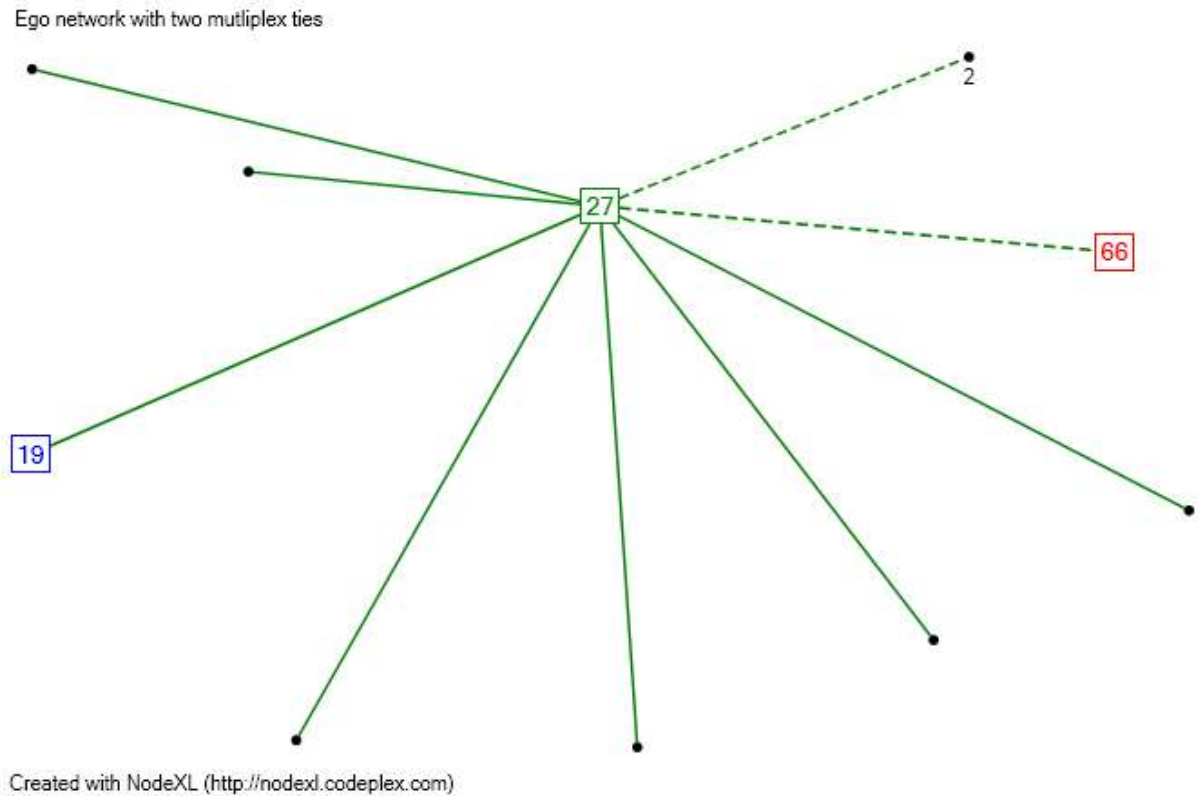
Created with NodeXL (<http://nodexl.codeplex.com>)

Source: Household Survey, Northern Province Zambia, September 2002.

Each respondent, or *ego*, is represented by a square with a number in the centre. Each alter is a black dot and the tie between ego and alter is given by a line (also called edges). Respondent 19 (blue) has a small reported social network, just 4 people. Respondent 27 (green) has 9 reported network partners and respondent 66 (red) has 39 reported network partners. Taking respondent 19 as the focal individual, for simplification, this person's employment networks can be seen by the black lines in the graph, both respondent 66 and respondent 27 provide employment to respondent 19. Respondent 19's social support networks with kin network partners are given by the Fuchsia line. One of respondent 19's network partners is kin providing social support. The links between the other two egos (27 and 66) could similarly be coded to denote a particular type of network relationship. For example, the connection between respondent 66 and respondent 27 could also be given by a Fuchsia line. Respondent 27 is the mother-in-law of respondent 66 and has stated a social support relationship. The graph can be extended to include a visualisation of multiple network relationships between egos and alters. For example, Figure 4.2 shows respondent 27's multiplex ties in terms of resources accessed through the network. This respondent has multiplex ties

with just two alters: Alter 2 and Respondent (alter) 66. In this case alter 2 is an employee and also provides information. Alter 66 provides social support and information. Multiplex ties are given by the broken line.

Figure 4.2 Egocentric Network with Two Multiplex Ties



The next section defines the social networks measures constructed from the dataset.

4.2 Social Networks Measures

The social networks survey instrument comprised three main components: the name generator, the position generator and the resource generator (see van der Gaag and Snijders, 2003a; 2003b; Lin and Dumin, 1986). The dataset was constructed as a tie-wise dataset, in which each case is an alter and the dataset contains information on alter attributes as well as the relationships between alter and ego. This allows network properties to be aggregated to the respondent and household levels for use in the econometric analysis. The tie-wise dataset was thus subsequently aggregated and structured into a net-wise dataset, where each case is the respondent or ego (described in

Chapter 3. See also Müller et al, 1999). The analysis in this chapter focuses on individual level social networks, with a sample size $N=211$. Analysis in subsequent chapters relating social networks to outcomes uses the household as the unit of analysis because the outcome variables in Chapters 5 and 6 are household level variables. Constructed social network variables capture three different dimensions of social networks: network size; prestige networks; and network diversity. The following subsections discuss each in turn.

4.2.1 Network Size

Network size can give an indication of the likelihood of a network containing an alter who possesses a resource that ego needs (Bourdieu, 1986; Burt, 1992). Network size measures can be constructed from the data. These include the total number of alters (network size) or the number of alters who fulfil specific roles such as providing employment, for example the size of employment network. Group affiliation contributes to network size, and the analysis includes a variable for membership of community-level groups and societies as a more ‘typical’ social capital variable. Table 4.2 summarises network size variables derived from the Tiewise dataset described in Chapter 3.

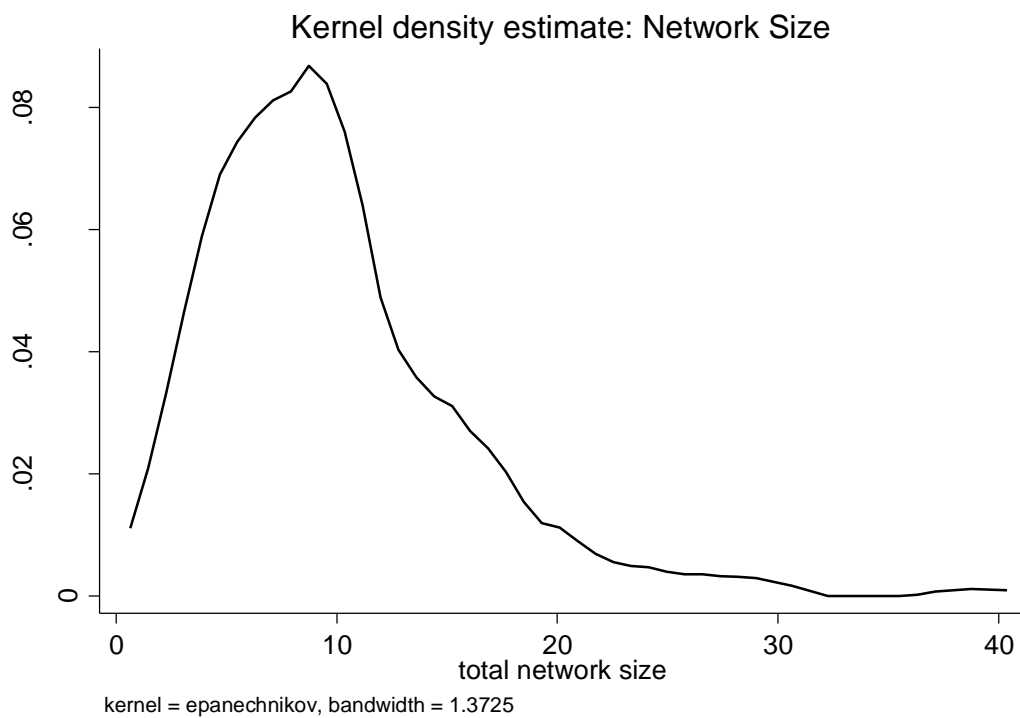
Table 4.2 Network Size Variables

Variable name	Definition	Mean	St. dev	Min	Max
size	Total number of network partners cited across all role relationships (interaction, employer, employee, migration, marketing, social support, information)	9.81	5.67	2	39
affil	Group Affiliation variable. Membership of village-level and other formal or informal groups	2.04	1.50	0	7

Size is the most basic measure of a social network, defined simply as the total number of network partners in an interpersonal environment. According to social networks studies, network size gives a “reasonably direct measure of social integration” (see, among others, Marsden, 1990; 1987; Marsden and Campbell, 1984). However, it is important to note that the number of alters might be misleading as a measure of ego’s “connectedness” in terms of accessibility of resources because it assumes each tie has the same sort of content, and does not tell us anything about embedded social resources. Using the household survey data, size of network is captured by i) the total number of alters (*size*); and ii) group affiliation (*affil*). Here, size is measured as the number of

different alters given across the following environments: interpersonal (interaction), labour (both employer and employee), migration, marketing, social support and information. Group affiliation is a variable capturing respondents' total number of memberships of formal and informal village-level groups and societies including women's groups, sports clubs, farmer groups and cooperatives. Figure 4.3 shows the kernel density estimate for network size. Larger networks (larger than 10 alters) are relatively uncommon in the sample.

Figure 4.3: Kernel Density Estimate – Network Size



4.2.2 Prestige and Influence

Social capital studies suggest that positive social capital results from accessing network members with high prestige, relative to ego and other alters in the network. This is because alters in 'powerful' positions are more influential and more likely to provide access to valuable and/or novel resources (financial, cultural and political) in their network (Lin, 2001). This relates to 'bridging' or 'linking' networks: links with individuals who are dissimilar to 'ego' and provide access to a more diverse flow of resources (See Woolcock, 2003). This goes beyond the number of ties or associations in

a network to ascribe also some kind of potential ‘value’ to alters and to the ties between ego and alter.

Unlike traditional position generator instruments, which tend to present respondents with lists of occupations from which to identify where they have alters, the research did not define prestige networks in relation to occupations. Instead, the social networks survey asked respondents whether or not they had influential network partners, phrased as “Do you have a friend/kin who: owns a business/ heads a committee/ community leader/ other influential position?”, along with their position, relationship to respondent, and their location (see Appendix A). This allowed the survey instrument to capture context-specific notions of prestige rather than imposing these from the outside. Prestige networks are captured as a dummy variable signifying the presence or absence of influential kin or friends, and a categorical variable for location: whether they live in ego’s own community (village) or elsewhere (Table 4.3).

Table 4.3 Prestige Network Variables

Variable Name	Definition	Mean	St.Dev	Min	Max
kininf	Whether or not respondent’s network comprises influential kin or friends	0.57	0.49	0	1
kinloc	Location of influential kin/ friend	0.75	0.76	0	2

4.2.3 Diversity in Composition and Function

The greater the diversity of the network (in terms of social resources accessed through the network, or range of network relationships), the more chance there is that someone in the network has something that the ‘ego’ needs. High diversity implies integration into several spheres of society or social circles/ contexts and this is considered to be advantageous for mobilising resources and for instrumental actions like gathering information (Lin et al, 1981; Campbell et al., 1986; Kadushin, 1982). Once again, by merely counting the number of different alters in a network we miss information about the relationship between ego and alter, how useful an alter might be for resource access or for providing support and information. Network diversity is another way of considering network heterogeneity – or how mixed networks are. Variables that capture network heterogeneity include: *diversity of composition* (sex and age heterogeneity of alters, kin versus non-kin network partners), and *functional diversity* (alters perform a variety of functions in the life of ‘ego’). Functional diversity can be captured in various

ways, for example as the number of ‘networks’ an alter operates in, in relation to the respondent, or the number of resources ego can access through the network, or the number of ties that are multiplex, that is, have more than one function or role relationship in ego’s network.

A. Diversity of Composition

Heterogeneity or dispersion measures of network partner attribute variables are the most direct measures of the diversity of alters that ego can contact within his/her personal environment (see Campbell et al, 1986). Respondents’ networks may be highly heterogeneous in some respects yet homogeneous in others, for example in the way they vary by age, sex, kin. For interval measures of network partner characteristics (for example, age) the standard deviation of alter attributes is used. For nominal or categorical variables, such as ethnic group, sex etc, analysts use an Index of Qualitative Variation (IQV).

The Index of Qualitative Variation (IQV) is a measure of variability or dispersion that represents diversity within a distribution, measuring variation by comparing the cases (scores or observations) to one another. It is appropriate for any level of measurement once it has been categorised (see Agresti and Agresti, 1977). IQV is a standardised version of the Index of Diversity, which indicates the likelihood that two observations drawn at random from a sample are from different categories of the variable. The IQV is the ratio of observed to maximum heterogeneity, and ranges in value from zero to 1. A value of zero means there is no diversity at all. A value of 1 signifies a distribution that is maximally diverse. For our sample we have calculated the IQV of each respondent’s sex of network partners using the formula:

$$IQV = \frac{k(N^2 - \sum f^2)}{N^2(k-1)} \quad (1)$$

where:

- k = number of categories (male/female)
- N = number of cases (network size)
- $\sum f^2$ = sum of squared frequencies of each category

However, when constructed in this way a network of two members, one male and one female, would get an IQV score of 1 and be maximally diverse, whereas a network of,

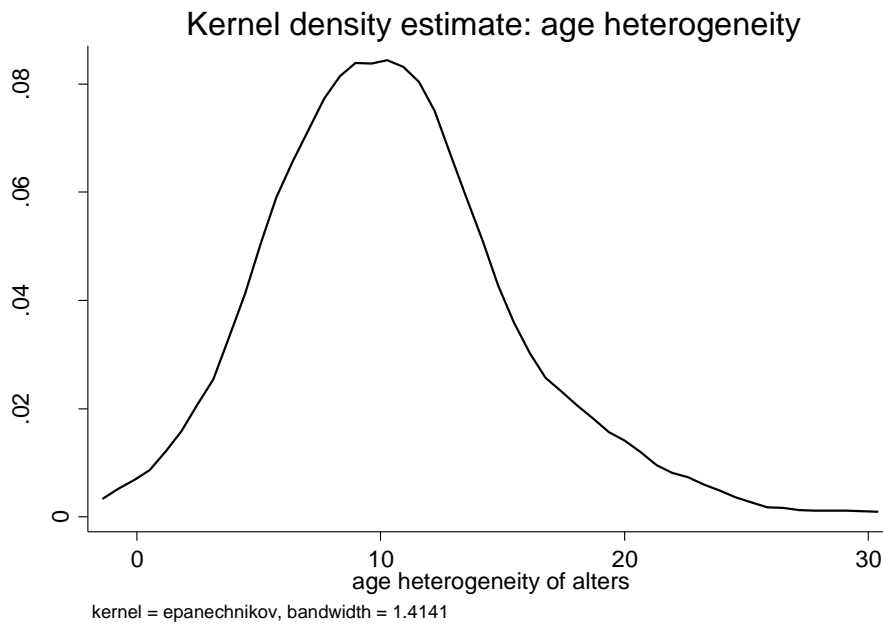
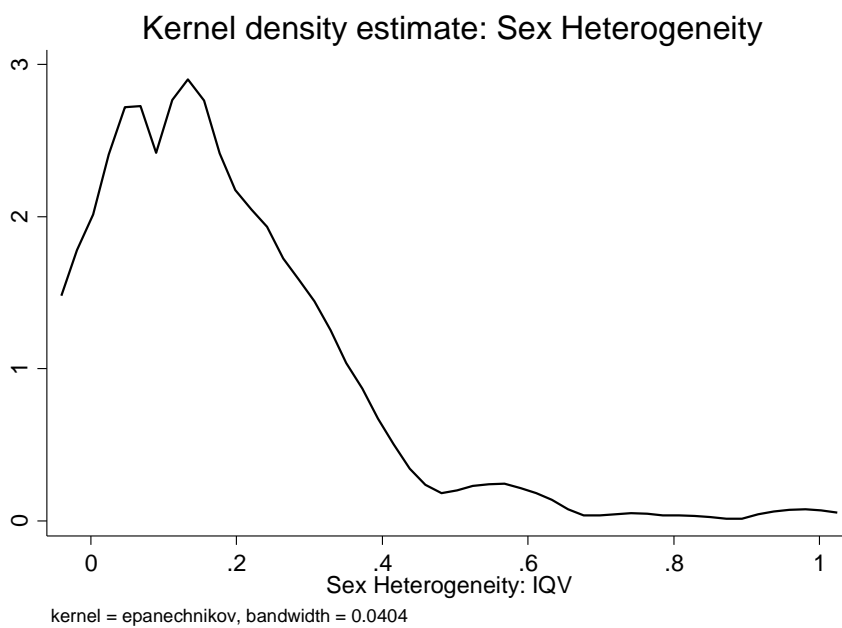
say, 20 people, 13 of whom are male and 7 female, would get an IQV score of around 0.9 and therefore appear less diverse than the smaller network. This does not make logical sense so the measure is standardised by weighting it by the size of the respondent's network as a proportion of the largest network size in a respondent's location.

Heterogeneity gives some idea of the range of someone's social network, that is, potential access to social resources. Size can also be used as a measure of range. In general, heterogeneity increases with network range. In the research, network heterogeneity is captured by diversity in age and sex of network partners, and kin composition of networks (Table 4.4).

Table 4.4 Composition Diversity Variables

Variable Name	Definition	Mean	St.Dev	Min	Max
agehet	Age heterogeneity of network partners (standard deviation of alter ages, across each network)	10.5	4.91	0	28.99
sexhet	Sex heterogeneity of alters (Simpson's Index of Qualitative Variation in sex of alters, across each network)	0.18	0.17	0	0.98
propkin	Proportion of network members who are kin	0.33	0.26	0	1

Age Heterogeneity is measured as the standard deviation of the ages of each respondent's network partners. The kernel density estimate in Figure 4.4 shows that it is centred around the lower two-thirds of scores, suggesting very mixed networks in terms of age are relatively less common in the sample. Figure 4.5 gives the distribution of the sex heterogeneity variable. This appears to be skewed towards lower values of the variable suggesting more homogeneity in networks related to sex of network partner. This fits with the local context, Zambian society, especially in rural areas, is generally conservative, 'gender' relations are very traditional in that there is a tendency towards less interaction between members of the opposite sex who are not kin/ related. The twin peak in the sex heterogeneity plot suggests that there may be two underlying distributions here, possibly one for men and one for women.

Figure 4.4: Kernel Density Estimate – Age Heterogeneity of Network**Figure 4.5: Kernel Density Estimate – Sex Heterogeneity of Network**

Relational heterogeneity is given by the ratio of kin to non-kin alters in ego's network (see Marsden, 1987). A high proportion of kin in one's network implies a network that may be less diverse in terms of background of alters. It also implies that ties between ego and kin alters are strong ties, relating to Granovetter's "Strength of Weak Ties"

argument (Granovetter, 1973; 1983). Weak ties between alters provide access to other people's clusters of closely knit friends, that are different from one's own close circle of friends. They therefore potentially act as a crucial bridge to novel information and resources 'from distant parts of the social structure', which is seen to be an advantage in competitive environments (see Burt, 1992). By contrast, alters with strong ties to each other, such as kin, tend to know the same things ego knows, and are less likely to provide such access to information and resources. The distribution of the kin composition variable is given in Figure 4.6. The distribution suggests most networks have lower proportions of kin (less than 0.5): the average proportion of kin is just one third (Table 4.4). This is on the surface surprising given the importance of kin in the context of the three rural communities, although the inclusion of a resettlement scheme may be influencing the distribution here. Breaking down the *propkin* variable by community, this indeed appears to be the case (Table 4.5). The maximum kin proportion network in Lufubu is two-thirds compared with Kabila and Ngulula, where respondents reported entire networks consisting of kin; the average kin proportion in Lufubu is also lower.

Figure 4.6: Kernel Density Estimate – Kin Composition of Network

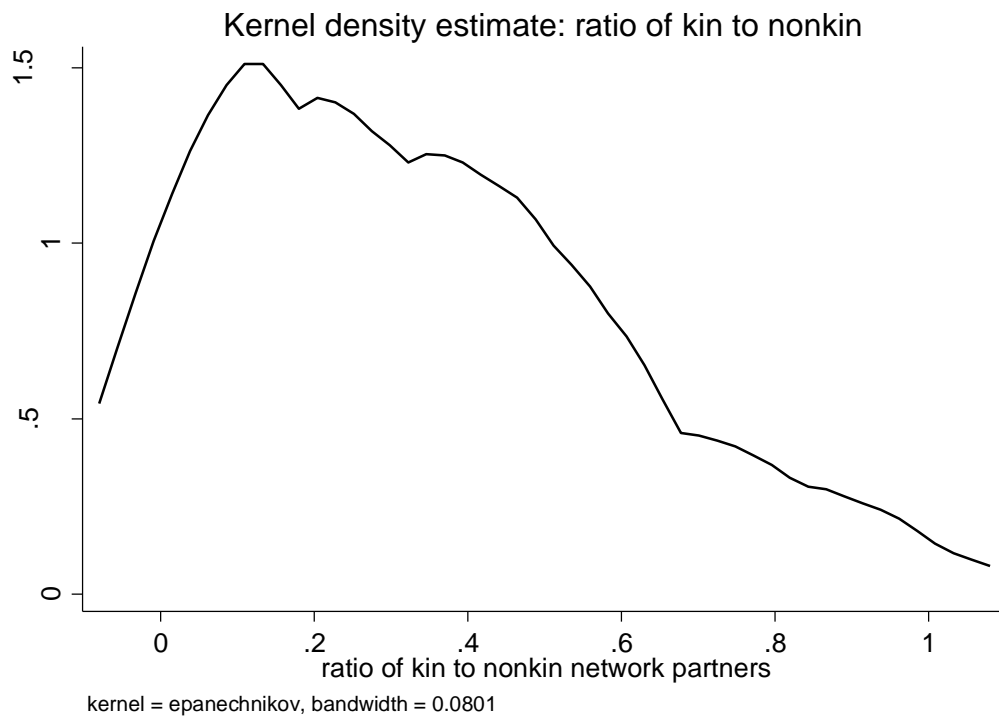


Table 4.5 Kin as proportion of Network, by Community

Propkin	N	Mean	St.Dev	Min	Max
Kabila	61	0.49	0.27	0	1
Ngulula	51	0.36	0.27	0	1
Lufubu	99	0.22	0.19	0	.67

B. Functional Diversity

Functional diversity of a network can be captured by measures on tie attributes, describing the quality of relations between the respondent and network partners. Concepts here include direction, frequency, duration, multiplexity, and intensity of relationship (Marsden, 1987; Wasserman and Faust, 1994; Scott, 1991). Many measures at the network level are inductive and often used as indicators of more than one concept:

Multiplexity: or ‘multi-stranded’ relationships. Relationships between ego and alters can operate in different contexts or with different roles, for example as employer-employee, and/or as friends who offer each other social support, and/or as a source of information (see also: Mitchell, 1969; Marsden, 1987; Skvoretz and Agneessens, 2007). Multi-stranded relationships “tend to be more intense because they are more diffuse in character” (Scott, 1991: 32). Positions in one set of relations may reinforce or contradict positions in another. Moreover, actors may be tied quite closely in one relationship network but be quite distant from one another in others.

Direction: relationships can involve a transaction or exchange, encompassing reciprocal arrangements, or they can be one-way. For example the respondent employs Person B, but B does not employ the respondent. Therefore in the respondent’s employee network the relationship is uni-directional.

Intensity: The strength of obligations involved in a relationship. This can be measured in a number of ways: the strength of commitment to obligations or the multiplexity of the relationship or whether the relationship is with kin or non-kin, or by the general level of ‘closeness’ between the respondent and network partner, for example are they friends or merely neighbours. It can also be measured as the number of relationships a respondent and network partner share.

Frequency: how often does interaction take place? (mean frequency of contact for each tie).

Duration: Is the relationship a one-off or is it *constantly* being ‘activated’? i.e. transience of relationship.

Here, diversity in the *ties* between the respondent and their network partners is measured using variables capturing *multiplexity* in the relationships between the respondent and their network partners. Multiplexity can also signify strength of relationship. It can indicate the extent to which ego is strongly or weakly tied to alters and the extent to which alters might be obligated to support him/her, and vice-versa. Social networks studies measure multiplexity in a number of ways. Measures of multiplexity based on the *ties* between ego and alter are:

- the total number of multiplex ties (*multiplex*: the number of ego-alter relationships with more than one role or function);
- a dummy variable for the existence of these multiplex ties in ego’s network (*multidum*);
- and the number of multiplex ties as a proportion of the total number of ties (*propmult*), that is, as a proportion of network size (Kapferer, 1969).

Roles or functions are defined in the study as interactions along the following networks: social interaction, labour, migration, marketing, social support, information, group affiliation. When an alter is mentioned in relation to more than one of these functions, for example as an employer and also providing social support, then the tie is considered to be multiplex. For the individual therefore the variable *multiplex* can range in value from zero, to denote ego has no tie mentioned more than once, to all ties in the network being multiplex, which for the individual will correspond to the total number of ties in their network (*network size*).

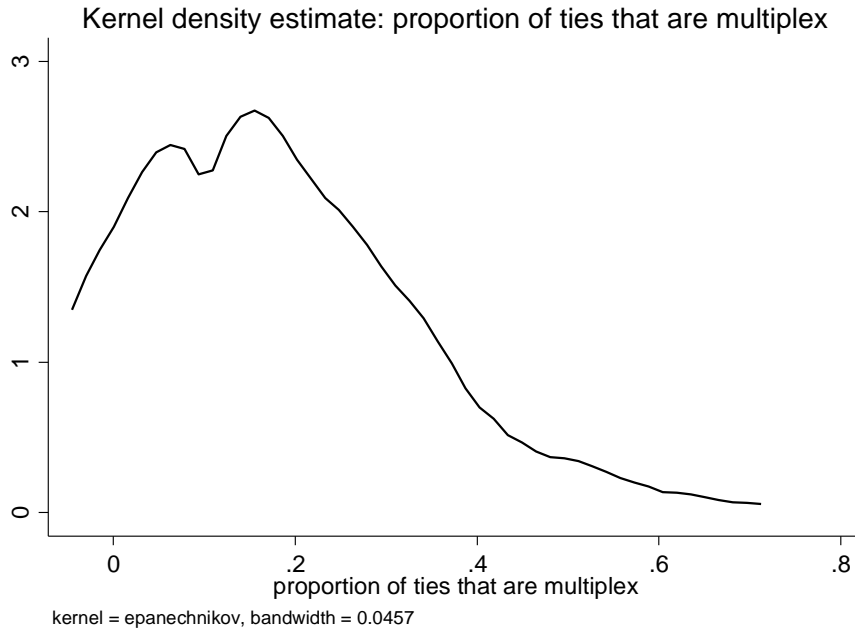
Multiplexity from ego’s perspective can be captured in two further ways. First, the mean number of roles found in ties between ego and alters. Alters can have ties to ego in each of the seven strands of instrumental network being measured, that is, the relationship has multiple contents. This can range from 1, the alter has only one role relationship with ego, to 7, the alter is mentioned in relation to all of the roles. The variable *multcont* is thus the average number of roles an ego’s alters have across the 7 networks types. A score of “zero” would denote no relationships and corresponds to complete isolation (a network size of zero). Second, multiplexity can be given by the

“number of specific relational contents...represented in a network” (Campbell et al, 1986:104). In the survey, respondents were asked about different network contents or resources focusing on specific types of social support or information in the ‘resource generator’ section of the social networks survey module, for example “who would you go to if you were ill in bed and needed help at home” (questions 7.25-7.29. See Appendix A). These network contents and labour network content from a separate labour markets survey module are combined (summed) to give a variable with potentially a maximum score of 21 and a minimum score of zero (*ressum*).

Functional diversity (multiplexity) measures constructed from the survey data based on tie characteristics are summarised in Table 4.6, and the kernel density estimate for the variable *propmult* is given in Figure 4.7. Once again, the twin peak is suggestive of two underlying distributions, possibly between commercialised and non-commercialised farmers – the former tend to score more highly on the *propmult* variable, with a higher average score.

Table 4.6 Functional Diversity Variables

Variable Name	Definition	Mean	St.Dev	Min	Max
multiplex	Total number of multiplex ties	1.67	1.43	0	7
multidum	Dummy variable for the existence of multiplex ties	0.78	0.42	0	1
propmult	Proportion of total number of ties that are multiplex (multiplex/size)	0.18	0.15	0	0.67
ressum	Sum of resources accessed across network (max=21)	12.00	4.43	0	20
multcont	Average number of ‘contents’ ie network roles per alter (max=7)	3.22	0.83	2	7

Figure 4.7: Kernel Density Estimate – Network Multiplexity (proportion of ties)

The next section sets out correlations between the social networks variables.

4.3 Correlations between different social networks measures

Correlations between the different social networks measures described above, including the variable for group affiliation, are examined to test whether the social networks variables are capturing different things. The correlation coefficients are given in Table 4.7. There is only one strong correlation, between the network size and sex heterogeneity of network partners ($\rho=0.72$). The positive sign suggests that larger networks are more mixed. There is a moderate correlation between network size and group affiliation ($\rho=0.47$) – group members have bigger networks; and between network size and two of the multiplexity variables: *ressum* (number of resources accessed through networks: $\rho=0.49$); and *multiplex* (total number of multiplex ties: $\rho=0.43$). A larger network means access to more network resources and more multiplex ties, that is more network partners performing multiple roles/functions, both relationships to be expected. The variable capturing the ratio of kin to nonkin in the network (*propkin*) has a weakly negative correlation with network size that is significant at 5 per cent. The multiplexity variables all have significant and mostly moderate positive correlations with each other, suggesting the research may wish to

select just one multiplexity variable in subsequent multivariate analyses. The *propmult* variable is the preferred multiplexity measure as this takes network size into account.

Table 4.7 Social Networks Variables: Pearson Correlation Coefficients

	size	affil	agehet	IQVst2	propkin	multcont	ressum	multiplex
size	1							
affil	0.47*	1						
agehet	-0.04	-0.08	1					
IQVst2	0.72*	0.21*	0.01	1				
propkin	-0.22*	-0.21*	0.16*	-0.003	1			
multcont	0.08	0.13	-0.07	0.08	0.05	1		
ressum	0.49*	0.41*	0.12	0.33*	0.01	0.19*	1	
multiplex	0.43*	0.27*	-0.07	0.32*	-0.02	0.52*	0.39*	1
propmult	-0.10	0.03	-0.01	-0.03	0.11	0.62*	0.19*	0.72*

* $p < 0.05$

The next section (4.4) presents bivariate analysis of social networks data and respondent attributes, before turning to multivariate analysis in Section 4.5.

4.4 Network characteristics and respondent attributes

This section analyses relationships between respondent networks and other individual attributes, reporting only those that are statistically significant, to test if social networks properties vary over the same individuals, and by implication does one draw different conclusions about how connected people are if different social networks measures are used. This section presents some simple bivariate associations before moving on to more formal testing using multivariate analysis in the next section.

Appendix B presents cross-tabulations of significant relationships between categorical versions of a selection of social networks variables. Different kinds of people do appear to have different network properties. Respondents with a male household head tend to report larger networks, as do those with a higher level of education (Table B.1 in Appendix B). Network size appears to be increasing with level of education. This may be related to the greater opportunity to expand one's network as a direct result of attending secondary school, which tend to be located in the nearest town, by providing the opportunity to mix with a (more diverse) set of people outside of one's community.

Males report larger networks than females: In the sample there are roughly equal numbers of men and women (slightly more men than women: 53.3 per cent male against 46.7 per cent female). Overall, male respondents are over-represented in the larger network categories and women in the smaller. Women make up 63.5 per cent of smallest network category and men 67.6 per cent of the >15 network partners category (Table B.2 in Appendix B). But women's networks are more likely to be mixed-sex than those of men. Most men (57.6 per cent) fall in the middle of the distribution, but are over-represented in the lower sex heterogeneity category, making up 77.6 per cent of those with low sex heterogeneity in their networks (Table B.10 in Appendix B). Women make up 79.3 per cent of those with sex heterogeneity of greater than 0.9 compared with 48.9 per cent of the sample as a whole. Men are likely to have network partners in more of the network types than women: almost 60 per cent of those involved in five, six or the maximum seven of the networks being measured are males compared with 51.1 per cent of respondents being male in the sample as a whole (Table B.11).

Employers of group labour have larger networks: Across the sample as whole, employers of group labour tend to have larger networks (70.6 per cent of those with networks of greater than 15 network partners compared with 41.6 per cent of sample). Three-quarters of those with 5 or fewer network partners are not employers of group labour compared with 58.4 per cent of sample (Table B.4 in Appendix B). Employers of group labour are also more likely to be involved in many kinds of network, and to report having influential kin or friends in their network. Those employing group labour are over-represented in the larger multiplexity category of five to seven of the instrumental networks being studied: 56.8 per cent of respondents involved in between five and seven networks employ group labour compared with 41.6 per cent of sample (Table B.11 in Appendix B).

The next section extends the analysis to consider multivariate relationships between respondent and household attributes and network characteristics to test whether the way networks are measured matters.

4.5 Determinants of Social Network Characteristics

The analysis above suggests it is important how we measure social networks: different individual and household characteristics have different relationships to social networks. This section presents the results of regression analysis taking in turn as dependent variable each social network variable: network size; group affiliation, age heterogeneity of network; sex heterogeneity of network; kin members as proportion of network; and the four multiplexity variables (*multiplex*, *propmult*, *multcont* and *ressum*). The purpose of the analysis is to see if we can identify what kinds of networks different types of people have, and whether or not there is any consistency across people in terms of their network characteristics. Summary statistics for the dependent variables and regressors are given in Table 4.8.

Table 4.8 Dependent Variables in Social Networks OLS estimations

Dependent Variables:	N	Mean	Std. Dev.	Min	Max
size	211	9.81	5.67	2	39
affil	211	2.04	1.50	0	7
agehet	203	10.50	4.91	0	28.99
IQVst2	211	0.18	0.17	0	0.98
propkin	211	0.33	0.26	0	1
multcont	211	3.22	0.83	2	7
ressum	211	11.95	4.43	0	20
multiplex	211	1.67	1.43	0	7
propmult	211	0.18	0.15	0	0.67
Regressors:					
Sex (memb_sex)	211	0.47	0.50	0	1
Age (memb_age)	211	40.17	13.74	15	80
Household Size (hh_size)	211	7.09	3.06	1	17
Sex of HH head (hh_sex)	211	0.15	0.35	0	1
Marital status (marital)	211	1.14	0.35	1	2
HH Head dummy (hhhead)	211	1.42	0.50	1	2
Education level (educ)	196	1.96	0.71	1	3
Communal labour dummy (commun)	211	0.82	0.38	0	1
Nonfarm income dummy (nonfarm)	211	0.40	0.49	0	1
No of large livestock (livelarg)	211	2.44	4.67	0	39
No of small livestock (livesmall)	211	5.76	6.25	0	45
Group labour employee (glpart2)	211	0.33	0.47	0	1
Group labour employer (glboss2)	211	0.44	0.50	0	1

The regressors are chosen to represent the main characteristics of individuals: sex, age, education level, marital status, whether or not the respondent is household head; as well some household-level demographic information: household size, sex of household head.

Variables are also included for livestock ownership as a proxy for assets. A distinction is made between large and small livestock because large livestock is more indicative of assets in terms of being a store of value, as well as potential farm inputs (draught animals; manure). Variables are also included for involvement in communal labour on community works such as road maintenance, for group labour involvement, and dummy variables for both employer and employee of group labour, as these might be expected to be closely related to social networks.

Leaving aside for now the issue of potential endogeneity, each estimation is compared to see how the coefficients change with the different dependent variables³⁵. While the data are collected at the individual level, in many cases multiple people are from the same household (just 57 households have a single respondent). A set of household dummies are included to capture fixed effects³⁶. The equality of coefficients between the equations is tested. Results of the Ordinary Least Squares regressions are given in Table 4.9.

An F- test of the equality of coefficients of variable k between regressions (m) , $\beta_{(k,m)}$, tests whether or not all coefficients in each regression are different from each other, under the null hypothesis $H_0: \beta_{(k,1)} = \beta_{(k,2)} = \dots = \beta_{(k,m)}$. The F test gives an F statistic of $F(128, 173) = 3.07$ (Prob > F = 0.0000), rejecting the null hypothesis, suggesting that the correlates of social networks characteristics do differ significantly across the different social networks measures and that different groups of individuals have different kinds of networks.

³⁵ Endogeneity may arise because of reverse causality between the social networks variables and other characteristics of respondents. For example, involvement in group labour may lead to larger networks but larger networks may also lead to more involvement in group labour, either through the information (for example, job search and screening) effects of networks, or through greater obligation towards networks partners to participate in reciprocal group labour. The research returns to this issue in detail and offers solutions in Chapter 6.

³⁶ See also the discussion of unobserved heterogeneity in Chapter 6.

Table 4.9 OLS Results: Determinants of Social Networks Characteristics

VARIABLES	Network Size		Network Diversity: Composition			Network Diversity: Function			
	(1) size	(2) affil	(3) agehet	(4) IQVst2	(5) propkin	(6) multcont	(7) ressum	(8) multiplex	(9) propmult
Ngulula	13.73 (1.01)	2.74 (0.77)	11.79 (1.05)	0.55* (1.37)	0.16 (0.26)	-2.20 (-1.14)	6.18 (0.60)	-1.91 (-0.58)	-0.52* (-1.47)
Lufubu	-9.86 (-0.86)	-3.61 (-1.20)	7.88 (0.84)	-0.14 (-0.43)	-0.15 (-0.30)	-0.63 (-0.39)	-0.54 (-0.06)	-5.02* (-1.81)	-0.25 (-0.82)
memb_sex	-1.10 (-0.10)	-0.36 (-0.12)	1.31 (0.14)	0.15*** (0.44)	-0.01 (-0.02)	-0.27 (-0.17)	-0.36 (-0.04)	-0.72* (-0.26)	-0.03 (-0.10)
memb_age	-0.10 (-0.22)	-0.02 (-0.18)	-0.09 (-0.24)	-0.00 (-0.35)	-0.01 (-0.46)	-0.01 (-0.10)	-0.14 (-0.41)	-0.02 (-0.17)	-0.00 (-0.08)
marital	-5.00 (-0.28)	-0.75 (-0.16)	3.38 (0.21)	-0.19 (-0.35)	0.13 (0.17)	-0.12 (-0.05)	-6.41** (-0.47)	-0.52 (-0.12)	-0.02 (-0.04)
hh_size	0.45 (0.24)	0.20 (0.40)	2.01* (1.27)	0.03 (0.56)	0.01 (0.11)	-0.19 (-0.72)	-0.13 (-0.09)	0.10 (0.22)	-0.00 (-0.01)
hh_sex	-10.27 (-0.59)	-3.10 (-0.68)	9.24 (0.61)	-0.26 (-0.50)	-0.08 (-0.10)	0.74 (0.30)	-5.76 (-0.44)	-0.38 (-0.09)	0.22 (0.50)
hhhead	-1.43 (-0.12)	-0.75 (-0.25)	-2.21 (-0.23)	-0.09 (-0.26)	-0.03 (-0.07)	-0.31 (-0.19)	-2.86* (-0.33)	-0.17 (-0.06)	-0.04 (-0.15)
nonfarm	1.08 (0.09)	-0.78* (-0.25)	0.01 (0.00)	0.03 (0.08)	0.09 (0.17)	-0.13 (-0.08)	0.02 (0.00)	0.36 (0.13)	0.03 (0.10)
upp prim ed	0.16 (0.01)	-0.23 (-0.08)	0.70 (0.07)	-0.03 (-0.10)	-0.15* (-0.29)	-0.33 (-0.20)	0.01 (0.00)	-0.39 (-0.14)	-0.05 (-0.17)
sec ed	0.97 (0.07)	-0.24 (-0.07)	1.76 (0.16)	0.02 (0.06)	-0.15 (-0.24)	-0.76** (-0.40)	0.19 (0.02)	-0.58 (-0.17)	-0.12* (-0.34)
commun	-1.56 (-0.10)	0.10 (0.02)	4.45* (0.33)	-0.02 (-0.04)	0.10 (0.14)	-0.02 (-0.01)	-1.91 (-0.16)	0.62 (0.16)	0.07 (0.17)
livelarg	-0.87 (-0.73)	-0.67 (-2.12)	2.98* (3.04)	0.00 (0.00)	0.01 (0.12)	-0.11 (-0.66)	-0.59 (-0.65)	-0.07 (-0.25)	0.02 (0.60)
livesmall	0.68 (0.75)	0.37 (1.55)	-1.38 (-1.84)	0.00 (0.16)	-0.01 (-0.17)	0.00 (0.01)	0.12 (0.17)	0.03 (0.12)	-0.02 (-0.77)
glpart2	-1.06 (-0.09)	0.03 (0.01)	2.47 (0.25)	-0.00 (-0.01)	-0.12 (-0.22)	-0.02 (-0.01)	1.45 (0.16)	-0.35 (-0.12)	-0.02 (-0.05)
glboss2	5.32*** (0.46)	0.82* (0.27)	-1.04 (-0.11)	0.09* (0.28)	-0.01 (-0.02)	0.33 (0.20)	3.21** (0.37)	0.25 (0.09)	-0.05 (-0.17)
Constant	18.47 (.)	4.48 (.)	-12.59 (.)	0.36 (.)	0.63 (.)	6.64** (.)	30.74** (.)	5.97 (.)	0.67 (.)
Observations	196	196	190	196	196	196	196	196	196
R-squared	0.842	0.792	0.763	0.782	0.803	0.777	0.809	0.816	0.791
Household Fixed Effects	F(120, 59) = 1.81 Prob > F = 0.0063	F(120, 59) = 1.10 Prob > F = 0.3396	F(119, 54) = 1.02 Prob > F = 0.4812	F(120, 59) = 1.45 Prob > F = 0.0576	F(120, 59) = 1.38 Prob > F = 0.0860	F(120, 59) = 1.34 Prob > F = 0.1055	F(120, 59) = 1.58 Prob > F = 0.0263	F(120, 59) = 1.70 Prob > F = 0.0122	F(120, 59) = 1.53 Prob > F = 0.0344

Normalized beta coefficients in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

$N=196$, due to the availability of education data, with the exception of the age heterogeneity regression where missing data on alter ages for five respondents reduces the sample size further to 190.

Household id dummies are included in the regressions to pick up household fixed effects: unobserved correlations within households, those characteristics that people in the same households share that might also affect their social networks properties, for example personality, popularity, status. F-tests of the hypothesis that the coefficients on the household id dummies are jointly equal to zero fail to reject the null hypothesis at the 10 per cent level for three of the regressions (*affil*, *agehet* and *multcont*) and for two at the 5 per cent level (*IQVst2* and *propkin*). This suggests that shared household characteristics may not affect respondents' network diversity of composition, group affiliation and multiplexity measured as the average number of network roles per alter, but may be important for other network characteristics, such as network size and resources accessed through networks. This could be due to unobservables not captured by the household survey, such as personality or popularity of the people the respondent shares a household with: more outgoing people may be more likely to have larger networks, and more popular or higher status people may attract more 'followers'. These may generate positive (or negative) externalities for ego, as people are likely to benefit (or otherwise) from the externalities of the social networks of those they live with. A variable for power in groups, which may proxy for status, is available in the household survey, based on responses to Question 7.16: To what extent do you participate in the group's decision-making? (Appendix A). Bivariate distributions of this variable by network size suggest that those with more decision-making power in groups tend also to have larger networks (Table B.12 in Appendix B).

From the OLS estimations it is interesting that women (*memb_sex* variable, base category male) are more likely to have smaller networks but these networks seem to be more heterogeneous, whereas men have larger networks but don't necessarily gain any additional information from the extra members as captured by the composition diversity measures for sex heterogeneity (*IQVst2*). Therefore whether or not men are considered to be better connected than women depends on the measure used.

People who are employers of non-household group labour (that is, are able to mobilise a following) appear to have larger networks (5 more alters than average), as well as network advantages over non-employers in relation to the resources they are able to access through networks (*ressum*: accessing 3 more resources through their networks than the average). This can be illustrated by the observation by participants in the

qualitative fieldwork that “the ‘famous’ always get their work done” and “those who are well known easily have their work done, which includes finding customers as well as employees”. By contrast, those employed in group labour activities have smaller networks. One reason may be that their networks for economic and for social activities overlap. However, there appears to be little difference between group labour employees and other people in terms of most network characteristics.

Owners of large livestock appear to have larger networks and more diversity in their network composition therefore potentially access to more novel information, but other than that have no additional network advantages over those who have small livestock, in terms of resources actually accessed through networks. This may be because livestock farming is still an agricultural activity, and is quite widespread, so the differences in types of people encountered and thus types of resources accessed may not be appreciably different.

In summary, the preceding analysis shows that it does matter how social networks are measured. Different findings emerge on who is more connected depending on which measure is used, and different types of connectedness will be important for different types of people. For example, the analysis suggests that while men tend to report larger networks than women, the sex composition of women’s networks is more mixed. This is important in the context of the research. There are two main modes of cultivation in the study sites: *ibala* or garden cultivation on mounds and *citemene* or slash and burn cultivation. For *citemene* in particular land preparation tasks are strongly gender disaggregated (discussed in Chapter 1). Men tend to cut the trees, while women collect, pile and burn the branches. Qualitative fieldwork suggests that while men may pile branches etc, it is very rare for women to cut trees, therefore access to male labour, and in particular male group labour, for these tasks is crucial. This suggests that a woman with a more mixed network, balanced between men and women, rather than being heavily made up of women, may be at an advantage when it comes to organising male labour for these tasks especially if there is limited or no availability of male labour in the household. Using just network size or group affiliation to capture social networks (or social capital) would miss valuable network characteristics such as this, leading to misleading results about connectedness and resource access.

4.6 Conclusion

This chapter has defined social networks variables constructed using the survey data, and analysed relationships between different aspects of social networks and respondents' individual and household characteristics to provide insights into the social networks of respondents in the study sites. Correlations, basic descriptive statistics and some simple multivariate analyses suggest significant associations between various individual and household-level attributes and social networks, and these differ according to the social networks measure used. This implies different groups of individuals have different kinds of networks. Some low correlations between the variables also suggest that different measures of social networks are capturing different things. This is backed up by the OLS results using the social networks measures as dependent variables. Here, estimates on the parameters change according to the dependent variable used. That different demographic groups have different kinds of networks is signalled by the way the signs on the coefficients for the same regressor changes across the different regressions. An F-test of the equality of coefficients between the regressions suggests that coefficients do differ significantly across the regressions.

Men have larger networks and are likely to report network partners in more of the network types than women do, women's networks are more diverse in terms of sex of their alters. Being an employer of group labour means the respondent has a bigger network. This is apparent in both the bivariate and the conditional correlations given by the multivariate analysis. The same property does not apply to employees in groups: it appears not to be group interaction itself that generates the network.

The analysis demonstrates that using social networks variables allows us to capture a more detailed picture of social integration going beyond membership of groups and societies. This allows us to uncover more nuanced insights into the role social networks might play in economic life. Not only that, the type of social network variable used matters. While we need to be cautious about ascribing causal relationships between social networks and individual and household characteristics, using only group affiliation or network size to capture social networks properties risks missing other,

important network properties, such as network diversity, that are useful to particular groups: for example the potential value to women of having more mixed sex networks over network size. This may result in misleading conclusions about people's connectedness and the role of their networks in effecting outcomes, suggesting that social networks and social capital research in economics needs to be more mindful of the subtleties between the different properties of social networks. This is in line with methodological work in social capital measurement using data from The Netherlands, which finds different dimensions of social capital are covered by different measures: "the extensity of a network, the variety of persons and their attributes in a social network, and their resources all concern different phenomena" (van der Gaag and Snijders, 2003b: 18). Further, the authors find "different aspects of social capital indicators show up as significant predictors for different returns to social capital" (*ibid*:18). The research turns to this in the next chapter.

.....

Having established that the way we measure social networks matters, the next chapter analyses how the local social environments and networks of small-holder farmers in the three study sites function in relation to accessing economic opportunities. Social networks variables described here allow the research to test hypotheses linking different social networks characteristics to specific outcomes.

Chapter Five

Social Networks and Access to Economic Opportunities

This chapter analyses the social networks of smallholder farmers in the three study sites, in relation to accessing economic opportunities. Social networks variables, described in Chapter Four, allow the research to test hypotheses linking different social networks characteristics to specific outcomes. Analysis in this chapter identifies relationships between social network characteristics and socio-economic status, and crop market participation: both entry into markets and degree of participation in terms of crop sales. By clustering households according to resource use in combination with crop marketing outcomes, the research examines how might social networks enhance households' existing resources, inputs such as labour for example, or make up for deficiencies in productive resources allowing households to overcome constraints to market participation. In this way, the research investigates whether social networks help to explain why some households do better or worse than might otherwise be expected.

Section 5.1 states briefly three hypotheses relating social networks to outcomes to test in the research, describes the social networks dataset and the outcome variables. Each hypothesis is taken in turn in Sections 5.2-5.4, and simple bivariate analysis is presented linking social networks to outcomes under the three hypotheses. Section 5.5 links smallholder commercialisation to social networks using multivariate analysis by creating a typology of farm households clustered according to command over resources and crop marketing levels and using this as a dependent variable in a multinomial Probit estimation incorporating social networks variables as regressors. Section 5.6 summarises the discussion and concludes.

5.1 Linking Social Networks to Outcomes

This section considers the potential links between social networks and outcomes, setting out hypotheses regarding beneficial network characteristics and couching these hypotheses in relation to outcomes specific to the rural Zambian context. The research takes an ego-centred (or personal) network approach (see Wasserman and Faust, 1994;

Hanneman, 2000). The network consists of a focal individual (ego) and the people they know (alters) and the focus is on structure, composition and shape of the network. Each ego-network is treated as a separate case, and relations between ego and alters form the variables. This approach is rooted in the Manchester School of anthropology (see, among others, Radcliffe-Brown; Mitchell, 1969). Econometric studies that have incorporated a social networks approach, linking social networks to outcomes in the sub-Saharan African context, include: Fafchamps and Minten's (2002) study of the effect of social networks on agricultural trader profitability in Madagascar; A study of technology adoption in rural Ethiopia and Kenya, based on a study of smallholder networks within agro-forestry projects (Darr and Pretzsch, 2006); and Conley and Udry's (2010) study of Ghanaian farmers' social learning and technology adoption through networks.

5.1.1 Social Networks Hypotheses

Social network theory and empirical studies of social networks and social capital highlight a number of 'standing hypotheses' linking (ego) networks to outcomes:

Hypothesis 1: Bigger is better. A bigger network is better for ego's "individual goal attainment" (Bourdieu, 1986; and Burt, 1992).

Hypothesis 2: There are benefits to having 'friends in high places'. Positive social capital results from accessing network members (alters) with high prestige (Lin 2001).

Hypothesis 3: The strength of heterogeneous networks. Those with heterogeneous networks are better able to attain their goals (Granovetter, 1973; 1983; Lin et al 1981; Campbell et al, 1986).

These hypotheses are examined in detail below. It is worth noting here that these points are arguable. There is evidence to suggest that larger networks arise in response to 'resources' available from a specific focal individual or ego (see Borgatti, 1998, for a summary), raising the issue of endogeneity. A large network can thus imply a greater drain on ego's resources as a result of obligations to a number of people, and in this case this individual could constitute a 'high prestige' network partner from the point of view of their alters. Social Resource Theory (Lin, 1982; 2001) states that tie strength is

linked to the kinds of resources provided: diverse social resources are necessary for *instrumental* actions such as buying goods or looking for work, and tend to be accessed through weak ties, whereas so-called *expressive* actions (such as emotional support) are more likely to be carried out by strong ties. Kin or relatives usually know each other and therefore are likely to mobilise quickly when certain types of assistance are needed, for example in response to illness (see Borgatti, 1998). Networks spanning generations can be a rich source of resources, for example, young people providing older network partners with physical labour, and older people providing younger people with knowledge or skills (see Borgatti, 1998). In communities like the study sites in this research people traditionally have strong kin networks. An empirical question here is whether or not a large kin network is likely to contain more heterogeneous alters, in terms of sex and age, given such networks tend to span generations and contain cross-sex links that otherwise would be less socially acceptable? And following from this, is this the kind of diversity in networks that is useful in economic life in such settings? In general, the larger ego's networks, the more support they receive, and more support is given by network partners to focal individuals (egos) with larger networks. It is not clear why this should be the case but it has been put forward that perhaps alters in such networks perceive ego to be more important and therefore deserve more help (Borgatti, 1998).

The hypotheses provide a starting point for conceptualising properties of networks and their implications for socio-economic and livelihood outcomes. They provide theoretical and empirical foundations to help to guide the choice of variables constructed using social networks data captured in the household survey (see van der Gaag and Snijders, 2003a; 2003b; 2004; Wasserman and Faust, 1994). To test these hypotheses, variables are needed to capture: i) network properties, for example: size; network heterogeneity; group affiliation; ii) properties of the ties contained within the network, for example: multiplexity (multiple relations among sets of actors); and iii) outcomes.

5.1.2 Social Networks Variables

Variables to capture properties of social networks and of the ties contained within the network constructed from the dataset are described in Chapter 4. These are:

- i) *Network Size*: total number of alters; total number of groups in which household is a member;
- ii) *Prestige Networks*: dummy variable for influential alter; location of influential alter;
- iii) *Diversity in Composition*: age and sex-heterogeneity of alters; proportion of kin in network;
- iv) *Diversity in Function*: network multiplexity measures capturing number of multiples ties, proportion of ties that are multiplex, number of resources accessed through networks, average network role relationships of alters.

5.1.3 Outcome Variables

The hypotheses above suggest some “gain” or “benefit” attached to social networks: bigger is better, in relation to goal attainment, for example, but the meaning of “better” needs to be defined and outcomes (or ‘goals’) need to be relevant in the context of rural Zambia. These are predominantly smallholder farming communities, where even people who tend to work for others nevertheless engage in some cultivation on their own account. A pair-wise ranking exercise was carried out in focus group discussions in all three study sites, during fieldwork for the research, designed to uncover what villagers themselves considered to signify ‘getting on’ in terms of socio-economic success. Structured discussion encompassed: what people considered ‘getting on’ to be; the factors they believed to be necessary in achieving this; and the ordered preference of these factors in relation to each other (see Appendix C for an explanation of the exercise and field-notes).

Participants universally agreed that ‘getting on’ means “having a better life” through achieving a successful livelihood, and saw the route to this to be income-generation through farming. In each focus group, participants, unprompted, listed all the elements they thought necessary for farming success, which could run into twenty or more factors. These included the obvious assets and inputs necessary for cultivation, and also

encompassed personal relations: cooperation (between people both within and outside the household, working together); people: relatives, workers and other people more generally; ‘being famous’ i.e. popular (thought to be necessary for attracting and keeping customers and attracting labour). Personal attributes also appeared in the list produced by Lufubu participants: honesty, sobriety and cleanliness, with a discussion about the importance of these for doing business with other people in general. The participatory pair-wise ranking exercise therefore suggests that people consider their social networks to be an important factor in determining success in smallholder farming enterprises in the three study sites.

An overarching question is what determines households’ differential ability to sell crops in the market, if on the surface at least there is little variability between households’ control over other productive resources. Do the (characteristics of) a household’s social networks play a role, and through which mechanisms? In relation to the three hypotheses described above, the research hypothesises therefore that social networks reduce transaction costs in economic participation, enabling people to access markets (local, regional, national, and international) for their crops. For example, networks can be important sources of information and, through greater levels of trust, can lower supervision and other transactions costs, making it easier for some farmers to access markets for their produce.

The research thus considers three outcome variables:

1. Socio-Economic Status;
2. Degree of Market Participation given as value of crop sales ; and
3. Crop Market Participation (yes/no).

The socio-economic status measure is an index constructed using survey data on: ownership of consumer durables, agricultural implements, livestock units, housing quality, transport and water source. It ranges from zero to 1 (see Appendix D for an explanation of how the index is constructed)³⁷. Here, a categorical variable of the index

³⁷ The way the socio-economic status variable has been constructed is a potential limitation of the study. A simple construction using aggregation and averages across key variables, driven by theoretical and empirical considerations, was chosen over a statistical method that selects and weights variables arbitrarily, such as factor or principal components analysis (see Vyas and Kumaranayake, 2006).

taking three values, ‘high’, ‘medium’, ‘low’, is used. This is a relative measure based on percentiles, with those scoring in the top third of the sample falling in the category ‘high’, the next third are classed as medium socio-economic status, and the bottom third in the low socio-economic status category.

Degree of crop market participation (or level of commercialisation) is a categorical variable based on total value of sales across all crops sold, and an aggregate of round 1 and round 2 crop data to give sales over a year³⁸. A table of percentage of households growing each crop by community, and average prices per kg are given in table E2.8 in Appendix E, and descriptive statistics are in Table 5.1. Level of commercialisation is given as either: zero, low, medium, or high value of crop sales, in order to explore associations with categorical social networks variables. Cut-off points were determined by previous studies on agricultural commercialisation in Northern Province, Zambia, where a farming household was considered to be commercialised if they sold more than 30 bags of maize and ‘emerging’ if they sold between 1 and 30 bags of maize (see Table 1.1 in Chapter 1 and Moore and Vaughan, 1994). This was converted into a value at current prices to give a cut-off point of crop income of 1.8 million Kwacha, beyond which a farmer is considered to have a ‘high’ level of commercialisation. The category ‘emerging’ was further split into ‘low’ and ‘medium’ commercialisation with the cut-off point between the two set at 800,000 Kwacha, the amount needed to invest in an asset such as a head of cattle plus a margin for other farm inputs and some consumption. For the households in the research commercialisation is considered across sales of *all* output sold, regardless of traditional classifications of crops as cash crops (for example, maize) or cash crops (for example, cassava). Crop market participation is also given by a dummy variable for market entry, with a value of zero for no crop sales and a value of 1 for crop sales greater than zero, in order to analyse determinants of whether or not a household sells, as well as how much.

³⁸ The crops sold variable is an aggregate of sales in thousand kwacha across the following crops: Maize; Hybrid Maize; Cassava; Millet; Sorghum; Mixed beans; Soybean; Sweet potato; Irish potato; Ground nut; Cabbage; Mango; Banana; Plantain; Squash; Orange; Tomato

Table 5.1 Outcome Variables: Summary Statistics N=136 Households

Outcome Variable	Unit	Mean	Std. Dev.	Min	Max
Socio-Economic Status (SES)	Scale: 0 to 1 Low=42 households Medium=46 households High=48 households	0.33	0.17	0.11	0.81
Total annual sales agricultural food crops	Thousand Kwacha Zero=22 households Low=63 households Medium=33 households High=21 households	806.0	1541.7	0	9439.48
Market participation	1=Yes; 0=No (% yes)	83.8

Outcomes in terms of both socio-economic status and crop income are measured at the household level. The socio-economic status variable is an index based on household assets, and crop income was reported at the household level by individual respondents during the survey, even though this part of the questionnaire was administered at the individual level. Social networks data, however, were reported at the individual level so to allow analysis in relation to outcomes the data are aggregated to give social networks scores at the household level. This is defensible in that it can be argued that an individual's networks are to the potential benefit (or detriment) of the entire household. The links of one household member provide access to all household members to the externalities associated with the connection (see Haddad and Maluccio, 2002). These can be picked up by household fixed effects, shown in the multivariate analyses at the individual level in Chapter 4. Qualitative fieldwork as part of the research suggests this is indeed the case. Thus, where there are network data for more than one individual in the household the maximum value is taken. For group membership, the variable is the total number of different groups in which the household has members.

The chapter now discusses each hypothesis in turn, describing the social networks measures constructed in relation to the hypotheses and presenting relationships between social networks measures and the outcome variables at the household level. Only statistically significant relationships are reported, unless stated otherwise.

5.2 Hypothesis 1: Bigger is Better

Hypothesis 1 states that a bigger network is better for "individual goal attainment" (see: Bourdieu, 1986; and Burt, 1992). But the type of goal may matter, and all types of resources in a network may not always be important at any given time. However, the larger the network the more likely it is to contain the very specific resources necessary for achieving certain goals, for example job search. Emerging evidence points to negative externalities associated with large social networks. This is related to forced redistribution under a social contract among kin and extended family (Hoff and Sen, 2006; Lewis, 1955). Examining the role played by extended family and kin in the context of the absence of formal insurance markets in KwaZulu-Natal, DiFalco and Bülte (2011) suggest tentatively that more extensive kinship networks may be associated with lower incomes. Analysing borrowing behaviour in Cameroon, Baland et al. (2011) find "that some people 'pretend to be poor' by both depositing and borrowing money. While excess borrowing is costly, it signals poverty and suggests an inability to respond to demands for financial assistance from kin members".

5.2.1 Social Network Measures: Network Composition

To test the hypothesis that a bigger network relates to better outcomes, network size variables used are: i) total number of alters (network size) and ii) Group affiliation (as a contributor to network size). Table 5.2 summarises network composition variables at the household level.

Table 5.2 Network Size Variables – Household Level

Variable name	Definition	Mean	St. dev
size	Total number of network partners cited across all role relationships (interaction, employer, employee, migration, marketing, social support, information)	11.3	6.2
affil	Group Affiliation variable. Membership of village-level and other formal or informal groups (number of groups household has membership in)	2.4	1.6

5.2.2 Network Size and Outcomes

This section considers relationships between the network size variables described above and the outcome variables: socio-economic status; market participation; and level of market participation. Only statistically significant relationships are reported. Table 5.3 sets out social network composition characteristics of individuals by socio-economic

status and level of crop marketing, reporting the mean values for each of the variables: network size; group affiliation (total number of groups).

Potential endogeneity cannot be ignored. Generally, more ‘successful’ people tend to attract more followers, while a bigger network may enable people to achieve a higher socio-economic status through greater resource access. The research returns to endogeneity in the multivariate analyses in Chapter 6.

Table 5.3 Mean Household Network Composition and Outcomes

Respondent attributes:	Total Number of Households	Network size	Group affiliation (number of group memberships)
Socio-economic status			
Low	42	8.9	1.5
Medium	46	11.7	2.9
High	48	13.2	2.9
Crop Marketing			
Zero	22	8.8	1.5
Low	63	10.9	2.3
Medium	30	11.9	3.1
High	21	14.6	3.0

Source: Household Survey, 2002-03

Those with a high socio-economic status tend to have bigger networks than those with a low socio-economic status, as indicated in Table 5.3. Breaking down the network size variable into a categorical variable with four categories of network size, people of medium socio-economic status relative to the sample as a whole are over-represented in the >15 and 6-10 network size categories, and under-represented in the 5 or less and 11-15 categories (Table 5.4). Correlation between network size and socio-economic status is only significant at the 10 per cent level.

Table 5.4 Distribution of Network Size and Group Affiliation by Socio-Economic Status (% households)

Socio-economic status	Network Size (number of Alters)				Group Affiliation	
	5 or less	6-10	11-15	>15	No	Yes
Low	31.0	35.7	23.8	9.5	14.3	85.1
Medium	10.9	39.1	23.9	26.1	2.2	97.8
High	10.4	35.4	25.0	29.2	2.1	98.0
Total	16.9	36.8	24.3	22.1	5.9	94.1

Chi-sq=11.6; Pr = 0.071

Chi-sq=7.75 Pr = 0.021

Source: Household Survey, 2002-03

The group affiliation variable here is a simple dummy variable taking a value one if the respondent is a member of community-level groups, clubs and societies, and zero if not.

It does not distinguish between those who are members of just one or many groups, between those who are group leaders or members, nor the degree of participation in groups. Group membership is commonly used as a measure of social capital. Group affiliation correlates significantly with socio-economic status, although those in groups are only slightly over-represented in the higher socio-economic categories (Table 5.4).

Most households (94.1 per cent) are involved in groups at the village level, and are slightly over-represented in the group of sellers (95.6 per cent). Cross-tabulations also show that sellers are also slightly over-represented in the group of farmers' group and cooperative members than the distribution across the sample as a whole would suggest, though some non-selling households are still members of such groups. Overall, only 37.5 per cent of households surveyed are members of farmers groups and co-operatives (Table 5.5).

Table 5.5 Distribution of Network size and Group Affiliation by Crop Market Participation (% households)

Participation in crop marketing	Network size (%)				Group affiliation (%)		Member of farmers group (%)	
	5 or less	6-10	11-15	>15	no	yes	no	yes
No	40.9	27.3	18.2	13.6	13.6	86.4	86.4	13.6
Yes	12.3	38.6	25.4	23.7	4.4	95.6	57.9	42.1
Total	16.9	36.8	24.3	22.1	5.9	94.1	62.5	37.5

Chi-sq=10.8; Pr = 0.013

Chi-sq=2.85; Pr = 0.091

Chi-sq=6.38; Pr = 0.012

Source: Household Survey, 2002-03

More non-sellers have fewer than six network partners compared with those selling crops (Table 5.5). As farmers move into higher levels of commercialisation measured in terms of value of crops sold, the associated network size grows (Table 5.6). Once again, the direction of causality is not clear here, but if higher levels of crop sales imply lower transactions costs (both fixed and variable) then logically a larger social network is also associated with lower transactions costs. Farmers operating at lower levels of commercialisation are over-represented in the category of 6-10 network partners; farmers at medium levels of commercialisation are over-represented in the category of 11-15 networks partners; and those at highest levels of commercialisation are over-represented in the category of more than 15 network partners.

Table 5.6 Distribution of Network Size and Group Affiliation by Crop Sales (% households)

Crop sales	Network size (%) (number of people)				Member of farmers group (%)		Group affiliation dummy (%)	
	5 or less	6-10	11-15	>15	no	yes	no	yes
Zero	40.9	27.3	18.2	13.6	86.4	13.6	13.6	86.4
Low	14.3	44.4	23.8	17.5	66.7	33.3	6.3	93.7
Medium	10.0	26.7	40.0	23.3	53.3	46.7	0.0	100.0
High	9.5	38.1	9.5	42.9	38.1	61.9	4.8	95.2
Total	16.9	36.8	24.3	22.1	62.5	37.5	5.9	94.1

*Chi-sq=22.4; Pr = 0.008**Chi-sq=12.2; Pr = 0.007**Chi-sq=4.33; Pr = 0.227**Source: Household Survey, 2002-03*

Sellers of higher volumes of crops also tend to be over-represented in farmer group or cooperative membership, although only 37.5 per cent of households reported such memberships (Table 5.6).

So far, the data on the whole appear to bear out hypothesis 1, that bigger networks are related to better outcomes, both in terms of socio-economic status and levels of crop marketing. Both larger networks and group affiliation are associated with higher socio-economic status, being a crop selling household and having higher levels of crop income.

5.3 Hypothesis 2: Friends in High Places

The second hypothesis states ‘there are benefits to having friends in high places’. Positive social capital results from accessing network members with high prestige (Lin, 2001). This relates to ‘bridging’ or ‘linking’ networks: links with individuals who are dissimilar to ‘ego’ and provide access to a more diverse flow of resources (See Woolcock, 2003).

5.3.1 Social Network Measures: Prestige and Influence

Prestige networks are captured as a dummy variable signifying the presence or absence of influential kin or friends, and a categorical variable for location: - whether they live in ego’s own community (village) or elsewhere. Summary statistics for the measures at household level are given in Table 5.7:

Table 5.7 Prestige Network Variables – Household Level

Variable Name	Definition	Mean	St.Dev
kininf	Whether or not respondent's network comprises influential kin or friends	0.71	0.46
kinloc	Location of influential kin/ friend	0.96	0.74

5.3.2 Influence and Outcomes

Table 5.8 reports the percentage of households reporting an influential alter in their networks, by socio-economic status and level of crop marketing. A higher proportion of households with high socio-economic status report having an influential alter compared with those falling in the medium or lower socio-economic status categories.

Table 5.8 Percentage of Outcome Group with Influential Alters

Respondent attributes:	% of Households	Influential alter
Socio-economic status		
Low	31	59.5
Medium	34	65.2
High	35	85.4
Crop Marketing		
Zero	16	45.5
Low	46	68.3
Medium	22	76.7
High	15	95.2

Source: Household Survey, 2002-03

Sellers are slightly over-represented in the group of those with ‘prestige’ networks captured as kin or friends in an influential position (Table 5.9). Locations of these influential alters were also recorded, that is, whether they lived in the respondents’ community or elsewhere, to get a sense of exactly how ‘bridging’ these relationships might be in terms of linking into wider networks in different locations. For non-sellers influential alters are split relatively evenly between own village and other locations, whereas sellers had a higher proportion of influential alters in their own community (47.4 per cent compared with 25.7 per cent living elsewhere), but were still slightly over-represented in the sub-sample of those with influential alters elsewhere than the distribution as a whole would suggest. This makes sense in the context of the research in signifying the value to smallholder farmers in remote regions of having people close by to help get things done in everyday cultivation activities, such as providing labour, transport, help with trading. Further, if selling is seen as a desirable attribute by other in the village this could attract influential alters.

Table 5.9 Distribution of prestige networks by crop market participation (% households)

Participation in crop marketing	Influential alter		Location of influential alter		
	no	yes	no influential alter	own village	other location
No	54.5	45.5	54.5	22.7	22.7
Yes	24.6	75.4	24.6	49.1	26.3
Total	29.4	70.6	29.4	44.8	25.7

Chi-sq=7.99; Pr = 0.005 *Chi-sq=7.43; Pr = 0.024*

Disaggregating by level of crop sales, medium and high sellers are relatively more likely to have kin in an influential position and in a location other than the households' own community than the distribution across the sample as a whole would suggest (Table 5.10).

Table 5.10 Distribution of prestige networks by level of crop sales (% households)

Crop sales	Influential alter		Location of influential alter		
	no	yes	no influential alter	own village	other location
Zero	54.5	45.5	54.5	22.7	22.7
Low	31.7	68.3	31.8	44.4	23.8
Medium	23.3	76.7	23.3	50.0	26.7
High	4.8	95.2	4.8	61.9	33.3
Total	29.4	70.6	29.4	44.9	25.7

Chi-sq=13.5; Pr = 0.004 *Chi-sq=14.4; Pr = 0.026*

Prestige networks, measured as whether or not ego has an influential alter and the location of this alter, appear to be associated with higher socio-economic status as well as higher levels of crop income. In terms of location of influential alters, those with higher levels of crop income are also more likely to cite influential alters living outside of ego's immediate community (location), than those at lower levels of crop income. The bivariate suggest therefore that friends in high places matter, especially those living close to the respondent. People living close by are more likely to be accessible on a day-to-day basis and therefore potentially more useful when it comes to everyday cultivation and marketing activities, thus helping to reduce ego's transaction costs.

5.4 Hypothesis 3: The Strength of Heterogeneous Networks

This hypothesis states those with heterogeneous networks are better able to attain their goals. The greater the diversity of the network (in terms of social resources accessed

through the network or range of network relationships), the more chance there is that someone in the network has something that the ‘ego’ needs. High diversity implies integration into several spheres of society or social circles/ contexts and this is considered to be advantageous for mobilising resources and for instrumental actions like gathering information (Lin et al, 1981; Campbell et al., 1986; Kadushin, 1982). This relates to Granovetter’s “Strength of Weak Ties” argument (Granovetter, 1973; 1983). Weak ties between alters provide access to other people’s clusters of closely knit friends and thus act as a crucial bridge to novel information ‘from distant parts of the social structure’, an advantage in competitive environments (see also Burt, 1992). By contrast, alters with strong ties to each other tend to know the same things ego knows. The denser the ties in ego networks, the stronger these ties tend to be, implying that the network is more insular and therefore more homogeneous. This may restrict the different types of help a person can obtain via their network.

5.4.1 Social Network Measures: Diversity in Composition and Function

Variables that allow exploration of heterogeneous networks hypothesis can reflect the differing personal attributes of alters, that is *diversity of composition*, or differences in the ties between alter and ego: *functional diversity*. These are discussed in Chapter 4, and in relation to outcomes below.

B. Diversity of Composition

Heterogeneity or dispersion measures of network partner attribute variables are the most direct measures of the diversity of alters that ego can contact within his/her personal environment. Respondents’ networks may be highly heterogeneous in some respects yet homogeneous in others. This gives some idea of the range of someone’s social network - that is, potential access to social resources (based on the ‘strength of weak ties’ argument outlined above). Size can also be used as a measure of range. In general, heterogeneity increases with network range. Network heterogeneity is captured by diversity in age and sex of network partners, and kin composition of networks (Table 5.11).

Table 5.11 Composition Diversity Variables – Household Level

Variable Name	Definition	Mean	St.Dev
agehet	Age heterogeneity of network partners (standard deviation of network partner ages, across each network)	11.96	4.81
sexhet	Sex heterogeneity of network partners (Simpson's Index of Qualitative Variation in sex of network partner, across each network)	0.23	0.18
propkin	Proportion of network members who are kin	0.40	0.27

B. Functional Diversity

Functional diversity of a network can be captured by measures on tie attributes, describing the quality of relations between the respondent and network partners.

Diversity in the *ties* between the respondent and their network partners is measured using a variable capturing *multiplexity* in the relationships between the respondent and their network partners. Functional diversity (multiplexity) measures at the household level constructed from the survey data based on tie characteristics are summarised in Table 5.12.

Table 5.12 Functional Diversity Variables

Variable Name	Definition	Mean	St.Dev
multiplex	Total number of multiplex ties	2	1.5
multidum	Dummy variable for the existence of multiplex ties	0.9	0.4
propmult	Proportion of total number of ties that are multiplex (multiplex/size)	0.2	0.1
ressum	Sum of resources accessed across network (max=20)	13.3	3.5
multcont	Average number of 'contents' ie network roles per alter (max=7)	3.5	0.9

The next section sets out descriptive statistics for both composition and functional diversity in relation to outcomes.

5.4.2 Network Diversity and Outcomes

Table 5.13 sets out social network characteristics of households by socio-economic status and level of crop marketing, reporting the mean values for each of the variables: age and sex heterogeneity; network multiplexity.

Table 5.13 Mean Network Diversity and Outcomes

Outcomes:	Age heterogeneity	Sex heterogeneity	Ratio of kin to non-kin	Number of multiplex ties	Proportion of multiplex ties	Sum of accessed resources	Average no. of roles per alter
Socio-economic status							
Low	13.3	0.22	0.49	1.71	0.19	11.1	3.4
Medium	11.5	0.21	0.42	2.07	0.18	14.5	3.5
High	11.6	0.28	0.37	2.19	0.19	14.0	3.6
Crop Marketing							
Zero	14.8	0.20	0.40	1.41	0.18	11.2	3.1
Low	11.5	0.24	0.48	2.16	0.19	13.4	3.5
Medium	12.4	0.23	0.39	2.13	0.22	14.4	3.6
High	10.6	0.29	0.32	1.95	0.14	13.4	3.6

Source: Household Survey, 2002-03. Note: The higher the score the more diverse the network on the dimension in each column.

The network heterogeneity hypothesis suggests that the more diverse the network the more valuable it is, for example by providing novel information. Mean values of the network partner heterogeneity variables given in Table 5.13 show that while networks that are more mixed in terms of sex of network partner appear to be associated with higher levels of socio-economic status (significant at only 10 per cent), respondents have on average more *homogeneous* networks in terms of age of network partners. Households with higher socio-economic status and households selling higher values of crops also tend to have a lower ratio of kin to non-kin network members implying greater relational diversity. Higher socio-economic status also has slightly higher average levels of network multiplexity, in terms of number of multiplex ties. However this is not significant. While sellers appear to access more resources through their networks than non-sellers, there appears to be little variation in network multiplexity measured as proportion of multiplex ties and as average number of contents per alter.

Turning to crop market participation, non-sellers tend to have the most age heterogeneous networks, while those with highest levels of crop sales are over-represented in lower age-heterogeneity categories (Table 5.14).

Table 5.14 Distribution of age heterogeneity of alters by crop market participation and crop sales (% households)

Participation in crop marketing	Age heterogeneity (%)			
	<5	5-<10	10-<15	15+
No	0.0	18.2	31.8	50.0
Yes	6.1	33.3	41.2	19.3
Total	5.1	30.9	39.7	24.3
<i>Chi-sq=10.3; Pr = 0.016</i>				
Crop sales				
Zero	0.0	18.2	31.8	50.0
Low	6.3	30.2	47.6	15.9
Medium	0.0	36.7	40.0	23.3
High	14.3	38.1	23.8	23.8
Total	5.1	30.9	39.7	24.3
<i>Chi-sq=18.6; Pr = 0.029</i>				

The variable measuring sex heterogeneity in respondents' network partners – the index of qualitative variation in sex of network partner – takes values between zero and one. The more diversity there is between network partners the closer the score is to one. In cross-tabulations this variable had no statistically significant relationships at the 5 per cent level with the outcome variables, and significant relationships with just two of the respondent attribute variables, those related to sex and household status of respondent, and at the 10 per cent level with sex of household head. These variables are highly correlated with each other (see Appendix E for bivariate relationships between social networks variables and individual and household level attributes).

In examining social networks variables in relation to outcome variables for socio-economic status and crop marketing, broad patterns are beginning to emerge in relation to the types of network characteristic related to positive outcomes. Higher levels of commercialisation measured as value of crops sold, and higher socio-economic status, are associated with a larger network and group affiliation (Hypothesis 1; Bigger is Better). Large networks imply that ego has more chance of accessing resources needed to lower transactions costs of cultivation and marketing. Prestige networks also appear to be associated with higher socio-economic status and higher levels of crop income, and sellers are more likely to have influential alters living within their immediate community (village), than non-sellers. The bivariate appear to uphold Hypothesis 2, that 'friends in high places' matter for better outcomes, especially those living close to the respondent. Being able to draw on valuable network links locally can contribute towards higher crop sales by reducing ego's transaction costs in their daily cultivation

activities. Turning to network heterogeneity (Hypothesis 3: the strength of heterogeneous networks), higher socio-economic status is associated with having a more mixed network gender-wise. Non-sellers' networks are more age-diverse, in contrast to households with higher incomes from crop marketing whose networks tend to be less so. While the above bivariate relationships suggest positive associations between social networks and outcomes related to transactions costs effects, the direction of causality is still not clear.

The above analysis suggests broad patterns in the social networks characteristics of sellers and non-sellers. The research is interested in how social networks might provide access to resources that enable farming households to participate in markets and/or participate in markets at a higher level of commercialisation. Implicit in this is the question of how might social networks enhance households' existing endowments, inputs such as labour for example, or make up for deficiencies in productive resources allowing households to overcome constraints to market participation. Similarly, might an absence of certain types of networks or network resources act as constraints on crop marketing behaviour regardless of resource access?

Before turning to causal analysis of social networks in relation to transaction costs in agricultural commercialisation in Chapter 6, the remainder of this chapter delves further into the links between social networks and crop market outcomes by grouping households according to their command over resources in addition to their economic behaviour. This allows the research to identify households that might be doing better or worse than their resource status would suggest, and explore whether social networks help to explain why.

The next section uses a typology of households to create categories relating level of commercialisation with command over resources. These categories are then used in a multinomial Probit, which includes social networks measures as explanatory variables, to explore further the potential role of social networks in achieving positive outcomes.

5.5 Smallholder Commercialisation and Command over Resources

The preceding analysis has focused on general patterns in relationships between network attributes and different categories of socio-economic status and crop market participation. Interesting questions arise in relation to households falling outside the broad patterns emerging from the bivariate analysis and summarised above. One of the over-arching motivations for the research is to explore the relevance of social networks in the economic lives of households in rural sub-Saharan Africa, households that on the surface appear to be similar, especially in terms of their asset base or command over resources, but might have markedly different socio-economic outcomes. The focus here is on crop marketing income. What is it that sets certain households apart from their neighbours, specifically, households that, due to relatively poor access to resources such as land, labour and other types of capital, one would reasonably expect to find it more difficult to overcome transactions costs as barriers to entry to markets and thus have low levels of crop marketing, but instead engage in a high level of marketing? Are such households able to overcome these barriers because of social networks? Or are other factors at play? The next section explores this by deepening the analysis through categorising households in relation to both their output levels and input use. This allows the research to identify so-called ‘atypical’ households - badly-resourced high-commercialisers, well-resourced low-commercialisers - and potentially uncover the value social networks might have in overcoming deficiencies or cancelling out advantages in productive resource endowment.

5.5.1. *A Typology of Smallholder Farming Households*

The previous analysis suggests there are positive associations between social networks and outcomes of smallholder farming households in the study sites. However, the analysis so far does not distinguish among the group of commercialised farmers in terms of resources. Do the benefits of social networks go hand-in-hand with other positive attributes such as command over resources (labour, farm inputs), or do they reduce transactions costs in selling by mitigating resource constraints?

The descriptive statistics in Appendix E examining correlations between endowments and market participation suggest that while it is true that highly commercialised farmers

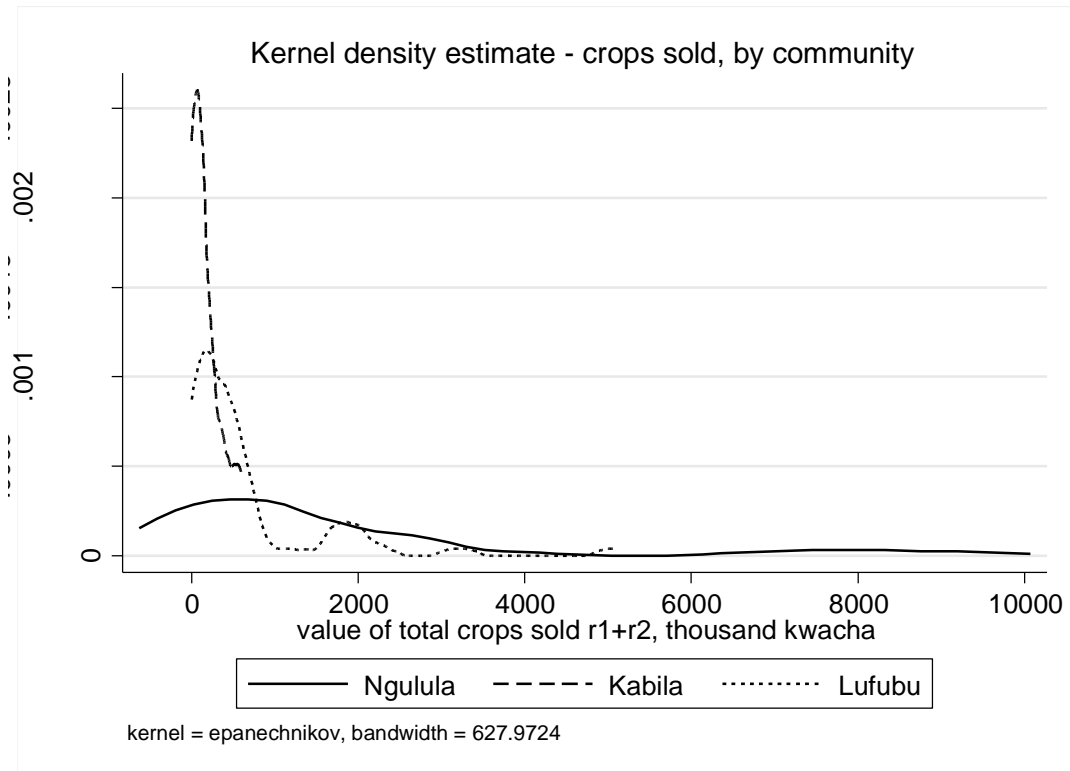
seem better endowed, and the low commercialised are not so well endowed, there are exceptions to this. Cross-tabulations between the crop market participation dummy and various productive inputs are given in Table E2.2 (Irrigation, fertiliser, hybrid seed) and E2.3 (storage and large tools). These show significant correlations between crop marketing and inputs. For these inputs, the largest groups are those without these assets but are still marketing produce, and there is a small percentage of respondents who do possess such assets but do not market at all. When the crop sales variable is broken down into the categories for level of crop marketing, a similar pattern emerges. While better-endowed households are over-represented in the categories of higher levels of commercialisation, there are once again notable numbers of households without such endowments who are also marketing at comparably high levels, as well as those endowed with resources who market at very low levels, if at all. Further, there is a very weak correlation between level of crop sales and total landholding ($\rho = 0.12$).

The differences may partly be explained by location. Communities were sampled purposively to represent three levels of market access: Ngulula with good market access; Kabila with poor market access; Lufubu somewhere between the two. The breakdown of the crops sold variable reflects this, with much higher average levels of crop sales for Ngulula, the geographically better-connected community, compared with Kabila and Lufubu (Table 5.15).

Table 5.15 Value of crops sold, by community '000 Kwacha

Community	Mean	St. dev	Min	Max
Kabila	162.83	183.98	0	607.83
Ngulula	1681.92	2400.9	0	9439.5
Lufubu	615.47	897.38	0	5074.14

The kernel density distribution of value of crops sold variable by community suggests three different distributions (Figure 5.1).

Figure 5.1 Kernel density estimate for crops sold, by community

The chi-square statistics from a Kruskal-Wallis equality-of-populations rank test indicate that there is a statistically significant difference among the three study sites ($\chi^2(2)=21.39$; and 21.49 with ties).

However, location does not tell the whole story. While none of the households sampled in Kabila are achieving “high” levels of commercialisation, there are households in Lufubu who are achieving levels of commercialisation comparable with the higher levels achieved by their more advantaged Ngulula counterparts. All three communities have households who grow but do not sell crops, and have households who are not well resourced (and in the case of Kabila and Lufubu, with disadvantaged market access) but do sell at medium and high levels of commercialisation (not shown). Social Networks may provide an explanation.

In order to examine further the role social networks might have in reducing transactions costs in the context of poor, smallholder households facing resource-constraints and incomplete or missing markets for key productive resources, alternative clusters of households are identified in the data by creating a typology based on categories for low and high levels of marketing and low and high command over inputs. A multinomial

probit is used to identify potential correlates for each group of households. This allows further disaggregating of the households into different kinds of groups that go beyond the very simple seller/ non-seller categories. Given obvious issues with endogeneity (importantly, circular causality and mutually reinforcing relationships, as well as selectivity bias) already touched on earlier in this chapter, the aim here is not to establish causal relations and thus endogeneity is not addressed econometrically at this stage. The research returns to the issue in the next chapter. For now, the aim is to draw initial insights from the data to use in the subsequent multivariate analysis in Chapter 6.

Asset accumulation enables smallholder farmers to respond to incentives to commercialise (see Chapoto et al, 2011; Moser and Dani, 2008; Leavy and Poulton, 2007; Jayne et al, 2003; Binswanger and Rosenzweig, 1986). If production and hence marketing are highly dependent on command over inputs, especially labour, then a typology of farmers can be given by:

Table 5.16 Farmer Clusters - Typology

Marketing	Command over inputs (labour and bought inputs)	
	High (H)	Low(L)
High (H)	A (HH)	B (LH)
Low (L)	C (HL)	D (LL)

One would expect most farmers to fall within categories A (HH) (high command, high marketing) or D (low command, low marketing). What is it about the household that leads them to be in this position? What types of farming households fall at the two extremes: struggling to get by and not selling anything i.e. very poor; or those who are relatively well-off and cultivating with every intention of selling? And what of discontinuities - that is, those farmers falling in categories B (low command, high marketing) or C (high command but low marketing)? Are such farmers the exceptions? Or can distinct groups be identified, with common characteristics related to their social networks? If social networks do mediate access to markets then one might expect social networks to have an effect in isolation from other assets and inputs.

Households were clustered according to their command over purchased inputs - fertiliser, hybrid seed and group labour – using simple dummy variables for each input. This was informed by commercialisation studies that consider farmers to be commercially oriented when they participate in commercial input markets i.e. purchased

inputs (von Braun and Kennedy, 1994. See also Jayne et al, 2003). Land was not included as an input given the customary land tenure system in two of the villages and the nature of land allocation in Lufubu resettlement scheme means land distribution has low variability between households. A household scoring zero would be labelled low command in that input, and a score of one denotes command over that input (labelled 'high'). Command over inputs is calculated as a simple additive term over the three input dummies. Scores of zero or 1 are considered low, a score of 2 or 3 is high. In terms of marketing, those whose value of crops sold was less than 800,000 Kwacha are considered to be 'low' marketers, those selling crops to the value of 800,000 Kwacha and above are 'high' marketers. As discussed in Section 5.1.3 above, this is a reasonable cut-off in the context of the research in that 800,000 Kwacha is the amount needed to invest in an asset such as a head of cattle plus a margin for other farm inputs and some consumption. Table 5.17 shows the numbers of households falling within each cluster.

Table 5.17 Farmer Clusters (number of households)

Marketing	Command over inputs (number of households)	
	High (H)	Low (L)
High (H)	66	19
Low (L)	15	36

Pearson $\chi^2(1) = 30.8$ Pr = 0.000

As expected, of the 136 households most farmers fall in either the HH or LL clusters. Nineteen households are high marketers but score low on the inputs variable (LH cluster); fifteen households are low marketers but score highly on command over inputs (HL cluster). These 'LH' and 'HL' households are those that seem to follow an unexpected path: one has resources but does not appear to sell (HL), and the other is selling but does not appear to have good resources (LH). What factors therefore are holding them back/helping them up?

5.5.2 Cluster Determinants

What determines which cluster households fall into? Do social networks play a role? To explore this, a multinomial Probit model³⁹ is estimated using as the dependent variable the cluster variable for the four categories shown in Table 5.17: LL (low command, low marketing); HL (high command low marketing); LH (low command high marketing); HH (High command high marketing). This model is intended only as an exploratory tool to examine patterns in the data rather than any causal relationships, and care must be taken to interpret results bearing this in mind (for example, the relatively small numbers of households contained in clusters would render many statistical tests invalid).

Theoretical models of crop market participation focus on household characteristics as well as transactions costs related to marketing (see literature review in Chapter 2). These include: information, transport availability and costs, input use, land cultivated, access to storage facilities (see: Key, Sadoulet and de Janvry, 2000; Bellemare and Barrett, 2006; among others). The qualitative exercise carried out during fieldwork provides interesting insights for the empirical analysis of correlates of crop market engagement in the three study sites (Appendix C). Important factors identified by the communities include having basic needs met (health and water), access to tools, and people – with an emphasis on relationships: cooperation between people. Determinants identified are a mixture of factors necessary to meet immediate, basic material needs, in terms of consumption and production, such as transport, cash, food, productive assets and inputs (land, tools, water/ irrigation), education – in agreement with the theoretical model. However, in contrast to the theoretical models, people also stressed the

³⁹ Initially, multinomial logit models were estimated but they failed the Small-Hsiao test of the IIA property – independence of irrelevant alternatives – which underlies the MNL. The Small-Hsiao test tests the hypothesis H0: odds (outcome J vs outcome K) are independent alternatives. An alternative test is the McFadden Hausman test of the IIA assumption (H0: difference in coefficients is not systematic). However, Hausman cannot be performed on weighted or clustered data so only the Small-Hsiao test was carried out. Sometimes the Hausman and the Small-Hsiao tests can give conflicting results: see Long and Freese (2006). In the absence of evidence to the contrary the research assumes the Small-Hsiao test results to be correct and in the case of these data one cannot plausibly assume that the four categories are distinct (and the alternatives are dissimilar). This suggests multinomial probit (or a nested logit) is more appropriate, potential identification issues notwithstanding, to avoid the IIA assumption especially without strong a priori beliefs about the relationships between the disturbances for the clusters/ categories (Maddala, 1983; Greene, 2003).

importance of more ‘intermediate’ factors, focusing on personal and relational attributes such as honesty, industriousness and cooperation. However, these personal and relational attributes are seen to be necessary for fostering and maintaining good relations with others, which in turn leads to improved access to assets, labour, information – in other words lowering transactions costs in market access.

Social networks potentially impact directly on outcomes, for example by providing access to a pool of labour at key points in the cultivation calendar, and also indirectly via their effect on transactions costs, for example enabling access to resources through networks such as information on prices and marketing or lowering search and ‘recruitment’ costs of farm labour. In this way, social networks could also proxy for transactions costs.

The household survey does not contain any direct measures of transactions costs, the direct costs to the household of market exchange such as transfer (transport and storage) costs, or those associated with making and enforcing contracts. The research relies instead on proxies. This is in line with other empirical studies of transactions costs in smallholder agricultural commercialisation (see also: Key, Sadoulet and de Janvry, 2000; Bellemare and Barrett, 2006; among others). Proxies for transactions costs in the household survey include: Ownership of radio (*radio*) – obtaining information about markets, prices, climatic conditions is easier to obtain if one possesses a radio; Access to crop storage facilities (*store*); Ownership of bicycle (*bike*); household employs farm labour (*employer*); Non-farm income (‘000 kwacha) (*nfy*); Proportion of land cultivated (*propland*); Use of hybrid seed (*hyseed*); Large livestock ownership (*livelarg*). Productive farm inputs such as hybrid seed and draught animals may help to increase productivity and thus marketing volumes, which in turn can lower transactions costs via economies of scale. The implication is that these factors have a positive effect, that is they are associated with lower transactions costs, thus there should be a positive relationship between the transactions costs proxies and output levels. If social networks do indeed lower transactions costs then we should observe a positive relationship between the social networks measures and these proxies for transactions costs.

The table of correlation coefficients below (Table 5.18) shows significant positive correlations between *propland* and social network size, and with one of the multiplexity

variables (*multcont*), between non-farm income and network size, and between large livestock ownership and relatively larger kin networks (*propkin*). There are some negative correlations but these are never statistically significant.

Table 5.18 Correlation Coefficients: Social Networks and Transactions Costs

	propland	nfy	large livestock
propland	1.00		
nfy	0.01	1.00	
livelarg	0.03	-0.09	1.00
agehet	-0.01	-0.06	-0.06
IQVst2	0.06	0.08	-0.01
propkin	0.22*	-0.13	0.19*
affil	-0.15	0.14	-0.06
size	-0.10	0.17*	0.01
propmult	0.15	0.04	-0.12
multiplex	0.06	0.07	-0.11
multcont	0.29*	0.14	-0.11
ressum	-0.01	0.10	0.01

* significant at 5 per cent

Chi-Square statistics suggest significant correlations between social networks and transactions costs as categorical/ dummy variables, with those possessing assets associated with lower transaction costs over-represented in the categories for higher values (bigger, more diverse etc) of the social network variable⁴⁰.

Certain features of social networks therefore do appear to go hand-in-hand with the ability to lower transactions costs as captured by the proxy variables. Because of these potential direct and indirect effects of social networks on crop incomes, they are included in the empirical estimations alongside other determinants of crop incomes, including the transactions costs proxies, and can also be considered as direct proxies for transactions costs.

⁴⁰ Chi-Square Statistics are as follows: Social network size and: crop storage (chi2(3) = 13.89); bicycle ownership (chi2(3) = 7.49: 10%); employee dummy (chi2(3) = 6.79: 10%); employer dummy (chi2(3) = 20.64). Group affiliation and: storage (chi2(1) = 3.80: 10%); bicycle ownership (chi2(1) = 8.51); employer dummy (chi2(1) = 7.61); employee dummy (chi2(1) = 2.93). Influential kin and: bicycle ownership (chi2(1) = 3.49); employer dummy (chi2(1) = 4.68); uses hybrid seed (chi2(1) = 6.75).

Bivariate analysis (see Appendix E for tables of descriptive statistics) identifies a range of household characteristics and proxies for transactions costs with significant relationships to crop sales, including: sex of household head, education level, access to inputs (fertiliser, hybrid seed, storage, large tools); transport (bicycle); information (working radio); access to labour including working in reciprocal groups. Of the social networks measures captured in the survey, total network size, network diversity (age and sex heterogeneity of networks), group affiliation, and various elements of kin networks (ratio of kin to non-kin, influential kin, location of influential kin) are significant. These encompass more ‘traditional’ social capital measures (group membership, for example, and location of influential kin as a bridging social capital measure) as well as measures rooted in social network analysis techniques.

These relationships of course do not point to any direction of causality, rather there is likely to be an association. In fact the relationship could run in either (both) directions – for example more successful households are so because they have influential kin or successful households are more likely to have kin who get into influential positions because of the commercial success of the household. This issue is revisited in Chapter 6 in the multivariate analysis.

5.5.3 Multinomial Probit

To analyse the determinants of a household falling into a particular cluster according to the typology set out above, a Multinomial Probit model is used.

Under the Multinomial Probit model, assume a multivariate normal distribution on the latent variables, $y^*_i = (y^*_{i1}, \dots, y^*_{ij-1})$.

The Multinomial Probit model can be given by the latent variable y^*_{ij} for the j th alternative:

$$y^*_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta}' + e_{ij} \quad (5.1)$$

where $e_i \sim N(0, \Sigma)$ and $y^*_i \sim N(\mathbf{X}_{ij}\boldsymbol{\beta}, \sigma^2)$ and $i=1, \dots, n$
Choices $j=1 \dots j; j>2$

Where:

X_i is a $(j - 1) \times k$ matrix of covariates

β is a $k \times 1$ vector of fixed coefficients

e_i is $(j-1) \times 1$ vector of disturbances; and

Σ is a $(j-1) \times (j-1)$ positive definite matrix.

For the model to be identified, the first diagonal element of Σ is constrained, $\sigma_{11} = 1$.

The response variable, y_i , is the index of the choice of individual i among the alternatives in the choice set. This is modelled in terms of this latent variable, y_i^* , via:

$$y_i(y_i^*) = \begin{cases} 0 & \text{if } \max(y_i^*) < 0 \\ j & \text{if } \max(y_i^*) = y_{ip}^* > 0, \text{ for } i = 1, \dots, n \text{ and } p = 1, \dots, j - 1 \end{cases} \quad (5.2)$$

where $y_i = 0$ corresponds to a base category.

The model is first estimated excluding social networks variables and then with social networks variables. Regressors include: household demographic characteristics (age of household head); human capital (whether or not maximum level of education in the household exceeds primary school); assets (crop storage; radio access/ownership; proportion of land cropped as fixed and variable transactions costs); a dummy variable for whether or not the household has non-farm income and one to denote whether or not the household is in Ngulula, the village with better market access. Social networks variables include: log of total network size (for Hypothesis 1: bigger is better); location of influential kin (for Hypothesis 2: benefits to having friends in high places); and sex heterogeneity of network partners as a measure of network diversity (for Hypothesis 3: the strength of heterogeneous networks). Because of the relatively small sample size in the HL and LH clusters, a parsimonious model was estimated. Other variables were tried but not included in the final estimation because they were not significant, in all likelihood due to lack of variation. These include: community categorical variable; livestock ownership; sex of household head; access to or ownership of bicycle; network diversity in terms of age of network partners; dummy variable for membership of village level groups and societies (affiliation networks); network multiplexity; sum of resources accessed/relationships across network.

Summary statistics for the dependent variable and regressors are given in Table 5.19:

Table 5.19 Descriptive Statistics: Dependent Variables and Regressors Multinomial Probit

	Unit	Mean/ dist	Std. Dev.
Dependent Variable:			
Farmer Cluster	Low Resource Low Marketing (LL)	48.5	..
	High Resource Low Marketing (HL)	14	..
	Low Resource High Marketing (LH)	11	..
	High Resource High Marketing (HH)	26.5	..
Regressors:			
Farm Capital: Assets and labour			
<i>Fixed transaction costs:</i>			
Ownership of radio	1=Yes; 0=No (% yes)	44.85	..
Access to crop storage facilities	1=Yes; 0=No (% yes)	30.88	..
<i>Variable transaction costs:</i>			
Non-farm income ('000 kwacha)	Number	34.29	83.57
Proportion of land cultivated	Share	0.35	0.35
Human Capital (variable transaction costs)			
Maximum education level in household	0 = Up to Grade 4	8.09	..
	1 = Grades 5-7	42.65	..
	2 = Secondary/ College	49.26	..
Social Networks			
<i>Network Diversity/Composition</i>			
Diversity of network in terms of sex of alters	Index of Qualitative Variation of alters' sex	0.23	0.18
<i>Size</i>			
Network Size	Number	11.35	6.21
Network Size (log)	Number	2.28	0.57
Household Demographics			
Sex of household head	1 = Female 0= Male (% female)	17.65	..
Location			
Categorical variable for location	1 = Kabila	27.21	..
	2 = Ngulula	29.41	..
	3 = Lufubu	43.38	..

Marginal effects are given in Table 5.20, the reference category is low marketing, low control (LL).⁴¹ Coefficient estimates are given in Appendix F⁴².

⁴¹ The model chi-square statistics suggest that neither model is a particularly good fit, chi-squares are slightly high. The multinomial logit estimates gave a much better fit – but these models violate the IIA assumption. Probit coefficients greater than 2, reported for some variables, should be treated with caution.

⁴² Coefficient estimates relate household characteristics to their preference for each cluster/ outcome, and give the effect of an infinitesimal change in each characteristic on the probability

Table 5.20 Multinomial Probit Results: Determinants of Farmer Cluster

Independent Variables	Multinomial Probit Results ^a – Marginal Effects					
	Without social networks			With social networks		
	HL	LH	HH	HL	LH	HH
HH access to storage	-.032	.193***	.090*	.103	.191***	.144**
Age of Household head	-.004	-.002*	-.005**	-.003	-.001	-.004 *
Non-farm income	1.49e-06	8.31e-07**	1.16e-06**	2.26e-06*	8.80e-07**	1.14e-06**
Proportion of land cropped	-.093	-.136**	-.074	-.095	-.057	-.052
Access to working radio	.012	.141***	.274***	.0107	.121***	.252***
Ngulula	.050**	.033**	.560***	.082***	.058***	.575***
Post-primary education	-.063	.019	.121	-.031	.012	.093
Network size (log)				.072	.074	.033
Sex heterogeneity of network				-.825*	-.328***	-.084
Influential kin nearby				.090*	.128**	.156**
Influential kin far location				.303***	.064*	.274***

Observations = 135; * significant at 10%; ** significant at 5%; *** significant at 1%

^a Cluster base category is LL

Taking first the basic model before introducing social networks variables, access to storage facilities has a significant and positive effect on the “choice” between LL and LH (1 per cent significance level), and between LL and HH (10 per cent level of significance), however there are endogeneity issues here. Low marketers may not be in need of crop storage in the first place, the demand for crop storage may be generated by having a larger marketable, or marketed, surplus. Non-farm income is also important for being able to market higher values of crops despite low command over inputs, as is radio ownership and living in Ngulula (significant for all three clusters compared with the base category). This suggests that even if a household has little command over farm inputs, in an advantageous market setting - with good, lower-cost access to markets, decent roads, short distances and available transport - they are more likely to succeed in terms of higher marketed volumes compared with a better endowed household in another location. Age of household head is negative (10 per cent level of significance) for LH households compared with LL – younger households with low control over

of the household being in a particular cluster over the base category (LL: Low command, Low marketing), holding all other things constant.

resources are more likely than their older counterparts to market high. For LH households, proportion of land cropped is also significant, though negative: when a household has a low command over resources, farming a smaller area more effectively may be the key to success, especially if this goes hand-in-hand with labour constraints. Whether or not the household has a member with post-primary education is not significant.

In the estimation including social networks, age of household head is only important for HH compared with LL households, and now only at 10 per cent. The sign is negative suggesting the older the household head the less likely the household is to be in the well-resourced high-marketing cluster compared to LL. This suggests that younger farmers have an edge over their older neighbours. This could be due to greater physical capabilities of younger farmers. The effects of non-farm income and radio ownership are the same as previous results. Storage access is now more significant for HH compared with LL. Household location in Ngulula, with better market access overall, is now significant (at 1 per cent) across the board. Land is no longer significant. Two out of the three social networks variables affect the probability of a household falling within a particular cluster – prestige networks and network composition captured as diversity of network in terms of sex of network partner. Network size is not significant. Sex heterogeneity of network is negative and significant for HL (10 per cent) and LH households (1 per cent) compared with the base category LL. Thus, sex heterogeneous networks lower the likelihood of being in LH compared to LL. In the context of the rural communities in the study a more sex heterogeneous network is likely to be associated with higher numbers of kin and therefore a more homogeneous network, as traditional social norms prevalent in rural areas mean it is considered to be somewhat inappropriate for members of the opposite sex to conduct friendships etc unless they are related. Therefore, rather than a diverse network in this respect overcoming poor resource access and enabling greater market participation, more kin network partners links to poor outcomes. This lends some support to emerging work suggesting a ‘dark side’ to social capital, with more obligations to kin creating a disincentive to accumulate or acting as a drain on resources (see: DiFalco and Bülte, 2011; Baland et al, 2011).

Influential kin variables (both nearby and in far locations) compared with having no influential kin are significant and positive for all clusters compared with the low-

control, low-marketing base category. This suggests prestige networks are important for command over resources as well as marketing outcomes (though Probit values greater than 2 on these coefficients, and the sex heterogeneity variable coefficient estimates, should be treated with caution as they imply extremely high probability effect, as illustrated by the marginal effects shown). An interesting result here is the lack of significance of the network size variable across all categories.

5.6 Conclusion

The analysis suggests that using social networks variables allows us to capture a more detailed picture of social integration going beyond membership of groups and societies, and helps to uncover not just relationships at a superficial level but the kinds of functions these relationships can potentially and actually do perform in relation to economic activity and, following from this, in achieving welfare outcomes. This can be insightful in circumstances where markets may be thin or under-developed and important institutions are missing. From the analysis so far, clear patterns appear to be emerging around social network ‘types’, welfare outcomes and the characteristics of respondents achieving these outcomes.

The hypotheses set out in the chapter suggest that larger, more diverse networks are associated with positive outcomes, and to a certain extent this appears to be borne out by a very preliminary analysis of the data. Network size and diversity measures (with the exception of age diversity of network partners) do indeed seem to be associated with good outcomes as captured by an index of socio-economic status and by level of crop marketing. People who have networks associated with good outcomes tend to be male, or in male-headed households, married, employers of group labour, with higher levels of education and have not always lived in their current community. Women tend to report more diverse networks in terms of sex of network partners (Appendix E).

Issues surrounding inequality of access to informal social institutions and social networks, and links to vulnerability and exclusion seem especially pertinent; particularly, who has access to which institutions and to what effect? Institutions here are broadly defined to encompass both informal and formal institutions and organisations, including social networks. What does this mean in terms of livelihoods

and welfare? The policy relevance lies in the usefulness of this information in targeting poor people. An issue that deserves attention is what is meant by exclusion. Lack of accessibility of formal institutions often impacts on a whole community in a covariate way, which can help to explain the role of informal mechanisms, for example to improve information flows.

There is a positive correlation between value of crops sold and network size, and with group affiliation, and a negative correlation with higher proportions of kin in the network, suggesting some support of the ‘strength of heterogeneous networks’ hypothesis. Novel and hence more valuable information is purported to be contained in networks where alters have weaker links between each other. A network consisting of many kin members implies not only strong links between the focal individual (respondent or *ego*) and alters, but also strong links between alters. Entry into crop sales markets is also positively correlated with larger social networks. The analysis of correlations between social networks variables and proxies for transactions costs (notably crop storage, bicycle and radio ownership, use of hybrid seed) suggests that higher values of the social networks variables do appear to go hand-in-hand with lower transactions costs. However, the analysis of household clusters according to command over resources and marketing levels, while pointing to the importance of networks in affecting both control over resources and the ability to sell higher volumes of output given resources available, suggests that network size per se is not the factor that differentiates between success or not in terms of crop marketing. How people use their networks is arguably more important. This is an interesting result and will be explored further in subsequent analyses in Chapter 6 using interactive variables. Households participating in markets and those selling higher values of crops appear to be more likely to have prestige networks, captured as kin in an influential position.

.....

Chapter 6 presents empirical results of a multivariate, causal analysis of the potential role of social networks in reducing transaction costs and mediating access to markets. It builds on the preliminary analysis contained here to consider whether the network variables have predictive value in the analysis of returns to social networks, captured as household crop market participation, and whether social networks may help or hinder

access to, or competition in, markets, focusing on trading. Endogeneity is likely to be an issue. This will be explored, and Instrumental Variable Estimation given as a possible solution.

Chapter Six

Empirical Model of Household Crop Marketing and Social Networks

This chapter presents the empirical results of a model of crop market participation, examining the role of social networks in market exchange of smallholder farming households. The overarching question is whether or not social networks help to determine household agricultural commercialisation in settings where there is little variability between households on endowments such as land, labour and farm inputs, in contexts of incomplete or missing markets. Do social networks, by mediating market access, help to explain differences between households?

Descriptive results in Chapter 5 showed sellers are more likely to have bigger social networks in general, with relatively more non-kin members, more mixed-sex networks (though less diverse in terms of age), and be members of multiple village groups and societies. These all go hand-in-hand with operating at higher levels of crop sales. Here, the research builds on this with causal analysis incorporating social networks variables into a model of agricultural commercialisation with transactions costs. Commercialisation is measured by the total value of output sold across all crops, using data collected during both rounds of the household survey⁴³.

The overarching question of the research is: what is the role of social networks in economic life? In the context of market participation this can be focused as what is the role of social networks in potentially reducing transactions costs thus allowing greater participation in markets. This chapter explores the value of being able to access resources through personal social networks. Does this improve the institutional environment for households, leading to better economic outcomes in the form of higher crop sales, through reducing transactions costs and mediating access to markets?

There is a debate about how useful networks are in economic terms. They may serve to help people access information, technical know-how, resources, but there is potentially

⁴³ Maize; Hybrid Maize; Cassava; Millet; Sorghum; Mixed beans; Soybean; Sweet potato; Irish potato; Ground nut; Cabbage; Mango; Banana; Plantain; Squash; Orange; Tomato. See Table

a ‘dark side’. This relates to the role of social networks in engendering obligations between network partners. As a result people may feel burdened by responsibilities and thus adopt strategies that purposefully do not lead to an excess of wealth or income (diFalco and Bülte, 2011; Baland et al, 2011). Social networks can also be a basis for exploitation and social marginalisation (Kabeer, 2000). The research posits that social networks may help reduce transactions costs enabling higher crops sales. The three hypotheses set out in chapter 4 suggest three channels through which social networks could have beneficial outcomes: 1) *Bigger is Better*: larger networks are associated with better outcomes; 2) *There are benefits to having ‘friends in high places’*, defined as links to persons with prestige or influence; and 3) *Heterogeneous networks*: network diversity means greater access to novel information and increased likelihood of being able to draw on diverse resources through networks. These hypotheses will be explored through the econometric models. Interaction terms could potentially highlight how social networks might reinforce other factors, and vice-versa, and these will be explored. A significant challenge in estimating the relationships between social networks and outcomes of smallholder farming households is in establishing direction of causality and correcting for other forms of endogeneity such as selectivity bias.

The chapter first sets out the empirical model of agricultural supply response (Section 6.1), followed by a discussion in Section 6.2 of challenges in estimating the model, focusing on: endogeneity, separability and identification. Section 6.3 sets out the methodology. Instrumental Variable Analysis as a potential solution to some of these challenges is discussed in Section 6.4. The empirical model is estimated using Generalised Method of Moments with instrumental variables and results are presented in Section 6.5. Section 6.6 concludes. Results of estimations assuming exogenous regressors but correcting for potential selectivity bias using i) Heckman selection model; and ii) Censored Tobit are given in Appendix K.

6.1 Empirical Model

The empirical model is based on a basic static model of the agricultural household incorporating transaction costs, following the methodology of Goetz (1992) and later

E2.8 in Appendix E for a breakdown of percentage households selling each crop and average prices. Table E2.9 shows mean values of each crop sold.

extended (see Key, Sadoulet and de Janvry, 2000; Heltberg and Tarp, 2002). Key, Sadoulet and de Janvry (2000) modify and extend the basic static agricultural household model to incorporate transactions costs. In this model, transactions costs are defined as being either *fixed* (costs are fixed regardless of the amount exchanged) or *variable* (costs vary with the amount exchanged).

Following Key, Sadoulet and de Janvry (2000), the empirical estimation is a simple model of agricultural supply response including transactions costs in determining crop market participation. The approach here differs from more mainstream approaches to modelling agricultural commercialisation in that it incorporates social networks in addition to more conventional variables as proxies for transactions costs. The theoretical model is set out in Appendix G. The empirical model is given by the reduced form equation:

$$y_i = \mathbf{X}_i\beta + u_i \quad (6.1)$$

where y_i is the measure of farm household commercialisation, value of aggregate crop sales; u_i is the error term; \mathbf{X}_i is a vector of observed explanatory variables (K, H, S, D, L), with:

K = Farm Capital (including assets, labour, transactions costs)

H = Human Capital

S = Social Networks

D = Household Demographics

L = Location

Social networks potentially impact directly on outcomes, and also indirectly via their effect on transaction costs, see discussion in Chapter 5. In this way, social networks could also proxy for transactions costs. Thus, in the notation above and in the subsequent analysis, social networks variables appear separately.

6.2 Challenges in Estimating the Empirical Model

In estimating the empirical model, analytical issues arise related to the underlying model, to sampling methods, and to the data themselves. This section discusses: i) Endogeneity; ii) Separability; iii) Identification. Other data-related issues including sample survey design, and those arising from the choice of estimation procedure will be examined later.

6.2.1 Endogeneity

An explanatory variable x_j contained in the vector X_i is said to be endogenous if it is correlated with the error term u_i , thus invalidating the orthogonality assumption. Endogeneity usually arises as a result of: i) sample selection bias; ii) simultaneity; iii) measurement error; and/or iv) omitted variables (see Durlauf and Fafchamps, 2005, for a review).

i. Sample Selection Bias

There is potentially a sample selection problem in the research to the extent that there is a restricted, non-random sample of sellers. The observed distribution of income from selling crops is a truncated distribution because the dependent variable for crop income y_i is only observed for those selling crops. People self-select into “sellers” and “non-sellers”. Selectivity bias arises if (6.1) is estimated by Ordinary Least Squares based on the observations for which crop income y_i is available because the dependent variable is censored by an unobserved latent variable, so estimators of the parameters will be inconsistent. Using least squares without correcting for selectivity bias leads to invalid estimates of the parameters for the full sample. The estimated intercept is also biased because means are not zero.

This can be dealt with using two types of econometric model: i) those taking a censored dependent variable, such as a Tobit model; ii) two-stage procedures such as Heckman selection model or treatment-effects models. The assumption underlying the Tobit model is that households are unconstrained. That is, the choice to sell or not to sell is exogenous. If zero values of the dependent variable are a rational choice of households then a censored Tobit model might be the more appropriate model. If this is not

plausible then a Heckman selection model might be more suitable. The rationale behind the Heckman selection model is to estimate market surplus conditional on market participation, where market participation is estimated with a reduced form equation (see also Goetz, 1992). The probability of participation is modelled using Maximum Likelihood Probit to obtain the Inverse Mills Ratio. This is included as a ‘selection term’ in the second stage, modelling the degree of participation given that the household participates at all. This provides consistent, asymptotically efficient estimators for all parameters (Greene, 2003; Maddala, 1983; Heckman, 1990).

ii. Simultaneity

This is a specific type of endogeneity problem, occurring when the explanatory variable is jointly determined with the dependent variable. Regressors included in the outcome model are thus also potentially choice variables. They are therefore likely to be correlated with unobservables contained in the error term. There is theoretical and empirical evidence that this is likely to be true of social networks (and other) variables. Literature on network formation, discussed in Chapter 2, suggests that networks tend to be fluid by nature. Depending on the circumstances, people may switch focus between networks, and may also react to a deterioration of their position in one network by joining another or by creating new networks. Further, are those with certain types of social network better able to participate in markets, or does more participation in markets result in better social networks; that is, networks facilitating successful market participation? Because of inherent simultaneity bias standard regression techniques will confound these two effects. If this is indeed the case then estimates of coefficients on variables will in general be biased leading to over- or underestimation of the true causal effects of the regressors on the outcome of interest.

iii. Measurement Error

In general, it is “assumed ... that the data used to estimate the parameters of our models are true measurements on their theoretical counterparts” (Greene, 2003: 83). In practice, many measurement problems may be present in empirical data. Observing only imperfect measures of variables can be due to: recall problems; ‘noisy’ measures; over- and under-estimation; and because there is simply no tangible measure of the household or individual-level characteristics to be included in the model. Measurement error can introduce at the very least less precision in estimated coefficients and a lower t-statistic,

and in some cases severe biases (Deaton, 1997). Evidence from social networks studies suggests they are inherently prone to measurement error, which is likely to be correlated with individual observable and unobservable characteristics. Studies estimate that as much as fifty per cent of what people report about their own particular interactions is incorrect. However, when the content of reported interactions is examined it tends to relate to their perceptions of stable, long-lasting relationships and not to specific, usually one-off, instances, thus giving a better snapshot of the core of people's social networks without the 'noise' of single, superficial interactions (Wasserman and Faust, 1994:57. See also Brewer, 2000). Other research suggests that what is reported by ego and alter usually corresponds (Marsden, 1990). Self reported crop incomes may also be prone to measurement error. In the research, income data were reported in survey modules administered at the individual level, with respondents (usually household head and spouse) interviewed separately. The reported volumes of crops sold by the household and prices received coincided. It can be assumed therefore that reported crop incomes are reasonably accurate and any measurement error will be non-systematic and therefore not correlated with the error term (see also Moore et al, 2000). The research also includes non-farm income as a regressor, which will be strongly correlated with household incomes, but not with measurement errors in crop incomes.

iv. Omitted variables

An omitted variable is one that might influence the dependent variable but is not included as an explanatory variable in the regression. Ordinary Least Squares (OLS) estimates will be unbiased as long as the omitted variables are uncorrelated with the included explanatory variables. However, if omitted variables are correlated with the included explanatory variables this means there is a compound disturbance term. It includes the unobserved variables as well as the standard random disturbance term. In this case OLS generally produces biased and inconsistent estimates of impact of observed variables. Estimates include not only their true effects but also effects of any correlated unobserved variables, for example unobserved social skills or entrepreneurial ability or other farmer characteristics. Usually such variables are omitted due to data unavailability. Many aspects of social networks are personal and individual choices, for example giving and receiving social support, becoming a member of a group or society. These will therefore depend on personal preferences, which may not be observable. They might also be things that do not vary over time, for example household ability, but

are not included in the model, also called unobserved (neglected) heterogeneity. This is relevant to the analysis here as unobserved heterogeneity might be introduced because the determinants of Social Networks=0 are not known. If this is the case then estimates, for example in the first stage Probit of a Heckman selection model, will be inconsistent, which has no impact on directions of effects or relative effects, but could be problematic for partial effects. This is only a problem where the magnitudes of β_i have meaning (See Wooldridge, 2002: page 471).

There are a number of ways of dealing with endogeneity empirically: If panel data are available, unobserved heterogeneity can be controlled for using random or fixed effects. Experimental approaches can also be used to establish causal links and account for selection bias. However, neither panel data nor experimental approaches are possible with the household survey data collected for the research. Sample selection bias can be corrected by using two-stage or censored models such as Heckman Selection Models (Heckman, 1979) or a Tobit. Other forms of endogeneity may be addressed by incorporating instrumental variables into the model. Instrumental variables can account for time-varying unobservable factors of households by inducing random variation in the endogenous social networks variables. A suitable instrument is one that is correlated with the endogenous explanatory variable, uncorrelated with the error term and does not affect the outcome of interest conditional on the included regressors. That is, it affects the outcome only through the endogenous variable (exclusion restriction).

Estimations include instrumental variables to attempt to control for reverse causality and unobservable household characteristics correlated with the social networks variables and outcomes. Finding suitable instruments is challenging and the chapter returns to this in more detail in section 6.4. The research also attempts to address potential selectivity bias by estimating a Heckman selection model and censored Tobit.

6.2.2 Separability

Do households make market participation and volume decisions simultaneously or sequentially? This is essentially an empirical question. If the decision is made sequentially the analysis needs to be broken down into sub-periods, otherwise the model will be misspecified. This makes it a two-stage decision problem – consumption and production decisions are separable. Most empirical studies of commercialisation assume

separability. First the producer decides whether or not to participate in markets as a seller and once this decision has been made, decides how much to sell, conditional on participation (see: Goetz, 1992; Key, Sadoulet and de Janvry, 2000; Heltberg and Tarp, 2002; Bellemare and Barrett, 2006).

The research is interested in this question insofar as the way one believes producers to make marketing, production and consumption decisions has implications for the way behaviour is modelled. Deciding whether the model should be based on separability or nonseparability conditions can be guided by empirical evidence. A useful summary is given by Sadoulet and de Janvry (1995):

“...If the market is used for a transaction, the household behaves *as if* it were deciding sequentially: production first and consumption/ work after. Production decisions are identical to those of a pure producer. Consumption decisions are affected by the level of income reached in production. For both decisions, *market prices* serve as decision prices. The relevant price is the sale price if the household is a net seller and the purchase price if it is a net buyer.”

Sadoulet and de Janvry (1995:151)

This implies that if there are no market failures and a study is only interested in farm household production, the separability condition holds and removes the need for the household approach. Instead resource allocation is assumed to take place “exactly as proposed in the pure production theory of the firm” (Sadoulet and de Janvry, 1995:159). While non-separability holds for those households not selling crops in market, for those who do, in the presence of market failures, the separability condition breaks down. The household approach is therefore useful as production and consumption must be estimated simultaneously.

Studies on market participation focusing on the separability question tend to look at three mutually exclusive production decisions. The household or individual is either: net buyer; net seller; or neither (autarkic). This thesis is not concerned with the production versus consumption decision so much as how social networks help determine how households can make a living out of farming – looking at the production decision alone. It makes sense therefore to approach the problem from the perspective of seller versus non-seller.

6.2.3 Identification

For a model to be identified the estimation must trace out the correct relationship between the variables. That is, we can actually calculate the coefficients from the observations. Failure to address identification has been a criticism of social capital empirical studies that purport to provide evidence of a role for social capital in socio-economic outcomes (Durlauf, 2002). Failure to identify the parameters of the model is also an effect of endogeneity. Thus, using relevant and valid instruments can help to identify the parameters of the model.

When modelling crop market participation decisions using a two-stage procedure in particular, such as a Heckman selection model, one must consider how to correctly identify the selection and outcome equations. In terms of identifying the selection (first stage) equation of the Heckman selection model there should be at least one continuous variable present in the selection equation that does not appear in the subsequent outcome equation (Deaton, 1997). That is, a variable that affects the decision to sell but not how much. Key et al (2000) distinguish between Fixed Transaction Costs (FTCs) and Variable Transaction Costs (VTCs) as determinants of crop market participation. Both types of transactions costs affect market participation but only VTCs affect the amount sold decision, conditional on participation. So FTCs are omitted from the outcome equation, and used econometrically to identify market participation (see also Heltberg and Tarp, 2002). The Inverse Mills ratio calculated in the first stage of the estimation is a nonlinear function of the regressors in the first-stage Probit model, so by including it in the second stage equation means the latter is identified but only if the assumption of normality in the Probit model holds. Therefore to make clearer the source of identification of the second stage equation one needs ideally to include at least one other variable in the second stage that is not present in the first stage.

The next section describes the methodologies for the multivariate analyses, first noting sampling weights used to take account of complex survey design. The choice of models is set out briefly, followed by a discussion of variables: dependent variables, exogenous regressors, endogenous regressors and instruments.

6.3 Methodology

6.3.1 Complex survey design⁴⁴:

Villages were purposively sampled, with households randomly sampled within each village so that each sample is representative of its village. Ignoring sampling design and assuming simple random sampling will likely underestimate standard errors, possibly leading to results that seem to be statistically significant, when in fact, they are not. The survey aimed to be representative of the areas surveyed, so sample weights based on the reciprocal of the ratio of sample size to village size are used and are given in Table 6.1 (see also the explanation in Chapter 3).

Table 6.1 Sampling weights

Community	Weight
Kabila	5.56
Ngulula	2.40
Lufubu	1.08

6.4.2 Choice of models

Two-stage procedures such as the Heckman Selection Model (Heckman, 1979), also called a generalised Tobit, can be used to estimate both market participation and the value of sales given market participation, thus addressing potential selectivity bias. The first step is to model selection into sellers and non-sellers (a discrete, binary decision). Second, model the continuous decision of how much to sell, conditional on participation. Quantity sold is therefore analysed for the sub-sample of selling households.

The Heckman model assumes lack of data on the subsample of non-seller households. However, in this research, even though the dependent variable ‘value of crops sold’ is not observed (is equal to zero) for the sub-sample of the data not selling crops during the survey period, the regressors are observed for these farm households. This suggests that a censored Tobit model might be an appropriate estimation procedure, rather than a selection model (Greene, 2003; Maddala, 1983). Here, it is assumed that data on the dependent variable is unavailable (or limited) but not data on the regressors. The

⁴⁴ A Chow test to test whether the parameter estimates are the same whether we pool the data or estimate the model for each village separately is not possible due to degrees of freedom problems. A categorical variable for community is sufficient.

assumption underlying the Tobit is that households are unconstrained. That is, in contrast to the Heckman model, the choice to sell or not to sell is exogenous and zero values of the dependent variable are a rational choice of households

Generalised Method of Moments (GMM) is a semi-parametric method that allows a relaxation of restrictive, parametric assumptions and generates estimators that are “robust to some variations in the underlying data generating process” (Greene, 2003: 525). In contrast to standard IV estimation GMM is efficient when ‘heteroskedasticity of unknown form’ is present, the effect of which is inconsistent estimates of standard errors rendering invalid diagnostic tests for validity (endogeneity) and relevance (see Baum et al 2003; 2007). Under GMM, test statistics are heteroskedasticity-robust. Unlike the Heckman and Tobit specifications, here selection is assumed to be exogenous. This assumption can be tested by comparing the results of the selection model with those of the GMM.

Choice of the most appropriate model will be made on consideration of the need for and success in addressing endogeneity concerns in the estimation, and minimisation of bias. This includes: significance of the inverse mills ratio in the second stage of the Heckman selection model; the ability to instrument for potential endogenous regressors in combination with dealing with complex survey design; general properties of the estimation procedures.

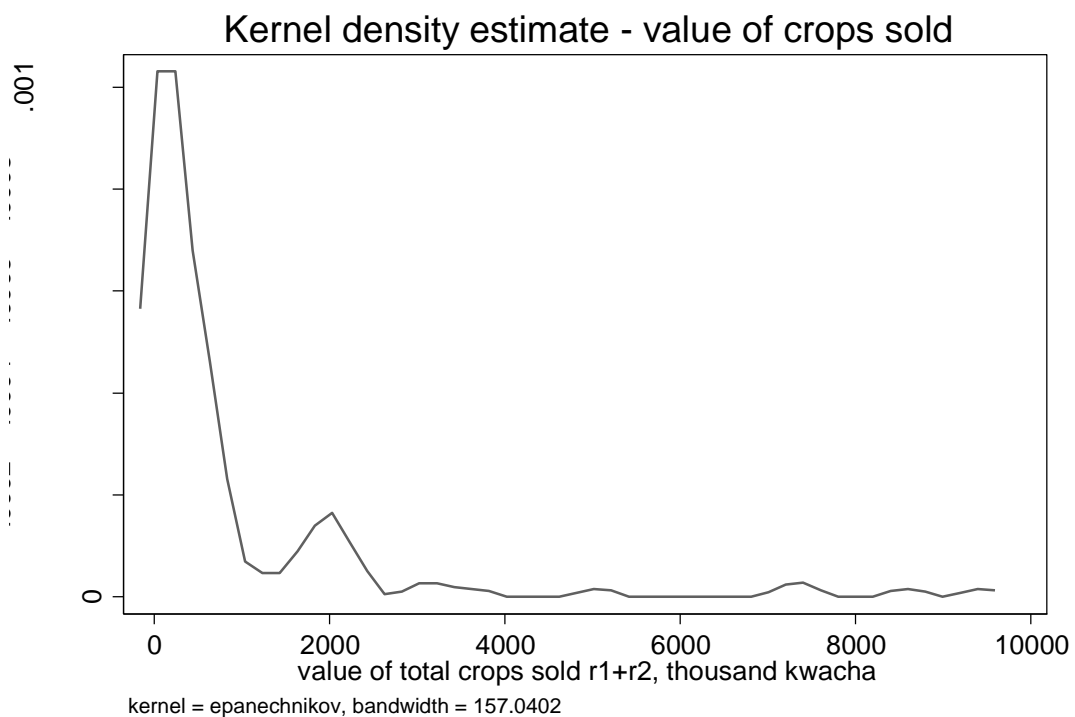
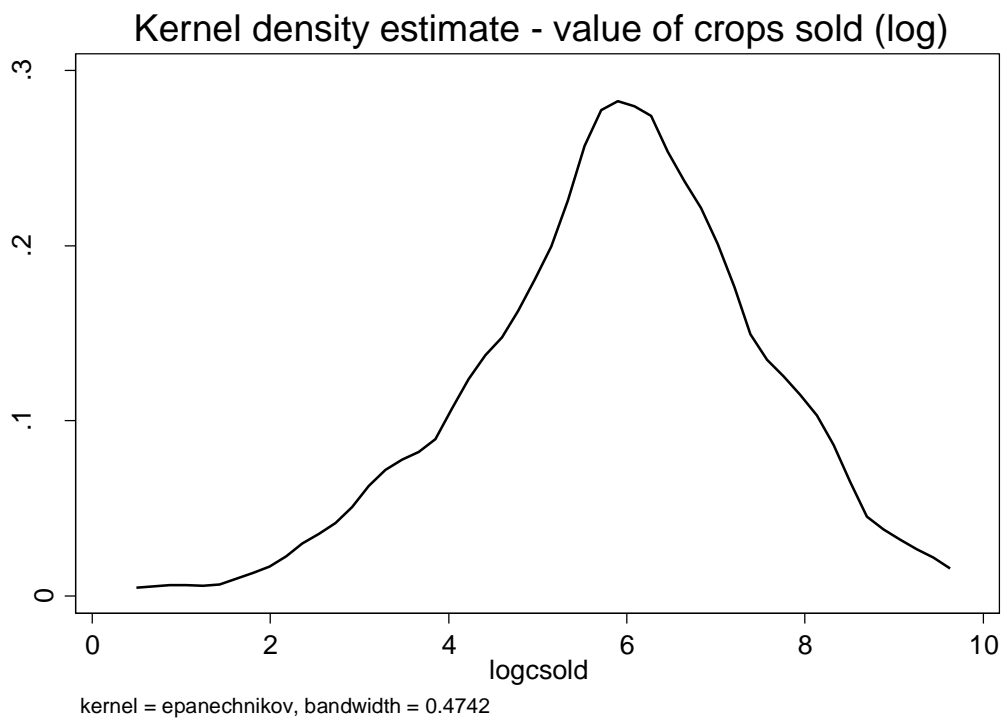
6.3.3 Choice of variables

Dependent Variables

The choice of indicator of household level commercialisation is largely an empirical question (see Chapter 1 of this thesis and also Leavy and Poulton, 2007, for a discussion). Following Heltberg and Tarp (2002), and in line with previous studies of smallholder commercialisation in Northern Province discussed in Chapter 5, the models in this chapter use y_i equal to the log of the aggregate value of crop sales as the endogenous (dependent) variable. Quantities of crops grown and sold, and prices received by the household were collected where possible from more than one family member and responses checked carefully for inconsistencies. Because household

portfolios consist of multiple crops, quantities sold cannot be aggregated across different crops in any meaningful way unless converted to values. This was done using market prices at the time of the two rounds of surveys, collected via price questionnaires administered at the community level. The household surveys also reported price data, but there were many gaps. The community level price data was checked against prices reported by the households and found to be consistent. These prices act as implicit weights. This approach, however, tells us nothing about underlying causal mechanisms for cultivation decisions (Heltberg and Tarp, 2002). In the selection stage of the Heckman two-stage model (Appendix K), the dependent variable is a dummy variable denoting whether or not the household sold crops at all.

The mean output of sellers compared with non-sellers, in terms of value of crops grown, is considerably higher (Table E2.7 in Appendix E). Sellers on average grow crops to a value of 2 million kwacha, with a maximum of 10.9 million kwacha, compared to an average crop value of 174 thousand kwacha for non-sellers, with a maximum of 1.5 million kwacha. Kernel density estimates for the value of crops sold variable suggest wide variability in crop incomes across households in the sample who sold over the survey period. These are given in Figures 6.1 and 6.2.

Figure 6.1: Kernel Density Estimate – Value of crops sold**Figure 6.2: Kernel Density Estimate – Value of crops sold (log)**

Regressors

Determinants of household crop marketing can be grouped as: household demographic characteristics; human capital; farm capital; social networks; location:

<i>Household Demographics:</i>	Household size and dependency ratio ⁴⁵ , sex and age of household head
<i>Human Capital:</i>	Maximum level of education in the household
<i>Farm Capital: FTCs</i>	Ownership of radio; Access to crop storage facilities; Ownership of bicycle
<i>Farm Capital: VTCs</i>	Labour dummies: household members work as farm labourers, household is an employer of farm labour; Non-farm income ('000 kwacha), Value of fertiliser used (Kwacha), Proportion of land cultivated ⁴⁶ , dummy variable for use of hybrid seed, Large livestock ownership (Number of animals)
<i>Social Networks:</i>	Network size, network diversity (by age and sex of network partners), membership of formal and informal groups, proportion of total network members who are kin, kin in influential position and their location, proportion of links that are multiplex; average content multiplexity of network partners; sum of resources accessed across network. These are all calculated as the maximum value for the household
<i>Location:</i>	categorical variable for community

⁴⁵ The household dependency ratio is calculated as the number of people in a household aged 15 or less plus those aged over 60 divided by the number aged 15-60.

⁴⁶ Proportion of land cultivated is used rather than total area cropped due to lack of variability in the latter variable. Out of 137 households, 34 households have landholdings of less than 2.5 hectares. There are some outliers – five households report between 30 and 50 hectares and one household 75 hectares of land. Forty five households report landholdings of exactly 10 hectares. These are all farmers in Lufubu Resettlement Scheme, reflecting the way land is allocated as a standard parcel of land. The standard deviations reflect differences in variability by location. Ngulula has a mean landholding of 9 hectares with a standard deviation of 11.04 for the full sample, and without outliers a mean landholding of 7.14 hectares, standard deviation 7.32. Kabila's mean landholding is 4.8 hectares, standard deviation 12.7, and without outliers 2.8 hectares on average with a standard deviation of 4.5. Lufubu Resettlement Scheme's average plot size is 12.5 hectares, standard deviation 6.45, and without outliers has a mean landholding of 11.5 hectares with a standard deviation of 3.45.

Following Key et al (2000), farm capital variables are subdivided into Fixed and Variable Transaction Costs, under the hypothesis that household relationships to the market can differ because of market transaction costs. Variable transaction costs include factors such as distances and transport. Fixed transaction costs include information variables. Transaction costs are at best only partly observable so proxies are used. Herbicide, pesticide and other chemical application to fields rarely appear in the sample and are therefore omitted from the estimation due to lack of observations. Because of cash constraints these are also likely to take place only when there is a problem, thus reactively rather than routinely, and so will be endogenous and have a negative correlation with crop output.

Social networks variables include: network size, network diversity (by age and sex of network partners), membership of formal and informal groups, proportion of total network members who are kin, kin in influential position and their location, network multiplexity measured as: number of multiplex ties, the average number of ties per alter, the sum of resources accessed/relationships across network. These are all calculated as the maximum value for the household. A variable is also included for group affiliation. Group membership as a proxy for social capital has been linked to outcomes, and the empirical literature provides examples of both positive and negative effects (see: Coleman 1990; Burt 1992; Massey and Espinoza, 1997, among others). Positive effects of group membership include: generating and pooling knowledge; providing information about reputation and reliability of agents; establishing institutional structures (norms and rules) for collective action (see Collier, 1998). Evidence of direct links between group membership and per capita household income is provided by: Haddad and Maluccio (2002), using longitudinal data for Kwazulu-Natal; Narayan and Pritchett (1999) for Tanzania, instrumenting for group membership with trust, (although Haddad and Maluccio find no evidence that trust is “contemporaneously important for income generation”); Grootaert, (1999); Grootaert et al, (1999). By contrast, groups may perpetuate hierarchies and inequalities to the detriment of its members, or non-members, and members may be bound to obligations with negative economic consequences.

Interaction terms are also included in the model to allow for shifts in the slope of agricultural supply function according to various characteristics, positing that discrete

jumps in supply are possible for those with access to information and/or greater potential to interpret and use information that also have larger social networks. This is captured with an interaction term between: i) radio ownership and network size; and ii) education level and network size. Table 6.2 sets out the variables used in the models and basic descriptive statistics.

Table 6.2 Descriptive Statistics: Dependent Variables and Potential Regressors.

N=136.	Unit	Mean/ dist	Std. Dev.
Dependent Variables:			
Market participation (<i>market</i>)	1=Yes; 0=No (% yes)	83.82	..
Total annual sales agricultural food crops (ln) (<i>logcsold</i>)	Thousand Kwacha (ln)	5.87	1.55
Total annual sales agricultural food crops (<i>csold</i>)	Thousand Kwacha	805.98	1541.69
Regressors:			
Farm Capital: Assets and labour			
<i>Fixed transaction costs:</i>			
Ownership of radio (<i>radio</i>)	1=Yes; 0=No (% yes)	44.85	..
Access to crop storage facilities (<i>store</i>)	1=Yes; 0=No (% yes)	30.88	..
Ownership of bicycle (<i>bike</i>)	1=Yes; 0=No (% yes)	70.59	..
<i>Variable transaction costs:</i>			
HH member is farm labourer (<i>egoemp</i>)	1=Yes; 0=No (% yes)	65.44	..
HH Employs farm labour (<i>egoboss</i>)	1=Yes; 0=No (% yes)	77.21	..
Non-farm income ('000 kwacha) (<i>nfy2</i>)	Number	34.29	83.57
Value of fertiliser used ('000 kwacha) (<i>fertval</i>)	Number	57.26	146.23
Proportion of land cultivated (<i>propland</i>)	Share	0.35	0.35
Uses hybrid seed (<i>hyseed</i>)	1=Yes; 0=No (% yes)	33.83	..
Large livestock ownership (<i>livelarg</i>)	Number	2.25	4.40
Human Capital (<i>Variable transaction costs</i>)			
Maximum education level in household (<i>maxedhh</i>)	0 = Up to Grade 4	8.09	..
[also: Dummy variable max HH education is secondary and above (<i>education</i>)]	1 = Grades 5-7	42.65	..
	2 = Secondary/ College	49.26	..
Social Networks			
<i>Network Diversity/Composition</i>			
Diversity of network in terms of sex of alters (<i>agehet</i>)	Index of Qualitative Variation of alters' sex	0.23	0.18
Diversity of network in terms of age of alters (<i>IQVst2</i>)	St.dev of alter ages in years	11.96	4.81
Kin as proportion of network (<i>propkin</i>)	Proportion	0.4	0.27
<i>Size</i>			
Group Membership (dummy) (<i>affil2</i>)	1=Yes; 0=No (% yes)	37.50	..
Group Membership (No of groups) (<i>affil</i>)	Number	2.44	1.59
Network Size (<i>size</i>)	Number	11.35	6.21
Network Size (log) (<i>logsize</i>)	Number	2.28	0.57
<i>Functional Diversity</i>			
Proportion of ties that are multiplex (<i>propmult</i>)	Proportion	0.19	0.13
Average content multiplexity of alters (<i>multcont</i>)	No. of 'network functions' alter appears in	3.49	0.85
<i>Resource access – instrumental networks</i>			
Sum of resources accessed across network (<i>ressum</i>)	No of resource types accessed by ego through alters	13.27	3.53
<i>Prestige Networks</i>			
Influential Kin dummy (<i>kininf</i>)	1=Yes; 0=No (% yes)	70.6	..
Household Demographics			
Sex of household head (<i>hh_sex</i>)	1 = Female 0= Male (% female)	17.65	..
Age of household head (<i>hh_age</i>)	Years	44.50	13.58
Household size (<i>hh_size</i>)	Number	7.01	3.12
Household dependency ratio (<i>depend</i>)	Number	0.51	0.21
Location			
Categorical variable for location (<i>comid</i>)	1 = Kabila	27.21	..
	2 = Ngulula	29.41	..
	3 = Lufubu	43.38	..
Interaction Terms			
Network size*Radio Ownership			
Network size*education			

The next section discusses endogeneity in the proposed independent variables.

6.3.4 Regressor Endogeneity

Addressing endogeneity is the primary challenge in modelling the relationships in this study. It would be unusual in such a study for all of the variables in the cross-sectional model to be exogenous, that is, regressors statistically independent of the residuals, a standard assumption in regression analysis. Consequently, parameter estimates will be biased and inconsistent. Most elements of people's lives, their resource access, socio-economic characteristics, are going to be related to each other in some way. People's lives are complex, relationships between variables will be mutually reinforcing and it would be difficult, if not impossible, to disentangle effects and establish a direction of causality. For example, farm inputs such as fertiliser and other chemicals, tend to be endogenous, their use determined in part by income and also by unobserved factors such as shocks.

To model relationships econometrically it is necessary to establish causal links and therefore potential endogeneity needs to be addressed. If not, parameter estimates are likely to be biased and inconsistent. The likely presence of endogeneity between outcomes and regressors, notably, social networks, has been discussed briefly. Of the regressors described in Table 6.2 above, the following are likely to be exogenous and therefore the direction of causality clear: education, sex and age of household head, community, non-farm income. These are all likely to determine output/ outcomes.

Social networks variables related to attributes of network partners, network multiplexity, age and sex heterogeneity, proportion of network partners who are kin and prestige networks (dummy and location of influential kin), are assumed to be exogenous as they are largely beyond the household's control or predate substantially the time-frame of the model.

The exogeneity of other variables are more debatable. Labour and other purchased inputs (fertiliser, hybrid seed) are determined to a great extent by the availability of cash and therefore value of the output that the household manages to sell, while at the same

time the use of purchased inputs potentially increases production thereby increasing crop income. However, within the timeframe of the model one would argue that current labour, seed and fertiliser use is determined by past crop income. Therefore, these variables are assumed to be exogenous. Household size is also arguably endogenous. More successful households may be more likely to attract more people, especially extended family members/ kin. This can have the effect of increasing the household's available labour and there could be social benefits to having 'many followers' as discussed in Chapter 1. Conversely, it may have the effect of increasing the household's dependency ratio and thus have socio-economic costs – for example, more mouths to feed. Such obligations to kin may pressure households to increase incomes, but could also act as a disincentive to accumulate (see also diFalco and Bulte, 2011).

The proportion of land owned that is cropped is potentially endogenous. Land is relatively accessible in Northern Province, allocated by the Chief of an area to Headmen, who in turn allocate land to villagers. So on the whole households do not have a choice of landholding. However, if someone is able to cultivate more land then they can request more land from the Headman and therefore may receive a further allocation. This implies that more successful farmers will be those cultivating a greater proportion of their total land allocation, which in turn feeds back into their success as a farmer creating a 'virtuous' circle, though there will be a point at which additional land allocation will mean the more successful farmer may then be cultivating a smaller proportion of land than someone relatively less successful, so the relationship will not be linear. It could also plausibly be argued that as cultivated area is usually determined at the start of the season it is reasonable to assume that it is exogenous to the value of output sold for that season, which occurs at some point in the future.

Radio ownership is also potentially endogenous as households with higher crop incomes are more likely to be able to afford and hence have bought a radio. Generally, asset ownership tends to reflect accumulated past income. Radio ownership also potentially provides access to more information (on markets, prices etc) thereby helping to generate higher incomes. In their model of agricultural supply response in Mozambique, Heltberg and Tarp (2002) use radio ownership to proxy for fixed transaction costs (access to information). They do not instrument for it although they do note that while significant and related positively to food crop market participation, direction of

causality is indeterminate. Looking at the descriptive statistics, those participating in crop markets are just as likely to own a radio as not: fifty per cent of households selling crops have a working radio. Given the impossibility of finding an instrument for radio in the dataset, it will be assumed that radio ownership predates the period of the research, in that data on working radio ownership was gathered in round 1 of the survey whereas data for the crops sold variable was gathered in both round 1, referring to the previous 6 month recall period, and round 2, referring to the period since the initial survey, and aggregated.

Finally, endogeneity of social networks: One of the major criticisms of work on social capital is its failure to adequately address endogeneity.⁴⁷ Two of the social networks variables may be endogenous – network size and household membership of village level groups and societies. Network size is potentially endogenous because i) those engaging with markets are more likely to interact with more people thus increasing their chances of having a larger network and ii) ‘successful’ people are more likely to attract others into their network. However, one could also reasonably argue that social networks are, on the whole, established and embedded relationships and as such are likely to pre-date substantially the time-frame of the research, and have existed before commercialisation of the farm household began. Group affiliation may be endogenous to income given group membership usually involves paying a subscription or fee of some kind, hence income will determine whether or not a household can actually afford to be in a group. Further, many groups are farmer/ cultivation groups and therefore level of production is likely to have an effect on whether or not a smallholder farming household member decides to become a group member. Assuming there are commercial benefits to being a member of a farming group (through increased access to information, labour etc) then group membership may lead to higher crop incomes.

To summarise, network size, proportion of land cropped and group affiliation are therefore assumed to be endogenous and the research attempts to find suitable instruments for these potentially endogenous regressors. The dataset, despite its richness, holds limited scope for instruments that are appropriate econometrically but also plausible in a real world sense. The next section sets out conditions for instrument

validity and relevance, and associated statistical tests, followed by a description of potential instruments in the dataset.

6.4 Instrumental Variable Analysis

Instrumental variable analysis provides a general solution to problems of omitted variable bias, measurement error and endogenous explanatory variables. Observed explanatory variables believed to be correlated with the random disturbance term are replaced by their predicted values, thus removing correlations between observed explanatory variables and the disturbance term. These predicted values are obtained from a first stage regression of the endogenous variables on a set of observable variables (the instruments) that are not correlated with the random disturbance term.

6.4.1 Instrument Validity

Instruments must be both relevant and valid. For an instrument to be relevant it must be sufficiently correlated with the observed endogenous regressors to provide reasonably accurate predicted values. Thus the partial correlation between the endogenous regressor and the instrumental variable should be nonzero. If it is nonzero but very small then the instruments are possibly weak. Consequently they will be poor predictors of the endogenous variable and obtained predicted values will have very little variation (see Murray, 2006). For an instrument to be valid it must be orthogonal to the errors, that is, uncorrelated with the disturbance term of relation being estimated. Further, instruments should not already appear in the relationship being estimated and there must be at least one instrumental variable for every observed endogenous explanatory variable in the relation being estimated.

Choice of instruments is one of the main challenges in empirical work, and in many ways is an art. The next section sets out tests for instrument relevance and validity (exogeneity), followed by a discussion of the choice of instrumental variables in the research.

⁴⁷ See Durlauf's (2002) comprehensive exploration of estimation problems in empirical work on

6.4.2 Diagnostic Tests

Once suitable instruments are found, one needs to test for instrument relevance (F- test) and validity (exogeneity). Also of concern is whether or not instruments are only weakly correlated with the potentially endogenous variable, which will result in large standard errors and 2SLS results biased towards the OLS estimator (see Murray, 2006). Tests for instrument relevance, weak instruments and validity are:

- i) F-test of joint significance of excluded instruments in the first stage IV regressions to assess instrument correlation with endogenous regressors. H0: variables jointly=0;
- ii) F-test of the equation in the second stage. H0: variables jointly=0;
- iii) Cragg-Donald weak identification test (see Stock and Yogo, 2002). H0: equation weakly identified. This also provides a test statistic for weak instruments;
- iv) In over-identified models, the Hansen J-statistic has a χ^2 distribution under the joint null hypothesis is that the instruments are valid instruments, uncorrelated with the error term, and excluded instruments are correctly excluded from the estimated equation i.e. instruments are orthogonal to the error. This is also a test of over-identifying restrictions. H0: instruments are valid i.e. orthogonal to the error term; a significant statistic indicates that one or more of our instruments are not valid (assuming that the model is otherwise correctly specified; see also Sargan test statistic). Estimates should be treated with caution as either the model is misspecified or some instruments are invalid.
- v) Hausman test for exogeneity. H0: exogeneity.

The underlying dependent variable equation (6.1) can be rewritten to separate out variables contained within the vector X_i as an additional regressor(s) (Z), suspected to be endogenous:

$$y_i = \mathbf{X}_i\boldsymbol{\beta} + \gamma Z_i e_i \quad (6.2)$$

X is the vector of exogenous explanatory variables, dimensions of X is $(N \times k)$

Note that:

$$Z_i = W_i' \nu + u_i \quad (6.3)$$

W_i is a vector of exogenous variables including a set of identifying instruments, dimensions (Nxm). When $k=m$ the model is just identified; where $k>m$ the model is over-identified.

Assuming normality, we can express the random disturbance e_i as a function of the randomness in (6.3) and some other term:

$$e_i = \psi u_i + v_i \quad (6.4)$$

where ψ is a constant and v_i are normally distributed errors independent of u_i .

For instruments to be valid the following conditions must be met:

Relevance: $\text{Cov}(Z_i, X_i) \neq 0$ or $\text{plim} 1/N(Z'X) = \sum_{zx} \neq 0$

Exogeneity: $\text{Cov}(Z_i, \epsilon_i) = 0$ or $\text{plim} 1/N(Z'\epsilon) = 0$

These conditions ensure that the part of X that is correlated with Z contains only 'good' variation. However, while relevance is testable, as are over-identifying restrictions, exogeneity is not fully testable so it is necessary to argue plausibility.

The exogeneity hypothesis implies that the two sources of randomness are independent and this forms the basis for testing instrument validity. This ultimately reduces to a restriction $\psi=0$. This can be implemented by first estimating 6.3 and obtain the residuals \hat{u}_i , then estimate a second stage equation:

$$y_i^* = \mathbf{x}_i' \boldsymbol{\beta} + \gamma Z_i + \psi \hat{u}_i + v_i \quad (6.5)$$

A t-test is then constructed on the restriction $\psi=0$. If we fail to reject this null hypothesis then the exogeneity restriction is upheld. However, if the t-test is significant this confirms there is an endogenous variable present and instrumentation is required.

Returning to equation (6.3), the instrumenting equation is:

$$Z_i = W_i' \nu + u_i$$

Where W_i is a vector of exogenous variables including a set of identifying instruments.

The predicted value of the endogenous independent variable from the instrumenting equation, $Z^*=g(W)$, is used as a regressor in the main model.

6.4.3 Potential instruments

There are a number of options for valid and relevant instruments. If panel data are available then using lagged values of the regressor may address endogeneity, although some lagged variables may still be endogenous. Behaviour or outcomes in the past may be correlated with current outcomes. For example, ownership of radio one year ago is likely to be highly correlated with crop sales because information received by radio may have informed crop planting and marketing decisions today. This links to unobserved heterogeneity (Murray, 2006).

If appropriate lagged variables are not available then in general to find suitable instruments one needs to use a priori reasoning to choose instruments that make sense from a statistical point of view supported by theory and other evidence (quantitative or qualitative). Ideal instruments successfully capture an exogenous source of variation in Y that either derives from true or pseudo-randomisation, usually predetermined or fixed characteristics out of the control of the household. Using qualitative and quantitative materials together, “essential in untangling causality in social science” (Durlauf, 2002: 270), may help to develop the case for using a specific instrument or for asserting direction of causality. Given the often joint determination of social networks and socioeconomic outcomes, qualitative evidence can help to determine the conditions under which there are social network effects on outcomes. During fieldwork for this research the qualitative exercise described in Appendix C, with its discussion of determinants of successful commercial farming, gives clues to the role played by social

networks. This qualitative data suggests people ascribe a causal link between having particular network characteristics and outcomes: “you need people, no matter what” and “the famous always get their work done”⁴⁸. The models in this chapter allow this to be examined quantitatively and potentially make robust empirical claims about linkages.

Recall, instruments need to be independent of the residual (that is, exogenous or at least predetermined) and not impact directly on the value of crop sold, only affecting the dependent variable through their effect on the endogenous regressor. Ideal instruments when estimating relationships related to agriculture and rural livelihoods include variables for distances, weather, and prices. However, given that there are only three locations, the community identifier variable acts as a proxy for elements measured at the community level. So the usual challenge in finding possible instruments, variables exogenous within timeframe of the model, is exacerbated by the nature of the sample. Correlation coefficients of the potential instruments, endogenous regressors and the outcome variables help to identify potentially weak instruments. The correlation matrix for potential instruments in the research is given in table 6.3 below.

An instrument type used in relation to social capital measures is the non-self cluster mean. These are values on a given endogenous variable calculated as mean over all other households in a community, that is, without the household’s score. Thus the variable is a community average that varies at the household level (Christiaensen and Alderman, 2001). Recall, measurement error in social capital and, analogously, social networks variables, is likely to be correlated with individual observable and unobservable characteristics. However, the averages of social capital (or social networks) over the other members of the community are unlikely to be correlated with individual observable and unobservable characteristics. They are likely to be correlated with social networks “as they are the product of the same social context in which the individual lives” (see d’Hombres et al, 2010). As long as there appears to be a strong correlation between the instrument and endogenous variables in the first stage regression then we can assume the effect is not in fact a local average treatment effect.

⁴⁸ DPhil Fieldwork, Northern Province Zambia, 2002-2003

Table 6.3 Correlation Matrix dependent variable, endogenous regressors, instrumental variables

	Log csold	propland	Log size	Logsize *ed2	Logsize* rad	affil	areatot	IQVst2	Log meansize	Log yearcomm	iqvrad	iqved	fire	water	wild	dom time
logcsold	1															
propland	-0.07	1														
logsize	0.11	-0.09	1													
logsizeed2	0.3	-0.07	0.33	1												
logsize*rad	0.47	-0.07	0.33	0.35	1											
affil	0.29	-0.15	0.46	0.24	0.28	1										
areatot	0.29	-0.33	0.23	0.27	0.22	0.15	1									
IQVst2	0.02	0.06	0.71	0.25	0.12	0.29	0.08	1								
logmeansize	0.01	-0.42	0.22	0.17	0.19	0.32	0.26	-0.08	1							
logyearcomm	0.14	0.3	-0.16	-0.25	-0.12	-0.02	-0.22	-0.01	-0.49	1						
iqvrad	0.38	-0.03	0.41	0.32	0.81	0.3	0.16	0.45	0.13	-0.05	1					
iqved	0.17	-0.01	0.63	0.61	0.19	0.3	0.16	0.86	0.01	-0.08	0.46	1				
fire	0.09	0.17	0.01	0.12	-0.05	-0.03	0.02	0.07	-0.32	0.32	0	0.11	1			
water	0.18	0.02	-0.01	0.04	-0.01	0.11	0.08	-0.02	-0.03	0.04	-0.01	0.01	0.39	1		
wild	-0.06	-0.02	0.04	0.1	-0.02	0.04	0.08	0.03	0.01	0.22	-0.04	0.05	0.53	0.43	1	
domtime	0.03	0.05	0.03	0.11	-0.03	0.04	0.08	0.03	-0.1	0.26	-0.03	0.07	0.76	0.64	0.93	1

Notes: *logcsold*=log of value of crops sold; *propland*=proportion of land that is cultivated; *logsize*=log of household network size; *logsizeed2*=interaction term between log of network size and household secondary education dummy; *logsize*rad*= interaction term between *logsize* and *radio*; *affil*= number of household group memberships; *areatot*=total land area farmed; *IQVst2*= sex heterogeneity of network; *logmeansize*=log of non-self cluster mean of network size; *logyearcomm*= log of years in community; *iqvrad*= interaction term between *IQVst2* and *radio*; *iqved*= interaction term between *IQVst2* and *education*; *fire*= household time spent collecting firewood, average hours daily; *water*= household time spent collecting water, average hours daily; *wild*= household time spent collecting wild food, average hours daily; *domtime*: time spent on collecting firewood, water and wildfood, average hours daily.

Exogenous social networks variables may be used to instrument for endogenous SN measures, given that some social network features may determine network size directly, for example diversity of networks, and outcomes only indirectly through their effect on network size. For social network size a potential instrument is sex heterogeneity of network partners. Intuitively, more mixed networks will bring the household into contact with more people. Further, this variable does not correlate with the value of crop sold variable but does correlate highly (correlation coefficient of 0.7. Table 6.3) with the (ln) network size variable, suggesting it might be a suitable instrument for network size. Any effect of this variable on crop output sold is likely to be through overall network size.

A second potential instrument for the endogenous social network variable and for group affiliation is church membership. As an instrument for social networks and/or group membership this is likely to significantly predate the research. However, all but one household in the sample profess to be members of a church, with the majority (130 households) of Christian denomination, suggesting there is not enough variability in the variable to consider inclusion in the estimation. Instead, tithe payments could possibly stand as a proxy to denote active church membership, taken as a dummy variable because it is not clear the extent to which tithe payment is related to income. However, church membership may not be a valid instrument due to potentially a direct impact on outcomes through the work ethic (for example, see Weber's thesis linking a protestant ethic and a capitalist spirit: 1904/5, 1930; subsequent work by Tawney, 1925, as well as more recent studies relating religion to socio-economic outcomes, such as Barro and McCleary, 2003).

Following Haddad and Maluccio (2002), a further instrument for the group affiliation variable (measured as the number of groups the household is present in) is time spent in area (log). The rationale is that spending longer in an area increases exposure to the possibility of joining community groups, and thus increases likelihood of group membership.

There are potentially two instruments for proportion of land cropped: i) total land area owned – this is related to the proportion of land cropped, but not directly to crop sales (see Isham, 2002). The Pearson correlation coefficient suggests that this variable is

moderately correlated with the land proportion variable and only weakly correlated with the value of output sold (-0.42 and 0.3 respectively); and ii) average time spent per day collecting firewood (in minutes): households who spend more time collecting firewood are likely to have less family labour available to cultivate a larger proportion of their landholding (see Brück, 2004). The Pearson correlation coefficients between this variable and the land and output variables are 0.2 and 0.09 respectively, suggesting little or no correlation with value of output sold, and just a weak correlation with proportion of land cropped. This instrument is potentially weak and could prove problematic in the estimation. The model was estimated using as an IV first the firewood variable, and then in its place an aggregate measure that is the sum of time taken collecting firewood, water and wild food (however, time spent collecting water and wild food is likely to be endogenous to household size for example, as a larger family means greater demand, and in turn more time spent collecting). GMM results are reported in Appendix H and diagnostic tests in Appendix I from estimations that ‘rotate’ the instruments. These show that on balance omitting these instruments gives slightly better diagnostic test results in the first stage for the endogenous land variable.

Instrument sets for the potentially endogenous regressors are given in Table 6.4:

Table 6.4 Instrument Sets for Potentially Endogenous Regressors

Endogenous regressor	Instrument
Network size (<i>size</i>)	Diversity of network in terms of sex of alters (<i>IQVst2</i>) Non-self cluster mean network size (<i>logmeansize</i>)
Land cropped (<i>propland</i>)	Total land holding (<i>areatot</i>) Time spent collecting firewood/water/wildfood (<i>fire; water; wild; domtime</i>)
Group affiliation (dummy; number of groups) (<i>affil2; affil</i>)	Logarithm of time spent in area (<i>logyearcomm</i>) Church membership (<i>tithe</i>)
Network size*education	<i>IQVst2*education (iqved)</i>
Network size*radio	<i>IQVst2*radio (iqvrad)</i>

The next section present results and analysis of the empirical model of the relationship between social networks and crop market participation estimated by general method of moments (GMM).

6.5 Results and Analysis

The model was estimated using a Heckman Selection Model. The Inverse Mills Ratio (IMR) is not significant in the second stage regressions, suggesting there is no selection bias in the sample. However, just 21 out of 136 households reported no crop sales over the survey period, which may explain the selection term's lack of significance. Further, it is not possible to achieve convergence when estimating either the Heckman or the censored Tobit model with the data using survey weights and instrumental variables. Therefore the Heckman and Tobit models have been estimated under the assumption of exogeneity. Results are given in Appendix K.

6.5.1 GMM estimation

The model is estimated using Generalised Methods of Moments (GMM) estimation. This gives efficient estimates of the coefficients as well as consistent estimates of the standard errors because of its use of the optimal weighting matrix, the over-identifying restrictions of the model, and it relaxes the i.i.d. assumption. GMM is the most efficient estimator within the class of instrumental variables estimators.

Variables are selected in order to test the three hypotheses set out in Chapter 4 relating social networks to outcomes. Recall:

Hypothesis 1: Bigger is better

A bigger network is better for ego's "individual goal attainment" (Bourdieu, 1986; and Burt, 1992).

Hypothesis 2: There are benefits to having 'friends in high places'

Positive social capital results from accessing network members (alters) with high prestige (Lin, 2001).

Hypothesis 3: The strength of heterogeneous networks

Those with heterogeneous networks are better able to attain their goals (Granovetter, 1973; 1983; Lin et al, 1981; Campbell et al., 1986).

Choice of variables is driven by theory and previous empirical studies, but there are data-driven concerns to take into account. Given the sample size too many dummy

variables increases the likelihood of small cell sizes, leading to an identification problem. Many of the variables available in the dataset are dummy or categorical variables. As discussed above, many of these are also potentially endogenous. The research therefore omits some dummy and categorical variables for farm assets (fixed and variable transactions costs) that do not have significant bivariate relationships with the outcome variable and/or much variation across the sample (namely ownership of large livestock, ownership or use of bicycle, use of hybrid seed, employer and employee dummy variables). Omitting these variables has trade-offs. While their omission allows the model to be estimated and convergence to be achieved, IV estimation will be biased if an omitted relevant variable is correlated either with the included non-endogenous explanatory variables or the instrumental variables. Access to crop storage, and radio ownership are retained as proxies for fixed transactions costs, and non-farm income and proportion of land cultivated as proxies for variable transactions costs.

To test the first hypothesis that a larger network is linked to positive outcomes, variables included are: log of the total number of network partners; group affiliation (household membership on community groups). A dummy variable for whether or not the household has influential kin or friends is included to test the potential effect of prestige networks (hypothesis 2). To test hypothesis 3, the benefits of having a more diverse network, three measures of network heterogeneity are included to capture diversity along two dimensions: diversity in composition (age heterogeneity of network partners and proportion of alters who are kin), and functional diversity or network multiplexity (number of multiplex ties as a proportion of total number of ties). Bivariate analysis in chapter 4 shows that other multiplexity variables are not significant, and only very weakly correlated with outcomes. Sex heterogeneity of network is used as an instrumental variable.

In estimating the relationship between the regressors and value of crop sold, in relation to the above hypotheses one would expect network size, group affiliation and influential kin variables to have positive signs. In the context of the research, age heterogeneity of network members may signify a less diverse network in that a network composed of many kin alters is highly likely to be more diverse in age range while being relatively homogeneous in other respects. The same reasoning applies to proportion of kin network partners – a high proportion of kin implies a less diverse network. There is

evidence to suggest there is a relationship between obligations towards kin and disincentives to accumulate (see Lewis, 1955; diFalco and Bülte, 2011). This is discussed in the context of the research in the results section below. As multiplex ties tend to occur among strong ties, that is, alters more likely to be like ego, for example kin and close friends, similarly one would expect in the context of the research for the functional diversity (multiplexity) measure to have a negative sign.

Location is captured as a community categorical variable, to control for institutional and other factors common to households in the three communities. Given the purposive nature of sampling of the survey locations, the location variable also proxies for market access. Recall, Ngulula was selected for its relatively good market access as it is located close to the provincial capital and just off the tarmac road. Kabila, by contrast, has poor market access due to its remote location. Lufubu Resettlement Scheme falls between the two and is considered to have ‘medium’ market access as it is located along the road and has reasonable access to the district capital Luwingu. One would expect Ngulula category to have positive correlation, Kabila negative. Human capital (education) of the household is captured as a dummy variable for whether or not a family member has completed at least secondary level schooling. Household demographics are given by age of household head and the household dependency ratio. The sign on both of these variables is expected to be negative.

Two interaction terms are included to explore potential effects of social network size given differential ability to process information (education level) and differential access to information: log of network size interacted with education; log of network size interacted with radio ownership.

The model is estimated with probability weights for each village, with the set of seven instrumental variables for the five potentially endogenous regressors, including the interaction terms. Results are given in Table 6.6. The estimates with instruments are compared with OLS estimates without instruments, that is, assuming all regressors are exogenous. OLS results are given in Appendix J. Larger standard errors on the potentially endogenous variables in the non-instrumented regression may indicate weak instruments.

Diagnostic tests for instrument relevance (including weak instruments) and validity, as well as over-identifying restrictions, are carried out at both stages of the two-stage estimation procedure and given in Table 6.5 below. Instruments should be reliable and valid, that is, i) be significant in equations; ii) not predict current market participation. Instrument relevance is tested via the first stage F test. The test for over-identification is via the Sargan or Hansen J test. Exogeneity is tested using the test of endogeneity under the ‘endog’ option used with the `ivreg2` command in Stata 11.2. Like the Hausman test, the null hypothesis is exogeneity of regressors⁴⁹. The test statistic is distributed as chi-squared, with degrees of freedom equal to the number of regressors tested. If significant, reject the null hypothesis in favour of the alternative of endogeneity. Otherwise, uphold exogeneity and there is no need to instrument.

Table 6.5 Summary results of diagnostic tests for instrument relevance and validity: GMM with social networks variables

First-stage regressions	Proportion of land cultivated (propland)	Log network size (logsize)	Log network size* education	Log network size*radio	Group affiliation (affil)
Partial R-squared of excluded instruments	0.35	0.89	0.96	0.98	0.40
F-test of excluded instruments ~F(7, 92)	2.49	41.15	13.65	19.15	2.70
Under-id Angrist-Pischke Chi-sq(3)	14.59	26.12	14.09	67.79	3.32
Weak-id Angrist-Pischke F(3, 92)	3.99	7.15	3.86	18.56	0.91
Underidentification tests					
Kleibergen-Paap rk LM statistic Chi-sq(3)	2.41				
Weak identification statistics Ho: equation is weakly identified					
Cragg-Donald Wald F statistic	0.31				
F-statistic ~ F(7,92)	2.47				
Chi-sq(7)	21.07				
Second Stage Estimation: GMM					
F(17, 94)	7.62				
Endogeneity test of endogenous regressors Chi-sq(5)	7.18				
Sargan-Hansen J statistic Chi-sq(2) (overidentification test of all instruments):	0.64				
Instrumented: propland logsize logsize*education logsize*radio affil					
Included instruments: depend loghh_age propkin _lcomid_2 _lcomid_3 kininf education store radio nfy propmult agehet					
Excluded instruments: areatot IQVst2 logmeansize logyearcomm tithe iqvradi iqved					

⁴⁹ The Hausman test is not available with `ivreg2`. However “under conditional homoskedasticity, this endogeneity test statistic is numerically equal to a Hausman test statistic; see Hayashi (2000, pp. 233-34)” (Stata help for `ivreg2`, www.stata.com, last accessed 19th September 2011; see also Baum et al, 2003).

The endogeneity test of endogenous regressors (*propland logsize logsizeed2 logsizerad affil*), under the null hypothesis that the specified endogenous regressors can actually be treated as exogenous, fails to reject the null hypothesis of exogeneity ($\text{Chi-sq}(5)=7.18$). However, it is unlikely that these regressors are exogenous, and likely that instruments are weak. Diagnostics from the first stage equations for the land and group affiliation variables do indicate potentially weak instruments: first stage F statistics are less than five: 2.49 and 2.70 respectively (Table 6.5). Coefficients on the instruments in the first stage regression show mixed results in terms of significance and there are some large standard errors. There are also large standard errors in the OLS estimation (Appendix J). As a set, these instruments may prove to be problematic: weak instruments are poor predictors of the endogenous variable and obtained predicted values will have very little variation (Murray, 2006). However, there is a strong possibility of Type II errors in the estimation due to the sample size. Further, it is very unlikely that the estimation will proceed with the full set of instruments and further testing is carried out to see if the approach successfully addresses the potential endogeneity problem. Proceeding with instrumental variable analysis risks losing efficiency in estimation, and IV estimates will converge towards those of OLS.

Diagnostic tests reject under-identification, weak identification, and fail to reject Sargan-Hansen over-identification test of all instruments ($\text{Chi-sq}=0.64$), so instruments appear to be valid.

Results are given in Table 6.6 below. For comparison, the model was also estimated omitting all social networks variables, retaining just the group affiliation variable. Diagnostics test results are given in Appendix L.

Turning to the GMM estimation in the second stage, the results are somewhat surprising given the significant, positive associations identified in Chapter 5 between outcomes and social networks. The variable for kin as a proportion of network is significant at 5 per cent, and is positively related to output, which is a surprising result and contrary to hypothesis three, and the bivariate in Chapter 5, that those with heterogeneous networks are better able to achieve their goals. One would thus expect having a higher proportion of kin in ones network, signifying less network diversity, to be less ‘useful’ in terms of achieving goals, in this case crop output. This is contrary to recent work on

the ‘dark side of social capital’ (diFalco and Bülte, 2011), which suggests tentative links between extensive kinship networks and lower incomes. The authors find “that households try to evade their ‘sharing obligations’ by (i) accumulating durables that are non-sharable at the expense of durables that may be shared and (ii) reducing savings in liquid assets. By attenuating accumulation incentives, kinship sharing may come at the expense of income growth – if so, a culturally-induced poverty trap can possibly eventuate” (diFalco and Bülte, 2011: 1128).

Table 6.6 Estimation Results: GMM regressions with instrumental variables

VARIABLES	logcsold
Proportion of land cropped (<i>propland</i>)	-1.65 (1.132)
Log of Network Size (<i>logsize</i>)	-1.53* (0.879)
Network size*education (<i>logsizeed2</i>)	1.57* (0.933)
Network Size*radio ownership (<i>logsizead</i>)	0.35 (0.769)
Group Affiliation (<i>affil</i>)	0.72 (0.639)
Household dependency ratio (<i>depend</i>)	-0.46 (0.706)
Log of Age of household Head (<i>loghh_age</i>)	-0.12 (0.447)
Kin as proportion of network (<i>propkin</i>)	1.18** (0.510)
Ngulula (<i>comid=2</i>)	1.16* (0.610)
Lufubu (<i>comid=3</i>)	-0.79 (0.926)
Influential Kin dummy (<i>kininf</i>)	0.48 (0.336)
Secondary education dummy (<i>education</i>)	-3.30* (1.995)
Access to crop storage (<i>store</i>)	0.30 (0.473)
Radio ownership (<i>radio</i>)	0.15 (1.742)
Nonfarm income (<i>nfy</i>)	0.00 (0.000)
Multiplex ties as proportion of network (<i>propmult</i>)	-0.90 (1.130)
Age heterogeneity of network (<i>agehet</i>)	-0.02 (0.028)
Constant	7.60*** (2.002)
Observations	112
R-squared	0.331

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The results in the thesis research could suggest that kin networks, while homogeneous in terms of ethnicity, may in fact be more diverse in other respects (including in gender and age), than assumed here. Kin relationships between men and women for example means interaction is more socially acceptable. Another possibility is that the situation in the study sites in this research is different compared to the context of DiFalco and Bülte's study of poor black households in KwaZulu-Natal. However, relationships between other social networks variables and crop income potentially corroborate DiFalco and Bülte's story, discussed below. Other network diversity measures, age heterogeneity and multiplexity, are not significant, so cannot tell us anything about Hypothesis Three.

Estimates with a significance level of 10 per cent are also discussed. Log of network size is significant at 10 per cent and the sign is negative. This suggests rejection of the first hypothesis that good outcomes are associated with larger networks. The negative coefficient here may be picking up the disincentive for households to accumulate if there are many calls on someone's resources, in that crop income is lower for the average farmer with larger social networks. During qualitative fieldwork one participant described the effect of his relative prominence in the community, the household head is the elected chairman of Lufubu Resettlement Scheme (analogous to headman), has had on his household: "You would expect us to be top in this community but we're not, we're more in the middle because we have extra people in our household now, and other people in the community expect me to help them". Indeed, according to the survey data the household is relatively successful in crop farming, with the second highest value of crops sold in the community. However, this is not reflected in the household's 'medium' socio-economic status as calculated from the survey data using various asset indicators (see Appendix D).

Qualitative fieldwork in Kabila village also uncovered evidence of culturally and socially embedded constraints on agricultural production, related to belief in witchcraft. This goes beyond fear of 'imagined', 'supernatural' entities, and superstition, to the very real threat of action by other people, ruining crops, killing animals and making people sick. This acts as a deterrent both to the recipient of the 'witchcraft' and also to

those around them who witness their ‘bad luck’. Kabila villagers said that as a result of these fears they grow “just enough” and “not too much”. However, growing “just enough” does not leave much of a margin of error leaving households even more vulnerable to production risk including pests and disease. When questioned about holding these fears despite evidence of people in the village growing a substantial surplus and ‘getting away with it’, respondents explained that for many this was not enough to overcome the fear of reprisals for being seen to be too successful. Information about witchcraft at the personal level is very difficult to uncover; the subject is so sensitive it requires a degree of trust between the interviewers and interviewees. While witchcraft beliefs featured similarly in Ngulula and Lufubu, the extent to which it is linked to livelihood constraints in these sites is unclear. The social networks variables may be picking up this effect. While having many kin in one’s network may reinforce belief in and fear of superstition, having many kin may also mean a household is more able to risk a lower margin of error because kin can potentially help them out if they face (idiosyncratic) shocks.

Turning to the interactive terms between network size and education and radio ownership, do they offset the apparent negative effect of larger social networks on outcomes? Interaction of network size with radio ownership is not significant, and one might expect this to have only a weak effect on the value of crops sold in contexts such as this: small village communities where everyone knows what everyone else is doing and market information is readily available from traders who come to the village. The interaction between social networks and the secondary education dummy is only significant at 10 per cent, and is positive. The secondary education dummy is also significant at just 10 per cent but has a negative sign. Thus the effect of social network size on output also potentially depends on the ability to interpret information through education level. The result suggests that on average, there is a higher (positive) effect of social networks on output for farmers with higher education levels in their household compared to farmers in other categories. Households where the highest level of education is secondary and above, who also have larger social networks, are likely to sell more crops. Having larger social networks per se does not help (as demonstrated by the negative sign on the significant social network size variable) and the offsetting effects in the interactive terms may be capturing independence from these networks. The transaction costs argument here is linked to the way information lowers transactions

costs. More novel information is provided through a larger social network and the ability to interpret and use this information is enhanced by higher education levels.

The results of the GMM estimation do not provide enough information to be able to make a judgment on hypothesis 2 that prestige networks ('linking' social capital) are associated with good outcomes. The influential alters variable is not significant. One reason for this could be that the perception of 'influential' may not in practice amount to very much in terms of the function performed by such alters and the survey might have benefited from further probing questions related to the functionings of prestige networks. The group affiliation variable is also not significant in the estimation.

The only other non-social networks variable that was significant (again at only the 10 per cent level) was the coefficient on the community variable for category 2, *Ngulula*, relative to the base category *Kabila*, which was the village with the worst market access. This suggests that aside from social networks, nothing really matters for good commercialisation outcomes apart from being in a good location.

6.6 Conclusion

This chapter has attempted to establish causal links between social networks and commercialisation outcomes, measured by the value of crops sold, to explore the potential role of social networks in reducing transaction costs and mediating household access to crop markets. Do social networks, by reducing transactions costs and mediating market access, help to explain differences between households?

The three hypotheses introduced in chapter 4 suggest three channels through which social networks might have beneficial outcomes: 1) *Bigger is better*: larger networks are associated with better outcomes; 2) *There are benefits to having 'friends in high places'*, defined as links to persons with prestige or influence; and 3) *Benefits of heterogeneous networks*: network diversity means greater access to novel information and increased likelihood of being able to draw on diverse resources through networks. Bivariate analysis in Chapter 4 suggests patterns emerging relating social networks characteristics to outcomes in support of the three hypotheses: sellers operating at higher levels of

crop sales are more likely to have bigger social networks in general, with relatively more non-kin members, more mixed-sex networks (though less diverse in terms of age), and be members of multiple village groups and societies. These lend support for Hypotheses 1 and 3.

The analysis in this chapter builds on this with causal analysis incorporating social networks variables into a model of agricultural commercialisation with transactions costs. A significant challenge in estimating the relationships between social networks and outcomes of smallholder farming households is in establishing direction of causality and correcting for other forms of endogeneity, such as selectivity bias. The empirical model of crop market participation is estimated using Generalised Method of Moments with instrumental variables. This is chosen as the preferred estimation procedure as it allows an instrumenting regression combined with survey weights. A selection model and censored Tobit were also explored as potential strategies. The selection term in the second stage of a Heckman procedure is not significant (Appendix K). This may signify that selection bias is not a problem in the sample. However, the relatively small subsample of non-sellers is likely to be having an effect here. With this in mind, coupled with the difficulties in achieving convergence when estimating both the Heckman and the Tobit models with instrumental variables, the analysis is based on the GMM results. The small sample size and some potentially weak instruments may also be driving the failure to reject exogeneity in the endogeneity tests of the GMM estimation. It is highly unlikely in the context of the research that relationships considered to be potentially endogenous are in fact exogenous, notably between crop incomes and: i) social network size; ii) area cultivated iii) group membership. The research therefore proceeds in estimating the model using instrumental variables. The trade-off here is in efficiency terms. If exogeneity is indeed upheld, then by instrumenting the relationships the estimation is less efficient and results converge to OLS.

With these limitations in mind, do the network variables have predictive value in the analysis of returns to social networks, captured as household crop market participation? Once the analysis controls for other variables and possible endogeneity, the results suggested by the bivariate in chapter 5 fall down. In contrast to the descriptive statistics, the analysis here suggests larger networks in fact have a negative effect on

crop sales. This potentially provides support for findings emerging in other studies that find negative effects of social capital/ social networks (diFalco and Bülte, 2011; Baland et al, 2011). Interacting the network size variable with the education variable suggests that larger networks go hand in hand with higher levels of crop sales only for those households that have at least one member with secondary-level education. Households where the highest level of education is secondary and above, who also have larger social networks, are likely to sell more crops. Group affiliation and age heterogeneity of network partners are no longer significant in the multivariate analysis, in contrast with the bivariate in Chapter 5.

The relationship between having a larger proportion of kin in ones network and crop sales is now positive. This may suggest that in relation to Hypothesis 3, ‘*the strength of heterogeneous networks*’, either kin networks do have value in that they are potentially more easily mobilised than so-called ‘weaker’ non-kin ties, or that kin networks may not be as homogeneous as the research assumes in relation to the resources and networks contained, and accessed, within them.

.....

The next and final chapter concludes.

Chapter Seven

Conclusions

7.1 Summary of the study

In this thesis I have attempted to examine the relationship between social networks and economic life in rural Zambia. The overarching question is whether or not social networks help to determine household agricultural commercialisation in settings where there is little variability between households on endowments such as land, labour and farm inputs, in contexts of incomplete or missing markets. Do social networks, by mediating market access, help to explain differences between households?

The role of social networks in economic life presents an interesting empirical question. Research in anthropology, economic sociology and social capital research in economics suggest that social context and social networks play a crucial role in exchange behaviour where markets and institutions are missing or incomplete. In such contexts the literature, discussed in Chapters 1 and 2, suggests social networks mediate access to markets and resources in environments characterised by an absence of functioning credit and insurance markets, cash constraints and a lack of assets. They may serve to help people access information, technical know-how, and resources. However, emerging research that contributes to debates about the ‘dark-side’ of social capital suggests social networks can also potentially have negative economic consequences. This relates to obligations towards family, kin and friends that are embedded in social networks. People may feel burdened by responsibilities and such obligations can represent a drain on a household’s resources and thus lead households to adopt strategies that purposefully do not lead to an excess of wealth or income (diFalco and Bulte, 2011; Baland et al, 2011). Social networks can also be a basis for exploitation and social marginalisation (see, for example, Kabeer, 2000). Negative externalities of social networks/social relations may impact directly on welfare outcomes of smallholder farming households through non-optimal behaviours including: disincentive to accumulate assets, crime, taking health risks, educational underachievement (see also: Portes, 1996; Fine, 1999; Carroll and Stanfield, 2003; Torpe, 2003).

The inherent difficulties in formalising socially-embedded market transactions within the standard neoclassical economics framework provide further motivation for the study.

The empirical study examines rural production systems in three predominantly Bemba villages in Northern Province, Zambia, focusing on agricultural commercialisation and market participation. Primary data collection took place during fieldwork for the thesis research over three periods totalling four months between March 2002 and May 2003, using both a household survey and qualitative methods to gather data on social networks and other farm household attributes. Social network analysis techniques are used to construct variables capturing characteristics of people's social networks, which are incorporated into models of commercialisation based on a transactions cost framework (Bellemare and Barrett, 2006; Key et al, 2000; Goetz, 1992). The research is informed by in-depth qualitative material gathered during extended stays in the study sites. While the sample is by no means nationally representative, the case studies presented here yield insights applicable to other smallholder farmers living in similar conditions and in similar ways.

7.2 Challenges in undertaking the research

Challenges faced in undertaking the research were many, spanning the conceptual, methodological and analytical domains.

Conceptual: The research is premised on the notion that 'economic' exchanges take place within the social realm and social interactions, institutions and relations go beyond a narrow, binary way of thinking in terms of possessing or not possessing a certain attribute. Rather, the research, drawing on insights from anthropology and economic sociology, considers people to be embedded within their social sphere, operating on economic, political, cultural and social levels in relation to others who are also socially embedded, in their sense of 'self' and in their actions. It was hoped that using a social network analytical approach would allow features of the institutional environment, not just 'capital' aspects of social networks, to be captured. This is achieved through measuring interactions, transactions and exchanges through networks

themselves, as well as the attributes of the different actors and the relations between them. In this way the network is viewed not only as something an actor (ego) uses but also something they experience. A transactions costs approach to modelling commercialisation offers a conceptual framework that allows incorporation of the social context into the analysis. In practice, this still amounts to stripping back interactions and experiences through networks to possession or not of an attribute, in that the interaction becomes an attribute of ego (the household). The approach is necessarily reductionist in order to be able to relate social networks phenomena to outcomes within an economics framework. However, questions remain concerning whether or not we will ever really get to the heart of social interactions and what they mean, and consequently their true impact on, and relationships with, outcomes. It is true that interpretation of some of the results in this research would be difficult without qualitative findings to provide clarification. Does this matter? If we can achieve an approximation of the experiential, and it is informed by in-depth, contextual data, then surely this is enough?

Methodological: Individual, household, group and community level data were collected, including social networks data, using a combination of quantitative (survey) and qualitative methods and instruments. It was an ambitious exercise, requiring considerable resources and time. In terms of social networks analysis, resource constraints meant that the collection of complete network data was not possible, that is, data on all actors in a network and all of the links between them. Instead the research had to take an egocentric approach. This also meant focusing on only first-order direct ties and not ties between alters. However, one must bear in mind there may be good (or bad) resource access through second order ties. These are implicit in the research, rather than explicit. Resource constraints also ruled out attempting to uncover more ‘psychosocial’ information about households and individuals, such as personality traits, which may affect social networks.

Once the data were collected, data cleaning and construction of the social networks dataset were incredibly time-consuming. There was no roster of all possible network partners, so social networks data collection included recording names of network partners, self-reported by respondents. These subsequently were coded manually to convert the names to alter identification numbers in order to create a dataset where each case was a particular alter. What made the task particularly time consuming was

multiple spellings for the same name and the use of nicknames. The latter had not been foreseen in the design of the first round of the survey, so extra questions and prompts were included in Round 2 of the survey to capture people's nicknames and allow for accurate coding of alters. Recording names also lengthened the time it took to conduct interviews, trading off sample size for more in-depth information. Because the research aims to be generalisable to smallholder farmers similar to those living in the study sites, this is not believed to be a problem. However, the relatively small sample size (136 households in total) means that some of the diagnostic tests may not work as well as they could.

Analysis: Addressing endogeneity is the primary challenge in modelling relationships in this study. The focus on just three study sites ruled out the use of community-level variables as instrumental variables. An 'approximation' to lagged dependent variables was also considered, by using regressors from round 1 of data collection and crop sales outcomes reported in round 2 of the survey, with a recall period of the previous six months, since the last survey. Most crops were sold prior to round 1 of the survey, which took place in September, in the post-harvest period. The ability to sell all year round might signal a higher level of commercialisation; farmers able to hold back some of their output can take advantage of seasonality in prices: lower near harvest time in August, higher later in the year. However, the sample size was not large enough to do this.

Stepping away from standard economic theory in relation to endogeneity and causality, work originating in other disciplines, notably psychology and neural networks, suggests that rather than thinking about cause and effect in a model of the world embedded within a logic-based framework, one should consider actors as operating within a systemic world with complex chains of social interaction, and the importance of feedback in maintaining these systems.⁵⁰ These ideas were taken up by anthropologist Gregory Bateson in relation to study of human systems such as the family.⁵¹ Crucially, systems are believed to stay stable by constantly adapting to changing conditions and

⁵⁰ John Bowlby's work in psychology for example stems from the belief that to understand people one also has to understand their environment – the need for contextual data is very much in step with what is generally accepted in social sciences, including economics.

⁵¹ Bateson also asserts that "logic is a poor model of cause and effect" (from Bateson's exploration of epistemology 'Mind and Nature, a necessary unity', 1980).

they do this through feedback about what works and what does not. The system as a whole therefore can be seen as circular and not linear. Rather than breaking a system into separate parts, and treating them as if they function in isolation, it makes more sense to understand that systems are connected to and determine each other. It follows that cause and effect depend on one's position, and, crucially, on where one starts in the loop, as well as how much information is included or excluded. A neural networks approach to mathematical modelling may be appropriate in this context, particularly in modelling social network formation. Neural networks, in economics, have mainly been used for prediction in time series, for example forecasting in capital markets, but they can also be applied to network formation in relation to social institutions (See Angus et al, 2007). However, this is for future exploration and well beyond the scope of this thesis.

7.3 Summary of the findings

Economic exchange takes place on a reciprocal or kinship basis throughout sub-Saharan Africa. This certainly typifies market activity of individuals and communities living in the study area. Social networks variables, constructed using social network analysis techniques, provide insights into the social context, going beyond more simplistic measures typically used in social capital studies, for example group membership. The analysis in Chapter 4 demonstrates that how social networks are measured matters. Different social networks attributes are important for different people. The research then demonstrates that relationships between social networks and outcomes depend on the measure being used. The constructed social networks measures allow the research to test hypotheses linking different social networks characteristics to specific outcomes: The first hypothesis is concerned with network size: *'bigger is better'*. The second links influential network partners to positive outcomes: *'benefits to having friends in high places'*. The third hypothesis refers to network diversity: *'the strength of heterogeneous networks'*.

The hypotheses suggest that larger, more diverse networks are associated with positive outcomes, and to a certain extent this appears to be borne out by patterns emerging from bivariate analysis of the data, which identifies relationships between social network

characteristics and crop market participation: both entry into market and degree of participation in terms of crop sales. Network size and diversity on the whole do indeed seem to be associated with good outcomes captured as socio-economic status and higher levels of crop marketing. There is a significant, positive correlation between value of crops sold and network size, and with group affiliation, lending support for Hypothesis 1. By contrast there is a significant, negative correlation with higher proportions of kin in the network, suggesting some support of the ‘strength of heterogeneous networks’ hypothesis, which posits that novel and hence more valuable information is contained in networks where alters have weaker links between each other. A network consisting of many kin members implies not only strong links between ego and alters, but also strong links between alters. More mixed-sex networks (though less diverse in terms of age), and household membership in multiple village groups and societies also go hand-in-hand with operating at higher levels of crop sales. This lends support for Hypothesis 3. In line with Hypothesis 2, households participating in markets and those selling higher values of crops appear to be more likely to have prestige networks, captured as kin in an influential position.

However, the multinomial probit analysis of households clustered according to command over resources and marketing levels, while pointing to the importance of networks in affecting both control over resources and the ability to sell higher volumes of output given resources available, suggests that network size per se is not the factor that differentiates between success or not in terms of crop marketing (Chapter 5). The content of a network is arguably more important than its size.

The research explores these results further in Chapter 6 via multivariate, causal analysis of the potential role of social networks in reducing transaction costs and mediating access to markets, to consider whether the network variables have predictive value in the analysis of returns to social networks, captured as household crop market participation (the total value of output sold across all crops).

A significant challenge in estimating the relationships between social networks and outcomes of smallholder farming households is in establishing direction of causality and correcting for other forms of endogeneity such as selectivity bias. It is highly unlikely in the context of the research that relationships considered to be potentially endogenous

are in fact exogenous, notably between crop incomes and: i) social network size; ii) area cultivated iii) group membership. Endogeneity is likely to be an issue and the research therefore estimates the model of crop market participation using Generalised Method of Moments with instrumental variables, even though the endogeneity test fails to reject the null hypothesis of exogeneity. The trade-off here is in efficiency terms. If exogeneity is indeed upheld, then by instrumenting the relationships the estimation is less efficient and results converge to OLS. There are challenges too in finding appropriate instruments: those that are relevant and valid. Diagnostics tests suggest that the instrumental variables used in the research are valid but some may be weak.

With these limitations in mind, do the network variables have predictive value in the analysis of returns to social networks, captured as household crop market participation? Once the analysis controls for other variables and possible endogeneity the relationships between social networks and outcomes suggested in the bivariate analysis are not upheld. In contrast to the descriptive statistics, the multivariate analysis suggests larger networks in fact have a negative effect on crop sales. An interaction of the network size variable with education suggests that larger networks go hand in hand with higher levels of crop sales only for those households that have at least one member with secondary-level education. Group affiliation and age heterogeneity of network partners are no longer significant in the multivariate analysis. The relationship between having a larger proportion of kin in ones network and crop sales is positive once the research controls for endogeneity. This may suggest that in relation to Hypothesis 3, the strength of heterogeneous networks, kin networks do have value in that they are potentially more easily mobilised than so-called ‘weaker’ non-kin ties. Another explanation may be that kin networks are not as homogeneous as assumed in the research.

On balance, do social networks help or hinder access to or competition in markets, focusing on trading? Particular features of social networks certainly do seem to have a causal effect on the level of household crop market participation. Larger networks appear to have a negative effect on crop incomes, which could be ascribed to a large following translating into a greater call on a household’s resources, rather than being a conduit for resource mobilisation and hence greater returns. The negative effect could be due to a disincentive to accumulate as a response to higher levels of obligation towards network partners, or the levelling effect of social norms such as belief in

witchcraft. On the other hand, network composition is important. Caveats about potential endogeneity notwithstanding, Chapter 5's analysis of household clusters according to command over resources and marketing levels suggests that network size per se is not the factor that differentiates between success or not in terms of crop marketing. How people use their networks is arguably more important, as is the composition of the network demonstrated by the significant coefficients on the influential kin variables. The multivariate analysis of the determinants of level of crop sales in Chapter 6, controlling for endogeneity with instrumental variables, suggests that having a greater proportion of kin in the network has a positive effect on household crop incomes, implying that in this context strong ties are key.

7.4 Broader implications

The study is based on original, primary data collected by the author. While the role of social networks and social capital is widely researched in relation to technology diffusion, migration and risk and insurance, relatively little has been done in agricultural development. The research potentially makes three clearly defined contributions to the literature in social networks in the context of agricultural and rural development:

1. It is important how we measure social networks: different social network characteristics are important for different people;
2. Relationships between social networks and outcomes depend on the measure used;
3. Results from the empirical analysis potentially provide support for findings emerging in other studies that find negative effects of social capital/ social networks (see for example: diFalco and Bülte, 2011; Baland et al, 2011).

Methodological insights from the research therefore has potential implications for the way we research social networks and social network capital, in particular the importance of considering the nature of social networks when trying to look at the effect of "connectedness" on economic decisions. Once social network measurement and reverse causality issues are taken into account, the research also provides some interesting empirical insights into the extent to which social networks or connectedness matter. Notably, large networks do not necessarily lead to positive outcomes. These contributions have important implications for policies and programmes that explicitly

function through social structures, for example microfinance schemes operating through social models of lending, work linking social protection and informal networks, the role of networks in migration and remittances, among others.

Social networks as the basis for exclusion and social marginalisation has been discussed in the development literature (see for example Kabeer, 2000). Although the household survey did not contain the data to be able to analyse this quantitatively, and in any case was probably not the appropriate instrument to uncover this information, qualitative evidence from the fieldwork suggests that failure to participate in networks, often for reasons of poverty, can result in more disadvantaged community members remaining locked in a vicious circle of poverty and exclusion from livelihood opportunities. For example not having clean clothes to go to church where many calls for labour are made. One woman from the most disadvantaged household in one of the communities said she did not attend church even though she'd like to because "I fear my dirtiness will drive people away". As a result the family would often hear calls for work when it was too late. The ways in which people engage with, or are excluded from, networks thus impact crucially on wellbeing. Often it is the stories from people on the 'margins' like this that are most revealing and worthy of further, targeted research.

Turning to the broader, contemporary context for agriculture in Sub-Saharan Africa, work by Pingali et al (2005) couches the transactions costs challenge faced by smallholder farmers in terms of a broad, agri-systems view of agricultural markets. The authors argue that a highly integrated, modern agri-food system with its demands in terms of standards and contractual requirements, means that potential market participants, including smallholder farmers, face "a new set of transaction costs that emerge from dealing with a food system characterized by different rules, regulations, and players" (Pingali et al, 2005: 2). If certain types of social networks, and the functions they perform, are an important determinant in allowing smallholder farmers to access markets, are these going to be enough in the context of faster moving markets with more stringent entry requirements? Arguably the social networks of the average smallholder farmer will not be enough and governments must take responsibility for providing the missing or inaccessible institutions social networks appear to be stepping in for, to help reduce transactions costs to enable smallholder farmers to participate in the global food system. The field research for the thesis was conducted in the pre-

mobile phone context. An important additional element in couching the research in the contemporary setting is to consider the role of technology in relation to networks and to market access. How might technology have changed things for the participants in the research? What is the role played by mobile technology in reducing transactions costs? How might technology interact with social norms and social networks in determining who is well connected? The early anthropological literature highlights the importance of commanding a following and the key role of prestige in this, enabling people to access resources and mobilise labour leading to positive outcomes. Education has latterly become an important dimension of prestige and thus an attractor of others. As technologies become increasingly accessible and ubiquitous even in the most remote rural areas it will be interesting to see whether, and how, the way people use technology, in interaction with social norms, changes social roles and relations and the way people engage in economic life.

Bibliography

- Agresti, A. and B. F. Agresti, B. (1977) "Statistical analysis of qualitative variation". In: Schuessler K F (eds) *Sociological Methodology* (1978) San Francisco: Jossey-Bass.
- Allen W.D. (2000). 'Social Networks and Self-Employment'. *Journal of Socio-Economics* **29**(2000) 487-501.
- Angus Simon, Andrew Bell, Paul L. Hooper, Elizabeth Mullane, & Rafal Raciborski (2007) "A new approach to modeling social institutions using artificial neural networks." *Collected Papers of the Santa Fe Institute Complex Systems Summer School, 2007*. Santa Fe Institute.
- Annen, K (2003) 'Social capital, inclusive networks, and economic performance'. *Journal of Economic Behavior & Organization* **50**(2003) 449–463.
- Arrow, K., (2000). 'Observations on social capital'. In: Dasgupta, P., Serageldin, I. (Eds.), *Social Capital: A Multifaceted Perspective*. The World Bank, Washington, DC.
- Bala V. and S Goyal (2000). 'A Noncooperative Model of Network Formation'. *Econometrica* **68**(5), 1181-1229.
- Baland, J-M., Guirkinger, C. and Mali, C. (2011) Pretending to be poor: borrowing to escape forced solidarity in Cameroon. *Economic Development and Cultural Change* **60**(1): 1-16.
- Bandiera, O. and Rasul, I. (2006). 'Social Networks and Technology Adoption in Northern Mozambique' *Economic Journal* **116**: 869–902.
- Barahona C. and S. Levy (2003). 'How to generate statistics and influence policy using participatory methods in research: reflections on work in Malawi 1999-2002'. *IDS Working Paper 212*, Institute of Development Studies, Brighton, UK.

- Bardhan, P.K. (1980) 'Interlocking Factor Markets and Agrarian Development: A Review of Issues,' *Oxford Economic Papers* **32** (1): 82-98.
- Barr A, (2002) "The Functional Diversity and Spillover Effects of Social Capital", *Journal of African Economies*, **11**(1), 90-113, March 2002.
- Barro Robert J.; Rachel M. McCleary (2003) 'Religion and Economic Growth across Countries', *American Sociological Review*, **68**(5): 760-781.
- Baum C, M Schaffer, S Stillman (2003) 'Instrumental Variables and GMM: Estimation and Testing'. *Boston College Department of Economics Working Paper No 545*.
- Becker G (1976) *The Economic Approach to Human Behavior* Chicago: University of Chicago Press.
- Becker Gary S and Kevin M Murphy (2000) *Social Economics: Market Behaviour in a Social Environment* The Belknap Press of Harvard University Press: Cambridge, Massachusetts.
- Bellemare Marc F. & Christopher B. Barrett (2006) "An Ordered Tobit Model of Market Participation: Evidence from Kenya and Ethiopia," *American Journal of Agricultural Economics* **88**(2):324-337.
- Bernard H.R., P.D. Killworth, D. Kronenfeld and L. Sailer (1984). 'On the validity of retrospective data: The problem of informant accuracy'. *Annual Review of Anthropology* **13**: 495-517.
- Berry S. (1986). "*Fathers Work for their Sons*". Berkeley and Los Angeles: California University Press.
- Berry S. (1989). "Social Institutions and Access to Resources" *Africa* **59**(1) 41- 55.
- Berry S. (1993) "*No Condition is Permanent*". Wisconsin: University of Wisconsin Press

- Bevelander Pieter and Ravi Pendakur (2007) Minorities, Social Capital and Voting *IZA DP No. 2928 Discussion Paper Series*. Institute for the Study of Labor IZA July 2007.
- Binns, J.A. (1992) 'Traditional Agriculture, Pastoralism and Fishing' in M.B. Gleave (1992) *Tropical African Development: Geographical Perspectives*. Harlow: Longman.
- Binswanger HP and MR Rosenzweig (1986) 'Behavioural and Material Determinants of Production relations in Agriculture' *Journal of Development Studies* **22**(3): 503-539.
- Blau,P (1964). *Exchange and Power in Social Life*. New York: Wiley.
- Borgatti S (1998) <http://www.analytictech.com/networks/socsup.htm>; accessed 29th October 2010
- Bourdieu, P (1986) 'The forms of capital', in Richardson (ed.) *Handbook of Theory and Research for the Sociology of Education*, 1986 pp. 241-58. Westport, CT: Greenwood Press.
- Bourdieu, P. (1983). *The forms of capital*. In J.G. Richardson (Ed.) *Handbook of theory and research for the sociology of education*. Westport, CT: Greenwood.
- Brewer D.D. (2000). 'Forgetting in the recall-based elicitation of personal and social networks' *Social Networks* **22**:29-43.
- Brewer D.D. (2000). "Forgetting in the recall-based elicitation of personal and social networks." *Social Networks* **22**:29-43.
- Brück Tilman (2004) 'Coping Strategies in Post-War Rural Mozambique' *HiCN Working Paper 02 March 2004*.
- Bryman (1984) 'The Debate about Quantitative and Qualitative Research'. *British Journal of Sociology* **35**(1): 75-92.
- Burt R S (1992) *Structural Holes the Social Structure of Competitions*. Cambridge MA: Harvard University Press.

- Burt R S (2001) 'Structural Holes versus Network Closure as Social Capital' in N Lin, K Cook and RS Burt (eds) 2001 *Social Capital: Theory and Research* Hawthorn NY: Aldine de Gruyter pp 31-56 .
- Campbell K, P V Marsden and J S Hurlbert (1986) "Social Resources and Socio-Economic Status" *Social Networks* **8**: 97-117.
- Carroll, Michael C, and James Ronald Stanfield. (2003). 'Social capital, Karl Polanyi, and American Social and Institutional Economics.' *Journal of Economic Issues* **37**: 397-404.
- Carter, Michael R. and J. Maluccio (2003) 'Social Capital and Coping with Economic Shocks: An Analysis of Stunting of South African Children,' *World Development*, **31**(7), 1147-1163.
- Carvalho S. and H. White (1997) 'Combining the Quantitative and Qualitative Approaches to Poverty Measurement and Analysis'. *World Bank Technical Paper No. 366*. Washington D.C.: The World Bank.
- Casley D.J. and D.A. Lury (1982) *Monitoring and Evaluation of Agriculture and Rural Development Projects*. Baltimore and London: Johns Hopkins University Press.
- Castiglione D (2005) 'Introduction: Conceptual issues in Social Capital Theory' (pp13-21) in *The Handbook of Social Capital*. Castiglione, D., Van Deth, J. W., & Wolleb, G., eds., New York: Oxford
- Chambers R. (2002) *Participatory Workshops* London and Sterling VA: Earthscan Publications Ltd.
- Chambers R. (2003a) Participation and Numbers *PLA Notes 47* International Institute for Environment and Development, London.
- Chambers R. (2003b) in Kanbur R. (2003). *Q-Squared: Qualitative and Quantitative Methods of Poverty Appraisal*. Delhi: Permanent Black.

- Chapoto A, Diana Banda, Steven Haggblade, and Priscilla Hamukwala (2011) 'Factors Affecting Poverty Dynamics in Rural Zambia'. *FSRP Working Paper No. 55*. July 2011.
- Chizyuka, R., R. Kamona, C. Ufwenuka, and M. Phiri (2006). National Report – Zambia: Policies and Strategies for Agrarian Reform and Rural Development to Secure and Improve Access to Natural Resources. *Paper prepared for the International Conference on Agrarian Reform and Rural Development (ICARRD) 7-10 March. Porto Alegre, Brazil.*
- Christiaensen (2003) in Kanbur R. (2003). *Q-Squared: Qualitative and Quantitative Methods of Poverty Appraisal*. Delhi: Permanent Black.
- Christiaensen and Alderman (2001) 'Child Malnutrition in Ethiopia: Can Maternal Knowledge Augment The Role of Income?' October 2001 *Africa Region Working Paper Series No. 22* World Bank: Washington DC.
- Cliggett L (2005) 'Remitting the Gift: Zambian Mobility and Anthropological Insights for Migration Studies'. *Population, Space and Place* **11**, 35–48.
- Coase R H (1937) *The Nature of the Firm*.
- Coleman J S (1990) *Foundations of Social Theory*. Cambridge, MA : Harvard University Press.
- Coleman J S (1988) 'Social Capital in the Creation of Human Capital'. *American Journal of Sociology* **94** (suppl): S95-S120.
- Collier P (1998), *Social Capital and Poverty*, Social Capital Initiative WP No 4, The World Bank, Washington DC
- Deaton A. (1997) *The Analysis of Household Surveys* Washington D.C.: World Bank/IBRD: Washington DC.
- Conley T and C R. Udry (2010). 'Learning about a New Technology: Pineapple in Ghana," *American Economic Review*, American Economic Association', **100**(1), pages 35-69, March.

- Conley, T., and C.R Udry (2001). 'Social learning through networks: The adoption of new agricultural technologies in Ghana'. *American Journal of Agricultural Economics* **83**(3): 668–673.
- Cook, T. & Reichardt, C. (Eds.) (1979). *Qualitative and Quantitative Methods in Evaluation Research*. Beverly Hills, CA: Sage Publications.
- CSO (2002) *Living Conditions Monitoring Survey*. Lusaka, Zambia: Central Statistical Office.
- CSO (1998) *Living Conditions Monitoring Survey 1998*. Lusaka, Zambia: Central Statistical Office.
- d'Hombres B, Rocco L, Suhrcke M, et al. (2010) Does social capital determine health? Evidence from eight transition countries. *Health Econ* 2010;19 (1)56-74.
- Darr, D. and Pretzsch, J. (2006). *The Spread of Innovations within Formal and Informal Farmers Groups: Evidence from Rural Communities of Semi-Arid Eastern Africa*. Paper presented at the *Conference on International Agricultural Research for Development, Tropentag, University of Bonn, October 11-13 2006*.
- Dasgupta P (2003) Social capital and economic performance: analytics. In Ostrom E and T Ahn (eds) *Foundations of Social Capital*. Edward Elgar: Cheltenham pp 309-339.
- Dasgupta, P. (2000) Economic progress and the idea of social capital, in: P. Dasgupta and I. Serageldin (eds) *Social Capital – A Multifaceted Perspective* pp. 325–424 Washington DC: World Bank.
- de Bruin, A., and Dupuis, A. (1999). 'Towards a synthesis of transaction cost economics and feminist oriented network analysis' *American Journal of Economics and Sociology*. 58(4) Oct., 1999.
- De Haan A., K Brock and N. Coulibaly (2002) "Migration, Livelihoods and Institutions: Contrasting Patterns of Migration in Mali". *Journal of Development Studies* **38**(5), 37-58.

- De Weerd, J. (2004). 'Risk sharing and endogenous network formation.' In S. Dercon, ed. *Insurance against poverty*. Oxford: Oxford University Press.
- De Weerd, Joachim & Dercon, Stefan (2006). 'Risk-sharing networks and insurance against illness,' *Journal of Development Economics*, **81**(2)337-356.
- De Weerd, Joachim, (2002). 'Risk-Sharing and Endogenous Network Formation'. *UNU-WIDER Research Paper*, World Institute for Development Economic Research (UNU-WIDER).
- Delgado C (1997) 'The role of smallholder income generation from agriculture in sub-Saharan Africa' in Haddad L. (ed) *Achieving Food Security in Southern Africa: New Challenges, New Opportunities*. Washington DC: International Food Policy Institute.
- Delgado C (1995). 'Agricultural diversification and export promotion in sub-Saharan Africa'. *Food Policy* **20**(3): 225-243.
- Delgado C. (1991) 'Cereal Production and Agricultural Development Strategy in the Sahel' *Food Policy* **16**(2): 105-111.
- Delgado C. (1999) 'Sources of Growth in smallholder agriculture in Sub-Saharan Africa: the role of vertical integration of smallholders with processors and marketers of high-value-added items'. *Agrekon* special issue, **38**(May): 165-189.
- den Ouden, Jan H.B. (1995) 'Who's For Work? The management of labour in the process of accumulation in three Adja villages, Benin' *Africa* **65**(1) 1-35.
- Denzin N.K. and Y.S. Lincoln (Ed.) (1998). *Collecting and Interpreting Qualitative Materials*. Thousand Oaks CA: Sage Publications.
- Dercon S and P Krishnan (2000) Risk-sharing within households in rural Ethiopia. *Journal of political Economy* **108**(4): 688-727.
- Dercon, S., Krishnan, P.(2000). 'In Sickness and in Health: Risk Sharing within Households in Rural Ethiopi'a. *Journal of Political Economy* **108**:688-727.

- DFID/OPM (2000). Making Markets work better for the poor. A Framework Paper. UK *Department for International Development*, London.
- di Falco Salvatore & Erwin Bulte (2011) 'A Dark Side of Social Capital? Kinship, Consumption, and Savings' *Journal of Development Studies* **47**(8): 1128-1151.
- Dorward A, (1999). A risk programming approach for analysing contractual choice in the presence of transaction costs *European Review of Agricultural Economics* **26**(4):481-492.
- Dorward A, N. Poole, J Morrison, J, Kydd and I Urey (2003). Markets, Institutions and Technology: Missing links in livelihoods analysis, *Development Policy Review* **21**(3): 319-332).
- Dorward A, S Fan, J Kydd, H Lofgren, J Morrison, C Poulton, N Rao, L Smith, H Tchale, S Thorat, I Urey and P Wobst (2004) Institutions and Policies for Pro-Poor Agricultural Growth. *Development Policy Review* **22**(6): 611-622.
- Durlauf, S. (2000). 'The memberships theory of poverty: the role of group affiliations in determining socioeconomic outcomes.' In *Understanding Poverty in America* (eds. S. Danziger and R. Haveman). Cambridge: Harvard University Press.
- Durlauf, SN (2002a) On the Empirics of Social Capital. *The Economic Journal* **112**(483) Features (Nov., 2002), pp. F459-F479.
- Durlauf S (2002b) 'Bowling Alone: a review essay' *Journal of Economic Behavior and Organization* **47**(3): 259-273.
- Durlauf, Steven N. & Fafchamps, Marcel, (2005) 'Social Capital,' in Philippe Aghion & Steven Durlauf (ed.), *Handbook of Economic Growth*, edition 1, volume 1b, chapter 26 pp 1639-1699 Elsevier.
- Epstein A. L (1981) *Urbanisation and Kinship: The Domestic Domain on the Copperbelt of Zambia 1950-1956*. London: Academic Press.
- Epstein, A.L. (1969). "The Networks and Urban Social Organization." in *Social Networks in Urban Situations. Analyses of Personal Relationships in Central*

African Towns, edited by J. Clyde Mitchell. Manchester University Press, for the Institute of Social Research, University of Zambia.

Fafchamps, M. (1992), “Solidarity Networks in Pre-Industrial Societies: Rational Peasants with a Moral Economy,” *Economic Development and Cultural Change*, **41**(1): 147-174, October 1992.

Fafchamps M (2004) *Markets and Institutions in Sub-Saharan Africa*. London, England and Cambridge, Massachusetts: The MIT Press.

Fafchamps M and A Quisumbing (2002) ‘Control and ownership of assets within rural Ethiopian households’. *Journal of Development Studies* **38**(2): 47-82.

Fafchamps M and A Quisumbing (2003) Social roles, human capital, and the intrahousehold division of labor: evidence from Pakistan. *Oxford Economic Papers* **55**(1): 36-80.

Fafchamps, M. & S. Lund (2001). ‘Risk-Sharing Networks in Rural Philippines,’ *Economics Series Working Papers 010*, University of Oxford, Department of Economics.

Fafchamps, M. & S. Lund (2003). ‘Risk-Sharing Networks in Rural Philippines,’ *Journal of Development Economics* **71**: 261-87.

Fafchamps, M. and Minten, B. (2002). ‘Returns to Social Network Capital Among Traders’. *Oxford Economic Papers* **54**:173-206.

Fafchamps, M., and F. Gubert. (2007). “The formation of risk sharing networks.” *Journal of Development Economics* **83**:326-350.

Fine B (1999) “The Developmental State is Dead, Long Live Social Capital?” *Development and Change* **30**(1), 1-19.

Fine, Ben. (2001). *Social Capital versus Social Theory: Political Economy and Social Science at the Turn of the Millennium*. London: Routledge.

- Foster, A. D., and M.R. Rosenzweig (1995). 'Learning by doing and learning from others: Human capital and technical change in agriculture'. *Journal of Political Economy* **103**(6): 1176-1209.
- Freeman L.C., A.K. Romney and S.C. Freeman (1987). 'Cognitive structure and informant accuracy'. *American Anthropologist* **89**(2): 310-325.
- Friedland, R. and Robertson, A.F. (1990) 'Beyond the marketplace', in Friedland, R. and Robertson, A.F. (editors) *Beyond the Marketplace. Rethinking Economy and Society*. New York: Aldine de Gruyter.
- Fukuyama, F. (1995). *Trust: Social Virtues and the Creation of Prosperity*. NY: Free Press.
- Furstenberg, F. and Hughes, M. (1995). "Social capital and successful development among at risk youth." *Journal of Marriage and the Family*, **57**: 580-592.
- Furubotn, Eirik G. and Rudolf Richter (1997), *Institutions and Economic Theory: The Contribution of the New Institutional Economics*, Ann Arbor: The University of Michigan Press.
- Gabre-Madhin E. (2001). 'Market Institutions, Transaction Costs and Social Capital in the Ethiopian Grain Market'. *Research Report 124*, International Food Policy Research Institute, Washington D.C.
- Gatter P (1990) Indigenous and Institutional Thought in the Practice of Rural Development: A Study of an Ushi Chiefdom in Luapula, Zambia. *Unpublished PhD Thesis. University of London*.
- Glazier J (1985) *Land and the uses of tradition among the Mbeere of Kenya*. New York: University Press of America.
- Goetz S J (1992) 'A Selectivity Model of Household Food Marketing Behaviour in sub-Saharan Africa'. *American Journal of Agricultural Economics* **74**(2): 444-452.

- Goldstein, Markus Alain de Janvry, and Elisabeth Sadoulet. (2005). "Is a Friend in Need a Friend Indeed? " in Stefan Dercon (ed.) *Insurance against Poverty*, Oxford University Press.
- Govere J, J.J. Shawa, E. Malawo, and T.S. Jayne (2006) Raising The Productivity Of Public Investments In Zambia's Agricultural Sector *Working Paper No. 20 Food Security Research Project Lusaka, Zambia*
- Govere J, T Jayne and Chapoto (2008) 'Assessment Of Alternative Maize Trade Andmarket Policy Interventions In Zambia' *Working Paper No. 33 Food Security Research Project Lusaka, Zambia October, 2008.*
- Granovetter M. (1973) "The Strength of Weak Ties" *American Journal of Sociology* **78**: 1360-1380.
- Granovetter M. (1974) "*Getting a Job: A study of Contacts and Careers*". Harvard: Harvard University Press.
- Granovetter M (1985) Economic Action and Social Structure: The problem of embeddedness. *American Journal of Sociology* **91**(3): 481-510
- Granovetter, M. (1992), "Economic Action and Social Structure: The Problem of Embeddedness" reprinted from AJS 1985 in Granovetter, M. and Swedberg, R. ed. (1992), 53 – 81. *The Sociology of Economic Life*, Boulder, Col.: Westview Press.
- Greene, W. H. (2003). *Econometric Analysis*. 5th Edition. New Jersey NY: Prentice Hall.
- Granovetter M and Swedberg, (1992) (eds.) *The Sociology of Economic Life*, Boulder, Col.: Westview Press.
- Grier, Kevin B., and Gordon Tullock. (1989). An Empirical Analysis of Cross-National Economic Growth, 1951-80. *Journal of Monetary Economics* **24**: 259-76.
- Grootaert, C. (1999). "Social Capital, Household Welfare and Poverty in Indonesia." *Local Level Institution Study Working Paper No. 6*. Washington, D.C: The World Bank.

Grootaert C (2000) 'Social Capital, Household Welfare and Poverty in Indonesia' *Social Capital Working Paper Series # 6* World Bank: Washington DC.

Grootaert Cet, Gi-Taik Oh, and Anand Swamy al, (1999); Social Capital and Development Outcomes in Burkina Faso. *LLI Study Working Paper No. 7*, Social Development Department, World Bank: Washington DC.

Grosch M and P Glewwe (1999) *Designing household survey questionnaires for developing countries: lessons from 15 years of the Living Standards Measurement Study*. Washington D.C.: World Bank.

Haddad L and J Maluccio (2002) "Trust, membership in groups, and household welfare," *FCND discussion papers 135*, International Food Policy Research Institute: Washington DC.

Hanifan L (1916) "The Rural School Community Center". *Annals of the American Academy of Political and Social Science* **67**:130-138.

Hanneman, Robert. (2000) *Introduction to Social Network Methods*.
<http://faculty.ucr.edu/~hanneman/Soc157/TEXT/TextIndex.html>.

Harries-Jones P and J C Chiwale (1962). Kasaka: A Case-Study in Succession and Dynamics of a Bemba village. *Rhodes-Livingstone Journal* **33**:1-67.

Heckman J (1990) Varieties of Selection Bias American Economic Review, American Economic Association **80**(2): 313-18.

Heckman, James J. (1979). "Sample Selection Bias as a Specification Error." *Econometrica* **47**(1): 153-161.

Heltberg R and Tarp F (2002) Agricultural supply response and poverty in Mozambique. *Food Policy* **27** (2002) 103–124.

Hentschel (2003) in Kanbur R. (2003). *Q-Squared: Qualitative and Quantitative Methods of Poverty Appraisal*. Delhi: Permanent Black.

- Hirschliefer J 1973 *Exchange Theory: The Missing Chapter.* *Western Economic Journal (Economic Inquiry)* **11** (June 1973): 129-146.
- Hoang, L.A., J.-C. Castella, and P. Novosad. (2006). Social networks and information access: Implications for agricultural extension in a rice farming community in northern Vietnam. *Agriculture and Human Values* **23**: 513–527.
- Hoff, K. And A. Sen (2006). The Kin System as a Poverty Trap. In: Bowles, S. Durlauf and Hoff (eds) *Poverty Traps*. Chapter 4, pp. 95-115. Princeton University Press, Princeton.
- Holloway Garth Charles Nicholson Chris Delgado Steve Staal Simeon Ehui (2004) ‘A revised Tobit procedure for mitigating bias in the presence of non-zero censoring with an application to milk-market participation in the Ethiopian Highlands’ *Agricultural Economics* **31** (2004) 97–106.
- Holloway Garth , Charles Nicholson, Chris Delgado, Steve Staal, Simeon Ehui (2000) ‘Agroindustrialization through institutional innovation Transaction costs, cooperatives and milk-market development in the east-African highlands’. *Agricultural Economics* **23** (2000) 279–288.
- Homans George.C. (1974) “*Social Behavior*”. Revised edition. New York: Harcourt Brace Jovanovich.
- Homans, George C. (1961) *Social Behaviour: Its Elementary Forms*. New York: Harcourt, Brace & World, Inc.
- Howard, J.A., and C. Mungoma (1996) Zambia’s Stop-and-Go Revolution: The Impact of Policies and Organizations on the Development and Spread of Maize Technology. *MSU International Development Working Paper No. 61*. East Lansing: Michigan State University.
- IFAD (2001). *Rural Poverty Report 2001: The Challenge of Ending Rural Poverty*. Oxford: Oxford University Press.

- Ionnanides, Y., and L.D. Loury (2004) 'Job information networks, neighborhood effects, and inequality'. *Journal of Economic Literature* **42**:1056-1093.
- Isham Jonathan (2002) 'The Effect of Social Capital on Fertiliser Adoption: Evidence from Rural Tanzania' *Journal of African Economies*; **11** (1)39-60.
- Iversen V (2006) Segmentation, Network Multipliers and Spillovers: A Theory of Rural Urban Migration for a Traditional Economy. *Development Research Centre on Migration, Globalisation and Poverty. Working Paper T9*. Brighton: University of Sussex.
- Iversen V, K Sen, A Verschoor, A Dubey (2009) 'Job Recruitment Networks and Migration to Cities in India' *Journal of Development Studies* **45**(4) 522–543.
- Jackson, M. (2007). 'The study of social networks in economics.' Unpublished, forthcoming in *The Missing Link: Formation and Decay of Economics Networks*.
- Jayarathne T.E. and Stewart (1991) 'Quantitative and Qualitative Methods in the Social Sciences' in M. Fonow and J.A. Cook (eds) *Beyond Methodology*. Bloomington and Indianapolis: Indiana University Press.
- Jayne T.S. J Govereh, P. Chilonda, N. Mason, A. Chapoto and H. Haantuba (2007) Trends In Agricultural And Rural Development Indicators In Zambia. *FSRP Working Paper No. 24. Food Security Research Project Lusaka, Zambia June 2007*
- Jayne, T., T. Yamano, M. Weber, D. Tschirley, R. Benfica, A. Chapoto and B. Zulu (2003). 'Smallholder Income and Land Distribution in Africa: Implications for Poverty Reduction Strategies.' *Food Policy* **28**(3): 253-275.
- Kabeer, Naila (2000) 'Social Exclusion, Poverty and Discrimination: Towards an Analytical Framework'. *IDS Bulletin* **31**(4): 83-97.
- Kadushin C (1982). "Social Density and Mental Health". Pp in *Social Structure and Network Analysis*, Edited by L. Nan. P.V. Marsden, Beverly Hills, CA: Sage.

- Kanbur R. (2003). *Q-Squared: Qualitative and Quantitative Methods of Poverty Appraisal*. Delhi: Permanent Black.
- Kay, G (1967) *Social Aspects of Village Regrouping in Zambia*. Lusaka: Institute for Social Research, University of Zambia.
- Key N, E Sadoulet and A de Janvry (2000) 'Transactions Costs and Agricultural Household Supply Response'. *American Journal of Agricultural Economics* **82** (May 2000): 245-259.
- Knack, S. and Keefer, P. (1997). "Does social capital have an economic impact? A crosscountry investigation." *Quarterly Journal of Economics*, vol. **CXII**, no. 4, pp. 1252-88.
- Kormendi, Roger C., and Philip G. Meguire (1985). Macroeconomic Determinants of Growth. *Journal of Monetary Economics* **16**: 141-163.
- Kranton R.E. and D.F. Minehart (2001). "A Theory of Buyer-Seller Networks". *The American Economic Review* **91**(3), 485-508.
- Krishna (2001) 'Moving from the Stock of Social Capital to the Flow of Benefits: The Role of Agency.' *World Development* **29**: 925-943.
- Krishnan P And E Sciubba (2009) 'Links And Architecture In Village Networks'. *The Economic Journal*, **119**(4) (April), 917-949.
- Leavy J and C Poulton (2007) "Commercialisations in Agriculture: A typology" *Ethiopian Journal of Economics* **XVI**(1) 3-41.
- LeCompte, M.D., and Goetz, J.P. (1982). 'Problems of reliability and validity in ethnographic research.' *Review of Educational Research* **52**(2), pp. 31-60.
- Lewis, W. (1955) *The Theory of Economic Growth* (London: George Allen & Unwin).
- Light I., P. Bachu, and S.Karageorgis (1993). "Migration Networks and Immigrant Entrepreneurship". in I.Light and P. Bachhu (eds) *Immigration and Entrepreneurship*. New Brunswick, NJ: Transaction.

- Lin N, W M Ensel and J C Vaughan (1981) "Social Resources and Strength of Ties: Structural Factors in Occupational Status Attainment". *American Sociological Review* **46**: 393-405.
- Lin N. (2001a) *Social Capital – a theory of social structure and action*. Cambridge: Cambridge University Press.
- Lin N. (2001b). 'Building a network theory of social capital' in Lin N, K Cook, RS Burt (Eds) *Social Capital: theory and research*. New York: Aldine de Gruyter.
- Lin, N. (2008). 'A network theory of social capital'. In *The Handbook of Social Capital*, Castiglione, D., Van Deth, J. W., & Wolleb, G., eds., 50-69. New York: Oxford.
- Lin N & B Erikson (2008) eds *Social Capital: an international research programme*. Oxford:OUP.
- Lin, N, K Cook and RS Burt (eds) (2001) *social capital: theory and research* hawthorn NY: aldine de gruyter)
- Lin, N. and Dumin, M. (1986) 'Access to occupations through social ties.' *Social Networks* **8**:365-85.
- Lincoln, Y.S., and Guba, E.G., (1985). *Naturalistic Inquiry*. Newbury Park, CA: Sage Publications,.
- Lipton, M (1992) 'Economics and anthropology: Grounding models in relationships'. *World Development* 20(10): 1541-1546.
- Long J S and J Freese (2006). *Regression Models for Categorical Dependent Variables Using Stata*, 2nd Edition. Stata Press.
- Long Norman (1968) *Social Change and the Individual; A Study of the Social and Religious Responses to Innovation in a Zambian Rural Community*. Manchester: Manchester University Press.
- Longhurst R. (1988). Cash crops, Household food security and nutrition. *IDS Bulletin* **19**(2):28-36.

- Lucas Robert E B (1997) 'Internal Migration in Developing Countries'. In *Handbook of Population and Family Economics Vol 1b*, Edited by M R Rosenzweig and O Stark Elsevier Science.
- Maddala, G S (1983) Limited-Dependent and Qualitative Variables in Econometrics. Econometric Society Monographs. Cambridge: CUP.
- Mafeje, A., (1976). "The Problem of Anthropology in Historical Perspective: An Inquiry into the Growth of the Social Sciences", *Canadian Journal of African Studies / Revue Canadienne Des Études Africaines*, **10**(2):307-333.
- Magubane, B. (1971) "A Critical Look at Indices Used in the Study of Social Change in Colonial Africa", *Current Anthropology*, **12**(4/5):419-445.
- Maluccio, J. A., L. Haddad, and J. May. (2000). 'Social capital and household welfare in South Africa 1993–1998'. *Journal of Development Studies* **36** (6): 54–81.
- Marin, A (2004) 'Are respondents more likely to list alters with certain characteristics? Implications for name generator data', *Social Networks*, **26**:289-307.
- Marsden P.V. (1990) "'Network Data and Measurement *Annual Review of Sociology* **16**:435-463.
- Marsden P V (1987) "Core Discussion Networks of Americans" *American Sociological Review* **52**(1):122-131.
- Marsden P and K Campbell (1984) 'Measuring Tie Strength'. *Social Forces* **63** (2): 482-501.
- Massey D.S. (1990) "Social Structure, Household Strategies and the Cumulative Causation of Migration", *Population Index*, Spring 1990, **56**(1), pp. 3-26.
- Massey, D. and F. Garcia Espana (1987). 'The Social Process of International Migration,' *Science* **237**: 733-38.
- McNetting, Robert (1993) *Smallholders, Householders. Farms families and the ecology of intensive, sustainable agriculture*. Stanford: Stanford University Press.

- Merleau-Ponty M. (1962) *The Phenomenology of Perception*. London: Routledge.
- Minot N (1999) Effect Of Transaction Costs On Supply Response And Marketed Surplus: Simulations Using Non-Separable Household Models *MSSD Discussion Paper No. 36*. IFPRI: Washington DC.
- Mitchell, J Clyde (1969) ed *Social Networks in Urban Situations: Analysis of Personal Relationships in Central African Towns* Manchester: Manchester University Press.
- Mitchell J.C. (1974) "Social Networks". *Annual Review of Anthropology* **3**(1974) 279-299.
- Montgomery J. (1991). "Social Networks and Labor-Market outcomes: Toward an Economic Analysis". *The American Economic Review* **81**(5), 1408-1418.
- Montgomery J. (1992) "Job Search and Network Composition: Implications of the Strength-of Weak Ties Hypothesis". *American Sociological Review* **57**: 586-596.
- Moore H.L and M. Vaughan (1994) *Cutting Down Trees*. Portsmouth: Heinemann.
- Moore Jeffrey C., Linda L. Stinson, and Edward J. Welniak, Jr. (2000) Income Measurement Error in Surveys: A Review *Journal of Official Statistics* **16**(4): 331–361.
- Mortensen, D. and T. Vishwanath 1994. 'Personal Contacts and Earnings, It's Who You Know,' *Labor Economics* **1**: 177-201.
- Moser, C. and Dani, A. (2008).eds. *Assets, Livelihoods and Social Policy* (Chapter 1). Washington, DC: World Bank.
- Müller C, B Wellman and A Marin (1999). "How to use SPSS to study ego-centred networks". *Bulletin des Méthodologies Sociologiques* **64**: 63-76.
- Munshi, K. (2003) 'Networks in the modern economy: Mexican migrants in the US labour market'. *Quarterly Journal of Economics*, **118**(2), pp. 549–599.

- Murgai, R., Winters, P., Sadoulet, E., de Janvry, A., (2002). 'Localized and Incomplete Mutual Insurance'. *Journal of Development Economics* **67**, 245–274.
- Murray, Michael P.(2006) "Avoiding Invalid Instruments and Coping with Weak Instruments," *Journal of Economic Perspectives*, 2006, vol. 20, issue 4, pages 111-132.
- Myrdal G. (1957). "*Economic Theory and Underdeveloped Regions*". London: Duckworth.
- Narayan, D. and Pritchett, L. (1999). "Cents and sociability: household income and social capital in rural Tanzania." *Economic Development and Social Change*, **47**(4), pp. 871-897.
- Narayan, Deepa, (1997), *Voices of the poor: Poverty and Social Capital in Tanzania*," *ESD monograph series*. Word Bank: Washington DC.
- North D. (1995). 'The New Institutional Economics and Third World Development'; pp17-26 in Harriss J., J Hunter and C.M Lewis (eds) *The New Institutional Economics and Third World Development*. London and New York: Routledge.
- OECD (2006) *Promoting Pro-Poor Growth: Private Sector Development*. Paris: OECD.
- Okali, C (1983) 'Kinship and Cocoa Farming in Ghana'. In C Oppong ed *Female and Male in West Africa*. London; George Allen and Unwin.
- Omamo S W (1998a) 'Farm-to-Market Transaction costs and specialisation in small-scale agriculture: explorations with a non-separable Household Model'. *Journal of Development Studies* **35**:152-163.
- Omamo S W (1998b) 'Transport Costs and smallholder Cropping choices: An Application to Siaya District, Kenya'. *American Journal of Agricultural Economics* **80**(1) 116-123.
- Patton, M. Q. (1990) *Qualitative Evaluation and Research Methods*, 2nd edition. Newbury Park, Calif.: Sage Publications, 1990.

- Pingali, P. L (1997). 'From subsistence to commercial production systems: the transformation of Asian agriculture.' *American Journal of Agricultural Economics* **79(2)**: 628 - 634.
- Pingali P L and M W Rosegrant (1995) 'Agricultural Commercialisation and Diversification: Processes and Policies'. *Food Policy* **20(3)**: 171-185.
- Pingali P, Yasmeen Khwaja and Madelon Meijer (2005) 'Commercializing Small Farms: Reducing Transaction Costs'. *ESA Working Paper No. 05-08* October 2005. Rome: Food and Agriculture Organisation of the United Nations (FAO).
- Platteau, J (1991)., "Traditional Systems of Social Security and Hunger Insurance: Past Achievements and Modern Challenges," in *Social Security in Developing Countries*, E. Ahmad, J. Dreze, J. Hills, and A. Sen (eds.), Clarendon Press, Oxford, 1991.
- Poewe K O (1978) 'Matriliney in the throes of change. Kinship, descent and marriage in Luapula, Zambia'. *Africa* **48(3)** 205-219.
- Poewe K O (1979) 'Regional and village economic activities. Prosperity and stagnation in Luapula, Zambia'. *African Studies Review* **22(2)** 77-93.
- Polanyi K (1944), *The Great Transformation*. Beacon Press.
- Portes, A (1998). Social capital: its origins and applications in modern sociology. *Annual Review of Sociology* **24**, 1-24.
- Portes, A., Landolt, P.(1996). The Downside of Social Capital. *The American Prospect*, **26**: 18-23.
- Portes, Alejandro. (1998). "Social capital: its origins and applications in modern sociology." *Annual Review of Sociology* **24**: 1-25.
- Pottier J (1988). *Migrants No More: settlement and survival in Mambwe villages, Zambia*. Manchester: Manchester University Press.

- Powell and Smith-Doerr (2005) "Networks and Economic Life." In *The Handbook of Economic Sociology*, Second Edition, edited by N.J. Smelser and R. Swedberg. Princeton, NJ: Russell Sage Foundation/Princeton University Press. pp. 379-402.
- Putnam R (1995) Bowling Alone: The Strange Disappearance of Civic America. *Journal of Democracy* **6**:65-78.
- Rauch and Hamilton (2001) Chapter 1 in Rauch and Casella, 2001, *Networks and Markets* New York: Russell Sage Foundation.
- Richards A.I. (1939) "*Land, Labour and Diet in Northern Rhodesia: An Economic Study of the Bemba Tribe*". Munster; Hamburg : LIT 1995.
- Rietbergen-McCracken J and D Narayan (1998) *Participation and Social Assessment – Tools and Techniques*. Washington D.C.: IBRD/World Bank.
- Rob, R., Zemesky, P.(2002). Social Capital, corporate culture, and incentive intensity. *Rand Journal of Economics*. **33**, 243-257.
- Routledge, Bryan R. & von Amsberg, Joachim, 2003. "Social capital and growth," *Journal of Monetary Economics* **50**(1), pages 167-193.
- Sadoulet E and A de Janvry (1995) *Quantitative Development Policy Analysis* Baltimore: Johns Hopkins University Press.
- Sadoulet, E., A.de Janvry, and S. Fukui (1997) 'The Meaning of Kinship in Sharecropping Contracts,' *American Journal of Agricultural Economics* **79**: 394-406.
- Santos P and C Barrett (2008) 'Identity, Interest and Information Search in a Dynamic Rural Economy' *Published as* Santos P and C Barrett (2010) 'Identity, Interest and Information Search in a Dynamic Rural Economy' *World Development*, **38**(12), 2010.
- Scott J. (1991) "*Social Network Analysis – A Handbook*". London; Newbury Park; New Delhi: Sage Publications.

- Scully, Gerry. (1988). 'The Institutional Framework and Economic Development'. *Journal of Political Economy* **96**: 652-662.
- Sharp K., S. Devereux and Yared Amare (2003). 'Destitution in Ethiopia's Northeastern Highlands (Amhara National Regional State), *Final Project Report for Save the Children UK, Institute of Development Studies, Sussex*.
- Simon, H (1961) *Administrative Behaviour*. Macmillan New York.
- Skvoretz J and F Agneessens, 2007 'Reciprocity, multiplexity, and exchange: Measures'. *Quality and Quantity* **41**: 341-357.
- Sobel, Joel. (2002). "Can We Trust Social Capital?" *Journal of Economic Literature* **B**:139-154.
- Solow, R. (2000). Notes on social capital and economic performance. In: Dasgupta, P., Serageldin, I. (Eds.), *Social Capital: A Multifaceted Perspective*. The World Bank, Washington, DC.
- Spielman, David J. & Davis, Kristin E. & Negash, Martha & Ayele, Gezahegn (2008). "Rural innovation systems and networks: Findings from a study of Ethiopian smallholders," *IFPRI discussion papers 759*, International Food Policy Research Institute (IFPRI).
- Staal S, C. Delgado and C. Nicholson (1997); Smallholder Dairy under Transactions Costs in East Africa. *World Development* **25**(5): 779-794.
- Stiglitz, J.E. (1974) Incentives and risk sharing in sharecropping, *Review of Economic Studies*, **95**, 219-255.
- Stiglitz J (1986), "The New Development Economics," *World Development*, **14**(2), 257-265, February.
- Stiglitz, J.E. (1989) 'Rational peasants, efficient institutions, and a theory of rural organization: methodological remarks for development economics'. In: Bardhan, P. (Ed.), (1989) *The Economic Theory of Agrarian Institutions*, Clarendon Press: Oxford.

- Stock James H. and Motohiro Yogo (2002) "Testing for Weak Instruments in Linear IV Regression," NBER Technical Working Papers 0284, National Bureau of Economic Research, Inc.
- Strasberg Paul J., T. S. Jayne, T Yamano, J Nyoro, DKaranja, and J Strauss (1999) Effects Of Agricultural Commercialization On Food Crop Input Use And Productivity In Kenya. *MSU International Development Working Paper No 71*, 1999. Michigan State University.
- Strauss, J (1984). Marketed surplus of agricultural households in Sierra Leone. *American Journal of Agricultural Economics* **66**(3), 321–331.
- Stromgaard P (1985) ‘A subsistence society under pressure’. *Africa* **55**(1): 39-58.
- Swedberg, R (1991) “Main Traditions of Economic Sociology”, *Annual Review of Sociology*, **17**(1991):251-76.
- Szreter, S., Woolcock, M. (2002), ‘Health by Association? Social Capital, Social Theory and The Political Economy of Public Health’, *von Hugel Institute Working Paper*, WP2002-13.
- Tawney, R. H. (1926). Religion and the Rise of Capitalism, New York: Harper and Row.
- Timmer C. P. (1997). Farmers and Markets: The Political Economy of New Paradigms. *American Journal of Agricultural Economics* **79**(2): 621-627.
- Torpe, Lars. (2003). "Social capital in Denmark: a deviant case?" *Scandinavian Political Studies* **26**: 27-48.
- Tracy E.M., R.F. Catalano, J.K. Whittaker and D. Fine (1990) ‘Reliability of Social Network Data’. *Social Work Research and Abstracts* **26**: 33-35.
- UNDP (2003) *Zambia Human Development Report*. Lusaka: United Nations Development Program.

- Van der Gaag, Martin and Snijders, Tom (2005) 'The resource generator', *Social Networks* **27**:1-27.
- Van der Gaag, Martin P.J. and Snijders, Tom A.B. (2003a) 'An approach to the measurement of individual social capital'. pp. 199-218 in H. Flap and B. Völker (eds.), *Creation and Returns of Social Capital*. London: Routledge, 2003.
- Van der Gaag, Martin P.J. and Snijders, Tom A.B. (2003b) 'A comparison of measures for individual social capital' *University of Groningen/ Vrije Universiteit Amsterdam*. 27th November 2003.
- Varughese, G. and E. Ostrom (2001) The contested role of heterogeneity in collective action: Some evidence from community forestry in Nepal. *World Development* **29** (5): 747-765.
- Vaughan M (1998) 'Exploitation and neglect: rural producers and the state in Malawi and Zambia' in Birmingham D and P Martin (eds.) *History of Central Africa: The Contemporary Years, since 1960*. London and New York; Longman.
- Von Braun (1995) 'Agricultural Commercialization: impacts on income, nutrition and implications for policy'. *Food Policy* **20**(3): 87-202.
- Von Braun J. and E. Kennedy (1986). 'Commercialization of Subsistence Agriculture: Income and Nutrition effects in Developing Countries'. *Working paper on commercialization of agriculture and Nutrition I*. International Food Policy Research Institute, Washington DC.
- Von Braun, J. & Kennedy, E. (eds.) (1994) *Commercialization of Agriculture, Economic Development and Nutrition*, John Hopkins Press, Baltimore, MD.
- Vyas S and L Kumaranayake (2006) Constructing socio-economic status indices: how to use principal components analysis *Health Policy and Planning*. 21(6): **459-468**.
- Wahba, J. and Zenou, Y. (2005) Density, social networks and job search methods: theory and application to Egypt. *Journal of Development Economics*, **78**(2), pp. 443-473.

- Warren, M R, (2005), Communities and schools: a new view of urban education reform, *Harvard Education Review*, **75** (2), 133–75.
- Wasserman S. and K. Faust (1994) *Social Network Analysis: Methods and Applications* Cambridge: Cambridge University Press.
- Wasserman S. and K. Faust (1994) *Social Network Analysis: Methods and Applications* Cambridge: Cambridge University Press.
- Watson W (1958) *Tribal Cohesion in a Money Economy: A Study of the Mambwe People of Northern Rhodesia* Manchester: Manchester University Press.
- Weber, M. (1930) [1904/5]. The Protestant Ethic and the Spirit of Capitalism, Translated by Talcott Parsons, with and introduction by Anthony Giddens, 1992, London and New York: Routledge.
- White H (2002) 'Combining Quantitative and Qualitative Approaches in Poverty Analysis', *World Development*, **30**(3): 511-522.
- Williamson, Oliver. (1975). *Markets and Hierarchies*. New York: Free Press.
- Williamson, Oliver (1979) 'Transaction Cost Economics: The governance of Contractual Arrangements' *Journal of Law and Economics* **22** (Oct): 233-261.
- Williamson, Oliver. (1985). *The Economic Institutions of Capitalism*. New York: Free Press.
- Williamson, Oliver. (2000). "The New Institutional Economics: taking stock, looking ahead." *Journal of Economic Literature* **38** (3): 595-613.
- Winters, P., A. de Janvry and E. Sadoulet (2001). 'Family and Community Networks in Mexico-US Migration,' *Journal of Human Resources* **36** (1): 159-85.
- Wood A and E Shula (1987) 'The state and agriculture in Zambia' in N Bourenane and T Mkandawire (eds) *The state and agriculture in Africa*. Dakar: CODESRIA.

- Woolcock M. (2001). "The Place of Social Capital in Understanding Social and Economic Outcomes". *ISUMA Canadian Journal of Policy Research* (Spring 2001),11-17.
- Woolcock, M. (2003): Social Capital. In: Christensen, K. & Levinson, D. (Editors), *Encyclopedia of Community: From the Village to the Virtual World*. Sage, Thousand Oaks et al., pp. 1258-1262.
- Woolcock M. and D Narayan (2000). 'Social Capital: Implications for Development Theory, Research and Policy'. *The World Bank Research Observer* **15**(2), 225-249.
- Wooldridge J (2002) *Econometric Analysis of Cross Section and Panel Data*, MIT Press.
- World Bank (1998) *Republic of Uganda Global Social Capital Survey*. Washington D.C.: World Bank.
- World Bank (2002) *World Development Report 2002*. Washington D.C: The World Bank.
- World Bank (2007) *World Development Indicators* Washington DC: World Bank.
- World Bank (2008) *World Development Report 2008*. Washington D.C: The World Bank.
- Zaibet L.T. and E.G. Dunn (1998) 'Land Tenure, Farm Size and Rural Market participation in Developing Countries: The case of the Tunisian Olive Sector'. *Economic Development and Cultural Change* **46**(4): 831-848.
- Zak, Paul J. and Stephen Knack (2001) 'Trust and Growth'. *Economic Journal*, **111**(470), March 2001.

Appendices

Appendix A: Social Networks Survey Module

7. In this part of the questionnaire we are interested in the people you know.

7.1 First, do you interact with people outside of your immediate household?

0 No ⇒ Qn 7.8
1 Yes ⇒ next question

Who are the non-household members you interact with, socialising and/or working, most frequently? FIRST LIST NAMES OF ALL PEOPLE MENTIONED FOR QUESTION 7.2, THEN GO ON TO ASK QUESTIONS 7.3 TO 7.7 ABOUT EACH PERSON MENTIONED

PERSON ID	7.2 Please can you give me the names of all the people you interact with? 7.2a TOTAL NUMBER OF PEOPLE:		7.3 Sex 1 Male 2 Female	7.4 Age Estimated Years	7.5 What do you do with these people? ACTIVITY CODES (RECORD MAXIMUM OF 3. LET A1 BE ACTIVITY DONE MOST OFTEN)			7.6 Thinking about the activity you do most often with [NAME], how often do you do this? 1 Daily 2 At least once a week 3 Once a month 4 Less often 5 Other	7.7 What is the relationship of [NAME] to you? RECORD MAXIMUM OF TWO 1 She or he is a friend 2 She or he is my neighbour 3 She or he is related (but not household member) 4 She or he belongs to an organisation I belong to (e.g. Church) 5 She or he employs me 6 She or he is employed by me 7 Other (specify) R1 R2	
	First Name	Last Name			A1	A2	A3			
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										

Activity Codes:

- | | | |
|---|------------------|--------------------|
| 1 Eat meals with outside your home | 5 Play games | 10 Work |
| 2 People who visit you at your home or you visit them | 6 Music/dance | 11 Trade |
| 3 Attend religious festivals together | 7 Do arts/crafts | 12 Chores |
| 4 Organise festivals/celebrations | 8 Talking | 13 Other (specify) |
| | 9 Drinking | |

7.8 In your opinion who are the three most important people in your community, people that the community looks to for leadership?

NAME 1		NAME 3	
NAME 2			

7.9 Do you have a friend/kin who: owns a business/ heads a committee/ community leader/ other influential position?

0 No ⇒ Qn 7.14
1 Yes ⇒ next question

[RECORD MAXIMUM OF THREE PEOPLE]

7.10 Name	7.11 What is [NAME'S] Position?	7.12 What is [NAME'S] relationship to you? 1 Father/Mother 2 Son/Daughter 3 Brother/Sister 4 Other relative/kin 5 Friend 6 Other	7.13 where does [NAME] live? 1 In this village 2 Another rural area (name place) 3 Urban area (name place) CODE NAME OF PLACE
P1			
P2			
P3			

PARTICIPATION IN GROUPS: Now I would like to ask you about your involvement in local groups and clubs, both formal and informal.

	Church / other religious organisation	Neighbourhood association/ group	Cooperative	Work group	Women's group	Leisure club	Farmers' group	Traders association/ business group	Credit/finance group	Youth group	Basic services group (eg health, education, roads, water)	Special Projects (1) - Specify	Special Projects (2) - Specify
GROUP CODE:													
7.14 Do you belong to any of these groups? 0 I do not belong to such a group 1 I belong but never participate 2 I have participated once or twice 3 I have participated more than twice 99 group does not exist in this community													
PROMPT FOR INFORMAL GROUPS, IF NO GROUP MEMBERSHIP ⇒ 7.22 IF NO GROUPS IN COMMUNITY ⇒ 7.25													

RECORD GROUP CODE OF GROUPS RESPONDENT IS A MEMBER OF INTO THIS COLUMN. PROMPT FOR REGULAR, INFORMAL GATHERINGS	7.15 How long have you been a member of [NAME] group? MM YY	7.16 To what extent do you participate in group's decision- making? 0 Not at all 1 To a very small extent 2 To a small extent 3 Neither small nor large extent 4 To a large extent 5 To a very large extent	7.17 What is the name of the group leader? IF RESPONDENT IS GROUP LEADER THEN RECORD CODE 0 99 Not applicable Name	7.18 Do you contribute money to the group? 0 No IF Yes THEN ASK: How much per year? Kwacha	7.19 How many group members are there? Number

CODE

7.20 IF DO NOT BELONG TO ANY GROUPS: Many obstacles keep people from becoming as involved in their community as they would like. Thinking about your own life, are there any obstacles or barriers that make it difficult for you to be as involved in your community as you would like, or not?	0 No ⇒ 7.23 1 Yes ⇒ 7.21 2 Do not know ⇒ 7.21 99 No answer ⇒ 7.21	
7.21 Are there any groups which you are not a member of but would like to be?	0 No ⇒ 7.23 1 Yes ⇒ 7.22 2 Do not know ⇒ 7.23 99 No answer ⇒ 7.23	
7.22 What is the most important obstacle to your involvement in groups? 1 inadequate childcare 2 short of money to pay for what they need 3 lack of info or not knowing how to begin. 4 gender 5 age 6 feeling unwelcome 7 hard to find time to get together 8 Other (specify)		
7.23 In the past two years have you worked together with others in your immediate community to try and get someone to fix or improve something or to address a common issue?	0 No ⇒ 7.25 1 Yes ⇒ next question 2 Do not know ⇒ 7.25 88 No answer ⇒ 7.25	
7.24 What did you do? 1 Road Improvement 2 Health Issue 3 Worked on school 4 Applied for funds for your community or group 5 Work on water supply 6 Conservation work 7 Animal work 8 Other (Specify)		

SOCIAL SUPPORT: I am now going to read out some situations where you may need help from other people. Who are you mostly likely to turn to for help in each of the situations I am about to read out to you? This could be INDIVIDUALS OR GROUPS including family, neighbours, friends, priest, teacher, doctor, someone you pay to help, traditional healer, kin, migrant family members living elsewhere, government or social services agency, community or church groups, money lender, bank or credit union.

Who would you go to if you...		7.25 Name of Individual or group: IF INDIVIDUAL :CHECK TO SEE IF INDIVIDUAL IS NAMED IN 7.2. IF SO ⇒ 7.29 IF NOT ⇒ next question IF GROUP ⇒ 7.29 IF NO-ONE RECORD CODE 00	7.26 Sex: 1 Male 2 Female	7.27 Age EST'D YEARS	7.28 What is [NAME'S] relationship to you? 1 She or he is a friend 2 She or he is my neighbour 3 household member 4 She or he is related (but not household member) 5 She or he belongs to an organisation I belong to (e.g. Church) 6 She or he employs me 7 She or he is employed by me 8 She or he is my patron/kangalila	7.29 What other kinds of help could you get from [NAME]? 1 Advice on important decision 2 Advice/ information on farming matters 3 Advice/info on work-related matters 4 Advice/info on personal matters 5 Filling out forms 6 gift 7 Finding work 8 Lend money 9 No other advice 10 other [RECORD MAX OF 3] A1 A2 A3
A ...needed a lift urgently, to go to the health centre, say.	1					
	2					
	3					
	4					
	5					
B if you were ill in bed and needed help at home	1					
	2					
	3					
	4					
	5					
C If you needed to borrow quite a large sum of	1					
	2					
	3					
	4					

Who would you go to if you:		7.25 Name of Individual or group: IF INDIVIDUAL :CHECK TO SEE IF INDIVIDUAL IS NAMED IN 7.2. IF SO ⇒ 7.29 IF NOT ⇒ next question IF GROUP ⇒ 7.29 IF NO-ONE RECORD CODE 00 ⇒ next situation	7.26 Sex: 1 Male 2 Female	7.27 Age YEARS	7.28 What is [NAME'S] relationship to you? 1 She or he is a friend 2 She or he is my neighbour 3 household member 4 She or he is related (but not household member) 5 She or he belongs to an organisation I belong to (e.g. Church) 6 She or he employs me 7 She or he is employed by me 8 She or he is my patron/ kangalila	7.29 What other kinds of help could you get from [NAME]? 1 Advice on important decision 2 Advice/ information on farming matters 3 Advice/info on work-related matters 4 Advice/info on personal matters 5 Filling out forms 6 gift 7 Finding work 8 Lend money 9 No other advice 10 other [RECORD MAX OF 3] A1 A2 A3		
	3							
L want information about cultivation techniques	1							
	2							
	3							
M want information about ew crops	1							
	2							
	3							
N want information about farming schemes e.g. Outgrower, PAM.	1							
	2							
	3							
O want information on health matters	1							
	2							
	3							

Appendix B Descriptive Statistics: Social Networks and Respondent Attributes

The analysis presented here sets out relationships between respondent networks, using the categorical variable versions of the social networks variables described in Chapter 4, and other individual attributes, reporting only those that are statistically significant.

Network Size

Those with a male household head tend to report larger networks...

Male household heads are slightly over-represented in larger networks, under-represented in the smaller category (5 or less; see Table A4.3.1). Most respondents with a female household head have a network size of less than ten people (72.9 per cent of those with a female household head).

Table B1 Network Size, and i) Sex of household head; ii) Education level (% of respondents)

Network Size	Sex of household head		Level of education		
	male	female	Up to Grade 4	Grades 5-7	Secondary+
5 or less	71.2	28.8	43.2	47.7	9.10
6-10	88.0	12.0	22.3	52.1	25.5
11-15	86.7	13.3	21.4	57.1	21.4
>15	88.2	11.8	14.7	44.1	41.2
Total Sample (%)	84.0	16.0	25.2	50.9	23.8

Source: Household Survey, 2002-03

...as do those with a higher level of education

Table B1 presents also the cross-tabulation of network size by educational level of respondent. In general, people with a secondary-level education are over-represented in the larger network size category (27.5 per cent of the >15 category, compared with 23.8 per cent of the sample), and conversely, under-represented in the smaller network size category. Those with lower levels of education tend to have smaller networks, 43.2 per cent of people with education up to grade 4 fall in the 5 networks partners or less category compared with 25.2 per cent of the sample as a whole, with those with education to the level of grades 5 to 7 under-represented in the two tails. Network size therefore appears to be increasing with level of education. There are exceptions: In Lufubu Resettlement Scheme, while people with secondary education are over-represented in larger networks (three out of the five households), the person with largest stated network of all has education to grade 3-4.

People who have not always lived in their particular community tend to have larger networks

In all three communities a network size of more than 20 people coincides with not having always lived in that community (and the cross-tabulation, see Table A4.3.2, is also highly significant). In the sample as whole most people stated that they had always lived in their community, but a sizeable minority (35.7 per cent of respondents) claimed to have not always lived there, and these people are over-represented in networks of >15 people: 68.4 per cent with networks of 15 people or more had not always lived in that community. The same pattern emerges when the data are broken down by community. It is important to consider here what exactly “always lived here” means. From qualitative fieldwork it was evident that a response of ‘no’ to the question “Have you always lived here?” captures those who “went and came back”. This is entirely expected given the Province’s status, especially traditionally, as a labour reserve,

with a long history of migration from the area, for example to work in the Copperbelt, returning to 'the village' (although not necessarily the same village) on retirement.

Table B.2 Network Size and i) Residence status; ii) Sex of respondent (% of respondents)

Network Size	Always lived here?		Sex of respondent	
	no	yes	male	female
5 or less	24.2	75.8	36.5	63.5
6-10	22.0	78.0	48.0	52.0
11-15	51.9	48.1	62.2	37.8
>15	68.4	31.6	67.6	32.4
Total Sample (%)	35.7	64.3	51.1	48.9

Source: Household Survey, 2002-03

The nature of Lufubu as a resettlement scheme means that one would expect a far larger percentage of the people surveyed to have not always lived there compared to the other two study sites, and when the data are broken down by community this is borne out: 57.5 per cent of Lufubu respondents had not always lived there compared with 33.3 and 21.4 per cent in Ngulula and Kabila respectively. Nevertheless, at first glance this is not as many as one would expect given that most people have been allocated their plots since 1991. Respondents' interpretation of the meaning of 'here' to mean roughly that area rather than the scheme per se, provides an explanation: Settlers are not confined to those retrenched from railway and other industries, who have a high chance of originating from elsewhere, many people are settling in the scheme from neighbouring villages.

Males report larger networks than females

In the sample there are roughly equal numbers of men and women (slightly more men than women: 53.3 per cent male against 46.7 per cent female). Overall, male respondents are over-represented in the larger network categories and women in the smaller – see Table 4.18. Women make up 63.5 per cent of smallest network category and men 67.6 per cent of the >15 network partners category. When the data are broken down by community, once again the same pattern is evident.

Married people are more likely to report bigger networks than their unmarried counterparts

Most respondents are married (84.8 per cent) and those who are married are more likely to have bigger networks, making up 94.1 per cent of those with network size greater than 15 people (Table B.3). No unmarried people have more than 20 network partners. Married people are over-represented in all categories apart from the smallest (five or fewer network partners), 28.8 per cent of whom are unmarried compared with only 15.2 per cent of the sample as a whole. If larger networks are indeed more valuable than smaller networks in terms of accessing resources, in particular labour, especially in contexts of relatively thin or underdeveloped labour markets where personal connections perform the role of more marketised or institutionalised exchange of labour, then this has potential negative implications for households headed by single people, in particular female-headed households in the study site settings where access to male labour during key points in the cultivation season is crucial for carrying out essential, heavier (and therefore gendered) agricultural tasks. It raises important questions of how to counteract a socially-embedded 'disadvantage' in terms of vital resource access.

Table B.3 Network Size and Marital status (% of respondents)

Network Size	marital status	
	married	not married
5 or less	71.2	28.8
6-10	87.0	13.0

11-15	88.9	11.1
>15	94.1	5.90
Total Sample (%)	84.8	15.2

Source: Household Survey, 2002-03

Employers of group labour have larger networks

Across the sample as whole, employers of group labour tend to have larger networks (70.6 per cent of those with networks of greater than 15 network partners compared with 41.6 per cent of sample). Three-quarters of those with 5 or fewer network partners are not employers of group labour compared with 58.4 per cent of sample (Table B.4).

Table B.4 Network Size and Group labour involvement (% of respondents)

Network Size	employer of group labour	
	No	Yes
5 or less	75.0	25.0
6-10	67.0	33.0
11-15	42.2	57.8
>15	29.4	70.6
Total Sample (%)	58.4	41.6

Source: Household Survey, 2002-03

Related to the previous point linking network size and access to crucial resources, it follows that this potentially represents a reinforcement of the inherent relative disadvantage of households headed by females with smaller networks in accessing the right kinds of labour, in particular for heavy, land-clearing tasks, at the appropriate point in the season, with knock-on effects on the ability of the household to cultivate a marketable surplus.

Group affiliation and respondent attributes

Turning now to correlations between group membership and other respondent attributes, there are significant relationships between group affiliation and a respondent's age, as well as group affiliation and heterogeneity of network partners.

Older respondents are less likely to belong to community-level groups

From the survey data the majority of respondents are members of at least one village-level group or society including church groups, with only twenty-two of the 231 respondents not belonging to any group whatsoever. When cross-tabulated with the respondent attribute data the group affiliation dummy variable is significantly correlated (at the 5 per cent level) with just one attribute variable: age of respondent. From Table B.5 one can see that older respondents are more likely to NOT be members of groups than their share of the sample would suggest. The over-60s make up 30 per cent of the non-member category compared with just 10.5 per cent of the sample as a whole. The remaining two age groups follow closely the sample distribution.

Table B.5 Group affiliation and respondent age (% of respondents)

group affiliation dummy	Age of respondent		
	30 and under	31-60	>60
No	30.0	40.0	30.0
Yes	28.2	63.2	8.6
Total Sample (%)	28.4	61.1	10.5

Source: Household Survey, 2002-03

Networks of group non-participants are less diverse

Group affiliation correlates significantly with the network heterogeneity variables. Those not in groups are more likely to have less diverse networks in terms of both age and sex of network partners (see Table B.6 below).

Table B.6 Group affiliation and heterogeneity of respondent networks (% of respondents)

group affiliation dummy	Age heterogeneity				Sex heterogeneity		
	<5	5-<10	10-<15	15+	<0.01	0.01-0.9	>0.9
No	37.5	31.3	12.5	18.8	45.5	36.4	18.2
Yes	8.30	39.2	36.3	16.2	19.1	55.0	25.8
Total Sample (%)	10.5	38.6	34.5	16.4	21.6	53.2	25.1

Source: Household Survey, 2002-03

Group affiliation and the variables for participation in and employment of group labour, however, are not significantly correlated. This is surprising: qualitative data from the research suggest that group members get a better rate when employing a work group from their club or society. Following from this, one might assume that people would be more likely to employ these work groups. The data do not appear to bear this out.

Prestige Networks and respondent attributes

Other significant associations between the prestige network dummy variable and other respondent attributes are given in Table B.7:

Table B.7 Prestige networks and i) sex of respondent; ii) status in household; iii) employer (% of respondents)

Influential kin dummy	Sex of respondent		Respondent is household head	Respondent is an employer
	male	female		
No	40.2	59.8	48.4	39.1
Yes	61.2	38.8	65.9	55.0
Total Sample (%)	52.2	47.8	58.5	48.2

Source: Household Survey, 2002-03

Respondents reporting an influential friend or kin in their social networks are more likely to also be male and/or a household head and/or employ non-household labour for either farm or non-farm work.

Network Diversity: Composition

A. Age Heterogeneity

Turning first to the variable capturing diversity in age of network partners, the higher the score on the age heterogeneity variable, the greater is heterogeneity in age of network partners.

The networks of unmarried respondents are more diverse in terms of age of network partner than those of married respondents

Unmarried people are more than three times as likely as one would expect given their share in the sample as a whole to have a network age heterogeneity of greater than 15 (38.9 per cent of

highest 15+ category compared with 12.7 per cent unmarried in the sample as a whole; see Table B.8).

Table B.8 Network age heterogeneity and respondent i) Marital status;; ii) Education level (% of respondents)

Agehet	marital status		level of education		
	married	not married	To grade 4	Grade 5-7	Secondary +
<5	95.7	4.30	27.3	50.0	22.7
5-<10	91.8	8.20	8.60	54.3	37.0
10-<15	92.1	7.90	30.1	56.2	13.7
15+	61.1	38.9	50.0	33.3	16.7
Total Sample (%)	87.3	12.7	24.3	51.5	24.3

Source: Household Survey, 2002-03

The lower the level of education of the respondent the greater the age heterogeneity of their network partners

Conversely, the higher the respondent's level of education the lower the age heterogeneity of alters. Table 4.21 also sets out the distribution. Those with lower levels of education (up to grade 4) make up 50 per cent of those with network age heterogeneity of 15+ compared with just 24.3 per cent in sample as a whole. Those with secondary education and higher are over-represented in the lower age heterogeneity categories.

The older the respondent the more heterogeneous are their network partners in terms of age

This relationship warrants some discussion about the direction one would expect it to take as it is rather ambiguous: Are older people likely to have more or less age heterogeneity in their networks? A traditional concept of 'age-mates' means that one might expect an older respondent to have networks dominated by network partners who are similar in age, both to each other and to the respondent themselves. Contrary to this, are the networks of older people likely to be characterised by people of diverse ages for such reasons as i) older people are more likely to know more people by sheer dint of having been around for a lot longer compared to their younger compatriots, and therefore more chance of having people of many ages in their network; ii) Older people are more likely to be in a respected position with the result that people of all ages interact with them rather than 'social circles' being confined solely to one's own age group? iii) demographic factors, that is, there are fewer older people around, coupled with more reliance by older people on those who are younger than themselves i.e. the productive age group? Work on social support networks suggest that younger people tend to provide older people with their labour (Borgatti, 1998). The data show that there is in general more age heterogeneity in the networks of those respondents in the over 60 age group. Table B.9 shows that 38.9 per cent of over 60s have an agehet of 15+ compared with 9.5 per cent in sample as a whole, while younger respondents (age 30 or less) make up nearly half (47.8 per cent) of those with an age heterogeneity of networks partners of less than 5, compared with being only 28.6 per cent of the sample as a whole.

Table B.9 Network Age heterogeneity and i) Age of respondent; ii) Sex of household head (% of respondents)

Agehet	Age of respondent			Sex of household head	
	30 and under	31-60	>60	male	female
<5	47.8	43.5	8.70	95.7	4.3
5-<10	25.9	72.9	1.20	85.9	14.1
10-<15	28.9	65.8	5.30	92.1	7.9
15+	22.2	38.9	38.9	69.4	30.6
Total Sample (%)	28.6	61.8	9.50	86.4	13.6

Source: Household Survey, 2002-03

To shed light on this finding it may be helpful to break the network down by network type, and examine its component parts. In more traditional settings would expect age heterogeneity of alters to be low in interaction networks because of the concept of “age mates”, mentioned above.

People from a female-headed household have more age heterogeneity in their networks than those with a male household head

Respondents with a female head of household make up just 13.6 per cent of the sample. They are over-represented in the highest age heterogeneity category, comprising 30.6 per cent of those with age heterogeneity >15 (See Table B.9). The age heterogeneity of networks partners of those with a male household head is in general lower: of those with network age heterogeneity of less than five, 95.7 per cent have a male household head, compared to 86.4 per cent of male household heads in the sample as a whole.

B. Sex Heterogeneity

Women’s networks are more likely to be mixed-sex than those of men

Most men (57.6 per cent) fall in the middle of the distribution, but are over-represented in the lower sex heterogeneity category, making up 77.6 per cent of those with low sex heterogeneity in their networks (Table B.10). Women make up 79.3 per cent of those with sex heterogeneity of greater than 0.9 compared with 48.9 per cent of the sample as a whole.

Respondents who are not head of their household are also likely to have a greater variation in sex of network partner

One would expect this variable to follow closely the distribution of the variable for sex of respondent because females in the sample tend to be the spouse of a male household head rather than the head of a female-headed household. Female-headed households make up just 16 per cent of the sample. This is indeed the case: 58.6 per cent of respondents falling in the >0.9 category are not household heads compared with just 43.3 per cent non-household heads in the total sample (see Table B.10).

Table B.10 Network sex heterogeneity and i) sex of respondent; ii) status in household; iii) sex of household head (% of respondents)

Sexhet	Sex of respondent		Household status		Sex of household head	
	male	female	household head	other	male	female
<0.01	78.0	22.0	68.0	32.0	94.0	6.00
0.01-0.9	54.5	45.5	60.3	39.7	82.1	17.9
>0.9	20.7	79.3	41.4	58.6	79.3	20.7
Total Sample (%)	51.1	48.9	57.2	42.8	84.0	16.0

Source: Household Survey, 2002-03

Table B.10 also shows the relationship between diversity in sex of network partner and sex of household head. Networks of those with male household heads are likely to have lower sex heterogeneity compared with females, who are over-represented in the >0.9 category.

A very simple indicator of multiplexity is taken in which each network partner can have a tie to the respondent in each of the seven strands of instrumental network being measured (personal, migration, employer, employee, social support, information and market access). The indicator takes a value between zero and seven, according to the number of network strands for which the

respondent has a stated relationship with other people. A score of “zero” would denote no relationships.

Men are likely to have network partners in more of the network types than women

Almost 60 per cent of those involved in five, six or the maximum seven of the networks being measured are males compared with 51.1 per cent of respondents being male in the sample as a whole (Table 4.3.11).

Those who have not always lived in their village are more likely to be involved in more networks

Respondents are more likely to be in 5-7 networks if they have not always lived in their current community (see Table B.11). Those who have always lived in their community are over-represented in the 1-2 network category (85.7 per cent of those with only one or two network types fall within the ‘yes’ group, compared with 64.3 per cent of the sample as a whole).

Table B.11 Multiplexity of networks and i) sex of respondent; ii) residence status; iii) group labour employment (% of respondents)

Multego	Sex of respondent		Always lived here?		Employer of group labour	
	male	female	no	yes	no	yes
1/2	31.0	69.0	14.3	85.7	65.5	34.5
3/4	50.4	49.6	26.5	73.5	66.9	33.1
5/6/7	59.3	40.7	55.3	44.7	43.2	56.8
Total Sample (%)	51.1	48.9	35.7	64.3	58.4	41.6

Source: Household Survey, 2002-03

Employers of group labour are also more likely to be involved in many kinds of network

Those employing group labour are over-represented in the larger multiplexity category of five to seven of the instrumental networks being studied: 56.8 per cent of respondents involved in between five and seven networks employ group labour compared with 41.6 per cent of sample (Table B.11).

Table B.12 Network Size and Decision-making Power in groups

Network Size	To what extent do you participate in the group's decision-making?				
	Not at all	To a small extent	Neither small nor large extent	To a large extent	To a very large extent
5 or less	45.83	20.83	2.08	20.83	10.42
6-10	14.61	8.99	8.99	38.20	29.21
11-15	15.91	11.36	6.82	34.09	31.82
>15	3.33	0.00	10.00	40.00	46.67
Total Sample (%)	20.38	10.90	7.11	33.65	27.96

Appendix C Qualitative Evidence of Smallholder Market Participation

This appendix sets out findings from qualitative fieldwork in the three study communities, based on focus group discussions with formal village-level farming groups, interviews with smallholder farmers and their households, and participatory analysis using pair-wise-ranking to identify determinants of market participation.

During qualitative fieldwork (focus group discussions and semi-structured interviews with individuals and households) farm households, discussing the decision-making process behind labour hiring and crop marketing decisions, reported that they first put aside crops for household consumption needs. The remainder was divided between crops to keep as in-kind payment for farm labour, and crops to sell in the market. Household food security comes first, deciding what and how much to consume. If the household cannot meet its own food security needs then they have to buy what they need from the market or work in exchange for food. Part of the decision by some households to keep back crops as in-kind payments for farm labour rests on the premise that some of their neighbours will be in this position. Precarious food markets are not just a cross borne by rural residents. Many of the settlers in the Lufubu resettlement scheme reported the potential opportunity to achieve food security for the family as one of the main motivations behind settling on the scheme, relocating from small 'rural' towns, with one resident saying "Here at least we can feed ourselves" (Lufubu Resident, Fieldwork September 2002).

A pair-wise ranking exercise was carried out in focus group discussions in all three study sites during round 2 of the fieldwork, designed to uncover what villagers themselves considered as signifying "getting on", or achieving a successful livelihood. This encompassed what they considered "getting on" to be, the factors they believed to be necessary in achieving this, and what was the ordered preference of these factors in relation to each other. The exercise was conducted in a small group in each village of between ten and thirteen people (ten in both Ngulula and Kabila, thirteen in Lufubu), of mixed sex and ages, initially focusing on what did people consider 'getting on' to mean to them. Participants universally agreed that this meant 'having a better life' and saw the route to this as doing well in their farming. It is worth noting that the discussions that people engage in during the course of exercises such as this could be considered as more important than the outcome or 'results' themselves (Chambers, 2003).

To Kabila's participants, "getting on" means cultivating more in terms of increased crop output. In Ngulula, getting on means 'having a better life' and again the route to this was considered to be through farming. Of all three communities the Ngulula participants alone explicitly mentioned the longer term and what doing well meant to them in terms of the next generation. Lufubu participants took a broader view of "getting on", thinking about it in terms of how to access economic opportunities, for example jobs in town, paid work, growing better crops or increasing production or starting a new enterprise. However, the emphasis was very much on their farming enterprises.

In each group, participants, unprompted, listed all the elements they thought necessary for doing well in their farming, which could run into twenty or more factors. These included the obvious assets and inputs necessary for cultivation, and also encompassed personal relations: cooperation (between people both within and outside the household, working together); people: relatives, workers and others more generally; 'being famous' i.e. popular (thought to be necessary for attracting and keeping customers). Personal attributes also appeared in the long-list produced by Lufubu participants: honesty, sobriety and cleanliness, with a discussion about the importance of these for doing business with other people in general.

Participants discussed their list, grouping elements when there was repetition, and then choosing what they considered to be the ten most important. These ten elements or *determinants* were then taken two-by-two and participants were asked to state which was the most important element of the pairs in relation to having a successful agricultural livelihood, until all pair-wise combinations of the 10 elements had been covered – 45 combinations in all. Matrices produced in each of the study sites are given below. Cells contain the 'preferred' element in each pair. The elements were then ordered according to the number of times they were preferred over other elements, to give scores for the ten most important determinants to

participants. The list was discussed and participants were given the opportunity to rearrange the list if they wished.

Kabila

	Tools	Transport	Money	Food	Land	Bridge	Hard Work	Cooperation	Education	Health
Tools										
Transport	Tools									
Money	Money	Money								
Food	Tools	Food	Food							
Land	Tools	Land	Land	Land						
Bridge	Tools	Bridge	Money	Food	Land					
Hard work	Tools	Hard work	Hard work	Hard work	Hard work	Hard work				
Cooperation	Tools	Transport	Cooperation	Cooperation	Land	Cooperation	Hard work			
Education	Tools	Education	Money	Food	education	education	Hard work	Education		
Health	Health	Health	Health	Health	Health	Health	Health	Health	Health	

Ngulula

	Education/ skills	Health	Hard work	Money	Market	Transport	Educating Children	Assistance	Cooperation	Water
Education/ skills										
Health	Health									
Hard Work	Hard work	Health								
Money	Money	Health	Hard work							
Market	Education	Health	Hard work	Market						
Transport	Transport	Health	Hard work	Transport	Transport					
Educating Children	Education	Health	Hard work	Money	Market	Transport				
Assistance	Education	Health	Hard work	Assistance	Assistance	Assistance	Educating children			
Cooperation	Cooperation	Health	Cooperation	Cooperation	Cooperation	Cooperation	Cooperation	Cooperation		
Water	Water	Water	Water	Water	Water	Water	Water	Water	Cooperation	

Lufubu

	Education	Capital	Land	Tools	Health	Food	Water	Preparation	Honesty	People
Education										
Capital	Capital									
Land	Land	Capital								
Tools	Tools	Tools	Land							
Health	Health	Health	Health	Health						
Food	Food	Food	Food	Tools	Food					
Water	Water	Water	Water	Water	Water	Water				
Preparation	Prep	Prep	Prep	Prep	Prep	Prep	Water			
Honesty	Education	Capital	Land	Tools	Health	Food	Water	Prep		
People	People	People	People	People	Health	Food	Water	Prep	People	

Scores, rankings and re-ranked lists are given in Table C.1. Scores signify the number of times a particular element was chosen as preferred over another. Taking as an example the score of 7 for tools in Kabila, this means that on seven occasions 'tools' was chosen as the more important factor in determining how well someone does in farming when drawn in a pair with other factors. Initial ranking is based on these scores, and the final ranking is the result of participants re-ranking following discussion of the rankings based on scores.

Table C.1 Scores from matrix ranking of determinants of 'economic success'

Kabila	Ngulula	Lufubu
SCORES:	SCORES:	SCORES:
Health 9	Health 8	Preparation 8
Tools 7	Cooperation 8	Health 6
Hard work 7	Water 8	Water 6
Land 5	Hard Work 6	People 5
Money 4	Transport 4	Tools 4
Food 4	Education/ skills 3	Food 4
Education 4	Assistance 3	Money 3
Cooperation 3	Money 2	Land 3
Transport 1	Market 2	Education 1
Bridge 1	Educating children 1	Honesty 0
INITIAL RANKING:	INITIAL RANKING:	INITIAL RANKING:
1 Health	1= Health, water, cooperation	1 Preparation
2= Tools, hard work	4 Hard work	2= Water/ Health
4 Land	5 Transport	4 People
5= money, food education	6= Education, Assistance	5= food/ tools
8 cooperation	8= Money, Market	7= land/ capital
9= transport, bridge	10 Educating children	9 education
		10 honesty
FINAL RANKING:	FINAL RANKING:	FINAL RANKING:
1 Health	1 Health	1 Preparation
2 Tools	2 Water	2 Water
3 Hard work	3 Cooperation	3 Health
4 Land	4 Hard work	4 Tools
5 Food	5 Education	5 Food
6 Money	6 Transport	6 People
7 Education	7 Market	7 Land
8 Cooperation	8 Money,	8 Capital
9 Bridge	9 Educating children	9 Education
10 Transport	10 Assistance	10 Honesty

Kabila and Ngulula participants each included cooperation in their 'top ten', while Lufubu residents grouped all of the elements of their master list related to personal relations under the heading 'People'. In the two study sites with poorer market access – Kabila and Lufubu – having enough food and meeting basic needs scored highly, in the top five in preference ordering. This did not feature in the Ngulula list. Only Ngulula participants included available markets in their top ten. Cooperation came joint first in Ngulula, but much lower (8th) in Kabila, while people were ranked 4th most important for economic success by Lufubu participants. This dropped to 6th place when participants later re-ranked the list, putting basic needs (food) and productive assets (tools) ahead of people.

Determinants identified are a mixture of factors necessary to meet immediate, basic material needs, in terms of consumption and production, such as transport, cash, food, productive assets and inputs (land, tools, water/ irrigation), education – in agreement with the theoretical model. However, in contrast to the theoretical model, which focuses only on household characteristics as well as transactions costs related to marketing, people also stressed the importance of more 'intermediate' factors, focusing on personal and relational attributes such as honesty, industriousness and cooperation. However, these personal and relational attributes are seen to be necessary for fostering and maintaining good relations with others, which in turn leads to improved access to assets, labour, information – in other words lowering transactions costs in market access.

Potential biases notwithstanding, the pair-wise ranking exercise and accompanying discussion are potentially useful tools for exploring empirically and in some depth people's own assessment of what they consider to be the determinants of market participation and directions of causality in their particular context.

Appendix D Construction of Socio-Economic Status Variable

The socio-economic status variable is an index constructed using underlying indices based on the survey data⁵². The notes below refer to question numbers in the household survey questionnaire.

Definition of Constructed Variables

Agricultural Assets Index:

Using constructed variables for Agricultural Implements Index AGIMPI and Livestock Unit Index LUI, AGASSETSI is a simple average of the two. The resulting variable is in the range 0 to 1.

Agricultural implements index:

Using answers to question 3.3 [How many [ITEM]s do you own?]:

$$AGIMP_OWN = 0.5 \times 8 + 9 + 10 + 11 + 2(12 + 13) + 14$$

Using answers to question 3.4 [How many [ITEM]s do you own with another household or households?]

$$AGIMP_POWN = 0.5 \times 8 + 9 + 10 + 11 + 2(12 + 13) + 14$$

$$AGIMP = AGIMP_OWN + 0.5 AGIMP_POWN$$

Agricultural implements index is scaled: $AGIMPI = AGIMP/AGIMP(max)$, where $AGIMP(max)$ is the maximum value of AGIMP found in the data. The resulting variable is in the range 0 to 1.

Dependency ratio: the number of people in a household aged 15 or less plus those aged over 60 divided by the number aged 15-60.

Household size: the total number of people currently resident in the household.

Housing quality

Housing quality, HQ is then the sum of answers to questions 3.7 [What is the main type of flooring in your place?], 3.8 [What is the main type of roofing material in your place?] and 3.9 [What is the main type of wall material in your place?] plus ten times the number of rooms per person (ROOMSP). The index, $HQI = HQ/HQ(max)$, where $HQ(max)$ is the maximum value of HQ found in the data. The resulting variable is in the range 0 to 1.

Livelihood diversification: constructed as the sum of the following scores (at the household level) divided by 50:

2 if $LU_L > 0$ (large livestock unit)

1 if $LU_S > 0$ (small livestock unit)

2 for each current migrant

2 for each agricultural employee

1 for each type of non-farm work engaged in

0.5 for each type of good produced for outside consumption

0.5 for each crop produced

1 for owning fish pond

1 for collecting forest products

⁵² See: White H, J Leavy, M Mulumbi, G Mulenga, Venkatesh Seshamani (2005) *Rural Labour Markets in Africa: a channel for pro-poor growth? Preliminary findings from a research project in Northern Province, Zambia*. Discussion Draft. IDS Mimeo.

Livestock units:

Using answers to question 3.3 [How many [ITEM]s do you own?]:

$$LU_OWN = 24 + 0.8 (25 + 26) + 0.5 (27 + 28 + 29 + 30) + 0.2 \times 31 + 0.05 (32 + 33 + 34 + 35)$$

Using answers to question 3.4 [How many [ITEM]s do you own with another household or households?]

$$LU_POWN = 24 + 0.8 (25 + 26) + 0.5 (27 + 28 + 29 + 30) + 0.2 \times 31 + 0.05 (32 + 33 + 34 + 35)$$

$$LU = LU_OWN + 0.5 LU_POWN$$

Livestock unit index is scaled: $LUI = LU/LU(max)$, where $LU(max)$ is the maximum value of LU found in the data. The resulting variable is in the range 0 to 1.

Ownership of consumer durables (CD): the sum of the answers given to question 3.1 [Does your household own any of the following (9) items?] divided by 9. The resulting variable is in the range 0 to 1.

Rooms per person (ROOMSPP): the number of rooms in the house (question 3.11 How many rooms do you have in your place?) divided by HHS.

Transport index

Using answers to question 3.3 [How many [ITEM]s do you own?]:

$$TRAN_OWN = 16 + 2 \times 17$$

Using answers to question 3.4 [How many [ITEM]s do you own with another household or households?]

$$TRAN_POWN = 16 + 2 \times 17$$

$$TRAN = TRAN_OWN + 0.5 TRAN_POWN$$

Transport index is scaled: $TRANI = TRAN/TRAN(max)$, where $TRAN(max)$ is the maximum value of $TRAN$ found in the data. The resulting variable is in the range 0 to 1.

Water source: For question 3.10 [What is the main source of drinking water for your household for most of the year] “other” is re-coded as 8 “furrow”. $WATER$ is then the answer to question 3.10 divided by 7. The resulting variable is in the range 0 to 1.

Socio-Economic Status

Socio-economic status (SES): the simple average of CD , $AGRIMPI$, LUI , HQI , $TRANI$ and $WATER$. The resulting variable is in the range 0 to 1.

3. I would now like to ask you about various items which your household may own:

Does your household own any of the following items?		3.1a 0 No 1 Yes	3.1b How Many? Number
ASSET CODES			
1	Chair(s)		
2	Table		
3	Bed(s)		
4	Mattress(es)		
5	Stove/ Cooker/ Charcoal blazer		
6	Working Radio		
7	Working Watch or clock		

Now I'd like to ask you about things your household may own, possibly jointly with other households. FIRST ASK 3.2 FOR ALL ITEMS

3.2 Does your household own any of the following items? FOR EACH ITEM THE HOUSEHOLD OWNS GO ON TO ASK QUESTIONS 3.3-3.6 ONCE ALL ITEMS ARE LISTED		3.3 How many [ITEM]s do you own? Number	3.4 How many [ITEM]s do you own with another household or households? Number	3.5 Have you rented out any of these [ITEM]s in the last growing season? No Yes 99 N/A	3.6 How many [ITEM]s have you sold in the last growing season? Number
ASSET CODES					
	No Yes, own Yes, have use of				
8	Farm Tools e.g. Hoe, Scythes, Knives, axes, rakes, sickles, shovels				
9	Working Crop Sprayer				
10	Working Hammer Mill				
11	Cart (small – person pulled)				
12	Cart (large – ox pulled)				
13	Plough				
14	Fishing net				
15	Storage shed/Containers				
16	Working Bicycle				
17	Working Motorcycle/scooter/vehicle				

ANIMAL CODES:

24	Oxen					
25	Donkey					
26	Bull or heifer					
27	Calf					
28	Sheep					
29	Goat					
30	Pigs					
31	Piglets					
32	Duck and Geese					
33	Rabbits					
34	Chickens					
35	Other birds					
36	Other animals					

IF POSSIBLE RECORD ANSWERS TO QUESTIONS 3.7 – 3.9 BY OBSERVATION. IF NOT, ASK RESPONDENT.

<p>3.7 What is the main type of flooring in your place? [CIRCLE ONE]</p> <p>1 Dirt/earth</p> <p>2 Wood/plank</p> <p>3 Cement/tile</p> <p>4 Other</p> <p> (specify)</p>	<p>3.8 What is the main type of roofing material in your place? [CIRCLE ONE]</p> <p>1 Grass/Straw/thatch/</p> <p>2 Wood/planks</p> <p>3 Tin/metal</p> <p>4 Tiles/slates/concrete/cement</p> <p>5 Other</p> <p> (specify)</p>	<p>3.9 What is the main type of wall material in your place? [CIRCLE ONE]</p> <p>1 Brick/concrete</p> <p>2 Galvanised iron</p> <p>3 Adobe/mud</p> <p>4 Matting/wood/branches</p> <p>5 Other</p> <p> (specify)</p>
---	---	--

<p>3.10 What is the main source of drinking water for your household for most of the year [CIRCLE ONE]</p> <p>1 River, Lake</p> <p>2 Unprotected well</p> <p>3 Protected well</p> <p>4 Borehole</p> <p>5 Public tap or neighbour's tap</p> <p>6 Own tap</p> <p>7 Other</p> <p> (specify)</p>

3.11 How many rooms do you have in your place?	NUMBER	
--	--------	--

Appendix E Descriptive Statistics: 1. Social Networks by Location; 2 Household Characteristics by Commercialisation Status

E.1 Descriptive statistics of household social network characteristics by community

The purposive nature of study site selection means location is a proxy for market access.
N=136

Table E1.1 Social network characteristics of households by location

Respondent attributes:	Network size	Group affiliation	Influential kin*
Community			
	9.8	1.6	51.4
Kabila	10	2.3	77.5
Ngulula	13.2	3.1	78.0
Lufubu			

Source: Household Survey, 2002-03; all data reported are means, with the exception of the influential kin dummy variable, which is reported as the percentage of respondents reporting an influential friend or kin in their network.

Table E1.2 Mean Network Diversity by location

Outcomes:	Age heterogeneity	Sex heterogeneity	Ratio of kin to non-kin	Number of multiplex ties	Proportion of multiplex ties	Sum of accessed resources	Average no. of roles per alter
Community							
Kabila	13.9	0.24	0.61	2.1	0.21	11.9	3.4
Ngulula	13.2	0.23	0.37	1.7	0.20	12.8	3.7
Lufubu	10.0	0.23	0.29	2.1	0.17	14.4	3.4

Source: Household Survey, 2002-03. Note: The higher the score the more diverse the network on the dimension in each column.

E2 Descriptive statistics of characteristics of seller and non-seller households

Looking across the sample as a whole, crop market participants are overrepresented by male household heads, and households who have higher levels of education, despite the common perception among farmers interviewed during qualitative fieldwork that “you don’t need education to be a farmer” (Table E2.1).

Table E2.1 Percentage of households marketing crops by sex of household head and highest education level in the household

Participation in crop marketing	Sex of HH head (%)		Highest level of education in the household (%)		
	male	female	up to grade 4	grade 5-7	secondary/ college
No	59.1	40.9	13.6	63.6	22.7
Yes	86.1	13.9	7.0	38.3	54.8
Total	81.8	18.2	8.0	42.3	49.6
	Chi-sq=9.77		Chi-sq=7.46		

In terms of productive assets, crop market participants also tend to have irrigation (although this is not significant - only 15.3 per cent of households surveyed have irrigation and most of these

are in Ngulula), use hybrid seed and fertilizer (Table E2.2), have storage, large tools and own or have use of a bicycle (Table E2.3).

Table E2.2 Crop market participation by farm inputs: irrigation, fertilizer, hybrid seed

Participation in crop marketing	Irrigation (%)		Fertiliser (%)		Hybrid seed (%)	
	no	yes	no	yes	no	yes
No	95.5	4.5	77.3	22.7	90.9	9.1
Yes	82.6	17.4	50.4	49.6	60.9	39.1
Total	84.7	15.3	54.7	45.3	65.7	34.3
	Chi-sq=2.18		Chi-sq=5.19		Chi-sq=7.17	

Table E2.3 Crop market participation by farm equipment and transport: storage, large tools, bicycle

Participation in crop marketing	Storage (%)		Large Tools (%)		Bicycle (%)	
	no	yes	no	yes	no	yes
No	90.9	9.1	100.0	.0	50.0	50.0
Yes	64.3	35.7	75.7	24.3	25.2	74.8
Total	68.6	31.4	79.6	20.4	29.2	70.8
	Chi-sq=5.84		Chi-sq=6.50		Chi-sq=5.36	

Participating households are only slightly more likely to have non-farm income (not significant). Farmers who sell their crops are also more likely to have a working radio, an important source of market and farming information for many farmers living in remote rural areas such as these (Table E2.4). This is potentially a proxy for transactions costs.

Table E2.4 Crop market participation by access to non-farm income and information

Participation in crop marketing	Non-farm income (%)		Working radio (%)	
	no	yes	no	yes
No	54.5	45.5	86.4	13.6
Yes	47.8	52.2	49.6	50.4
Total	48.9	51.1	55.5	44.5
	Chi-sq=0.29		Chi-sq=10.3	

There is no significant difference between sellers and non-sellers in terms of their propensity to have worked for someone else in a non-farm activity, although not surprisingly sellers are more likely to employ non-household farm labour, as well as non-farm employees (Table E2.5).

Table E2.5 Crop Market Participation by Employment (as an individual)

Participation in crop marketing	Employs Farm labour (%)		Worked as farm labourer (%)		Employed someone else in non-farm activity (%)		Worked for someone else in non-farm activity (%)	
	no	yes	no	yes	no	yes	no	yes
No	72.7	27.3	63.6	36.4	95.5	4.5	86.4	13.6
Yes	31.3	68.7	62.6	37.4	72.2	27.8	80.9	19.1
Total	38.0	62.0	62.8	37.2	75.9	24.1	81.8	18.2
	Chi-sq=13.2		Chi-sq=0.01		Chi-sq=5.55		Chi-sq=0.39	

Crop market participants are likely to have both worked in a group and to have employed group labour for farm work, reflecting the need for larger groups of workers to carry out heavy work for planting, and the importance of being able to mobilise such groups in a timely fashion and reciprocate (Table E2.6). In order to employ group labour, households need adequate resources to feed the group and also to provide reciprocal labour suggesting that a certain command over resources – material and physical – is necessary from the outset and points to potential endogeneity between labour variables and welfare outcomes.

Table E2.6 Crop Market Participation by Employment (Group level)

Participation in crop marketing	Work in a group (%)		Employed group labour (%)	
	no	yes	no	yes
No	81.8	18.2	59.1	40.9
Yes	56.5	43.5	42.6	57.4
Total	60.6	39.4	45.3	54.7
	<i>Chi-sq=5.08</i>		<i>Chi-sq=1.93</i>	

The mean output of sellers compared with non-sellers, in terms of value of crops grown, is considerably higher (Table E2.7). Sellers on average grow crops to a value of 2 million kwacha, with a maximum of 10.9 million kwacha, compared to an average crop value of 174 thousand kwacha for non-sellers, with a maximum of 1.5 million kwacha.

Table E2.7 Mean output of sellers and non-sellers, '000 Kwacha

Participation in crop marketing	Value of crops grown ('000 Kwacha)			
	Mean	St.dev	min	max
No	174.87	344.31	0	1516.11
Yes	2000.81	2287.867	84.14	10929.34

The correlation between total value of crops sold and total area of land cropped is weakly positive at 0.3705, as is the correlation between total value of crops sold and total land area owned (0.3464). The customary land tenure system and the relative abundance of land in Zambia, as discussed previously, mean that land availability per se is not a constraint on production. Rather, farmers are constrained by lack of access to other inputs such as fertiliser and labour. In Northern Province land quality is also a constraint.

Table E2.8 Shows the percentage of households selling each type of crop by community over the two rounds of the survey, with average prices per kg weight for the six month period. This is an average of the maximum and minimum price reported at the community level. In round 1, with a recall period encompassing the immediate post-harvest period, prices were identical across the three communities. In round 2 prices in Kabila and Ngulula were the same in both locations (related to prices in the provincial capital Kasama) but prices in Lufubu differed for some crops, suggesting for some crops Lufubu farmers are facing different markets.

Table E2.8 Percentage of households selling each type of crop, by community

Crop	Round 1				Round 2				
	Kabila	Ngulula	Lufubu	Average* Price	Kabila	Ngulula	Lufubu	Average* Price: Kabila, Ngulua	Average* Price: Lufubu
Maize	2.4%	18.6%	19.7%	731	0.0%	14.0%	1.6%	900	528
Hybrid Maize	0.0%	14.0%	4.9%	731	0.0%	0.0%	0.0%	900	528
Cassava	4.9%	16.3%	11.5%	429	2.4%	14.0%	4.9%	327	365
Millet	17.1%	18.6%	32.8%	542	0.0%	0.0%	0.0%	484	397
Sorghum	4.9%	4.7%	0.0%	756	0.0%	0.0%	0.0%	1203	702
Mixed beans	56.1%	16.3%	77.0%	1149	7.3%	2.3%	27.9%	300	1500
Soybean	0.0%	23.3%	4.9%	600	0.0%	0.0%	0.0%	745	461
Sweet potato	0.0%	53.5%	9.8%	342	0.0%	20.9%	0.0%	619	1003
Irish potato	0.0%	9.3%	3.3%	328	0.0%	4.7%	1.6%	743	833
Groundnut	39.0%	16.3%	42.6%	1529	0.0%	0.0%	1.6%	1667	875
Cabbage	0.0%	32.6%	3.3%	619	0.0%	20.9%	0.0%	1000	625
Mango	0.0%	0.0%	0.0%	2000	0.0%	2.3%	1.6%	1000	320
Banana	0.0%	23.3%	4.9%	1500	0.0%	11.6%	0.0%	4000	4000
Plantain	0.0%	2.3%	0.0%	1500	0.0%	9.3%	1.6%	2000	2000
Squash	0.0%	0.0%	3.3%	450	0.0%	4.7%	0.0%	400	400
Orange	0.0%	0.0%	0.0%	750	0.0%	0.0%	0.0%	722	722
Tomato	0.0%	18.6%	13.1%	356	2.4%	9.3%	1.6%	732	732

*Average of minimum and maximum price reported at community-level.

E2.9 Average crop sales '000 Kwacha

Crop	Round 1				Round 2				
	Kabila	Ngulula	Lufubu	Average* Price	Kabila	Ngulula	Lufubu	Average* Price: Kabila, Ngulua	Average* Price: Lufubu
Maize	11983	52367	65128	731	0	196074	43	900	528
Hybrid Maize	0	55462	46742	731	0	0	0	900	528
Cassava	2812	5928	10763	429	240	5459	2291	327	365
Millet	9099	5834	45532	542	0	0	0	484	397
Sorghum	1291	6629	0	756	0	0	0	1203	702
Mixed beans	70125	26397	210194	1149	3462	140	14472	300	1500
Soybean	0	20386	1987	600	0	0	0	745	461
Sweet potato	0	84423	13431	342	0	14316	0	619	1003
Irish potato	0	5866	677	328	0	3464	1270	743	833
Groundnut	53492	192101	134789	1529	0	0	493	1667	875
Cabbage	0	98688	2233	619	0	241601	0	1000	625
Mango	0	0	0	2000	0	4651	1475	1000	320
Banana	0	427326	82377	1500	0	147093	0	4000	4000
Plantain	0	11302	0	1500	0	37209	1230	2000	2000
Squash	0	0	664	450	0	3349	0	400	400
Orange	0	0	0	750	0	0	0	722	722
Tomato	0	61106	8335	356	264	11085	2131	732	732

*Average of minimum and maximum price reported at community-level.

Appendix F Multinomial Probit Coefficient Estimates

Multinomial probit results ^a						
Independent Variables	Without social networks			With social networks		
	HL	LH	HH	HL	LH	HH
Access to storage facilities	0.29 (0.53)	1.75*** (0.53)	0.97* (0.47)	1.11 (0.63)	2.31*** (0.69)	1.82** (0.66)
Age of Household head	-0.03 (0.02)	-0.03* (0.02)	-0.05** (0.02)	-0.02 (0.02)	-0.03 (0.02)	-0.05* (0.02)
Non-farm income	0.00* (0.00)	0.00** (0.00)	0.00** (0.00)	0.00* (0.00)	0.00** (0.00)	0.00** (0.00)
Proportion of land cropped	-0.80 (0.70)	-1.89** (0.67)	-0.98 (0.67)	-0.66 (0.73)	-1.08 (0.68)	-0.81 (0.74)
Access to working radio	0.78 (0.52)	1.97*** (0.52)	2.23*** (0.53)	0.67 (0.52)	1.94*** (0.59)	2.36*** (0.58)
Ngulula	1.57** (0.50)	1.76** (0.63)	3.46*** (0.57)	1.89*** (0.56)	2.32*** (0.65)	4.03*** (0.65)
Post-primary education	-0.11 (0.42)	0.33 (0.52)	0.84 (0.43)	-0.00 (0.41)	0.29 (0.56)	0.88 (0.50)
Network size (log)				0.54 (0.61)	1.28 (0.78)	0.60 (0.74)
Sex heterogeneity of network				-4.89* (2.00)	-6.27*** (1.88)	-2.73 (1.86)
Influential kin nearby				0.94* (0.48)	1.97** (0.75)	1.86** (0.64)
Influential kin far location				2.19*** (0.59)	2.00* (0.89)	2.79*** (0.79)
Constant	-0.28 (1.09)	-1.71* (0.71)	-1.62 (0.97)	-2.07 (1.44)	-5.62** (1.95)	-4.72** (1.74)
Model chi-square	72.03			102.40		

Observations = 135

Robust Standard Errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

^a Cluster base category is LL

Appendix G Theoretical Model of Agricultural Household with Transactions Costs

The underlying theoretical model for the empirical analysis is a basic static model of the agricultural household incorporating transaction costs, following the methodology of Goetz (1992) and later extended (see Key, Sadoulet and de Janvry, 2000; Heltberg and Tarp, 2002).

Key, Sadoulet and de Janvry (2000) modify and extend the basic static agricultural household model to incorporate transactions costs⁵³. In this model, transactions costs are defined as being either *fixed* (regardless of the amount exchanged) or *variable* (vary with the amount exchanged).

In the absence of transactions costs, households ($i=1\dots N$) maximise utility, deciding how much to produce, consume and sell in the market:

$$U(c; z_u) \quad (G.1)$$

Subject to:

a budget constraint, which states that expenditure on all purchases must not exceed sales revenues plus transfers (t):

$$\sum_{i=1}^N p_i^m m_i + T = 0; \quad (G.2)$$

the resource balance (the amount consumed, used as an input and sold of a good is equal to production plus any endowment of that good):

$$q_i - x_i + A_i - m_i - c_i = 0; i = 1, \dots, N; \quad (G.3)$$

and production technology (the relationship of inputs to output):

$$G(q, x; z_q) = 0 \quad (G.4)$$

$$c_i, q_i, x_i \geq 0 \quad (G.5)$$

Where, for good i :

⁵³ The theoretical model developed by Key, Sadoulet and de Janvry (2000) ignores the role of risk and credit constraints. Neither does it consider the case of households that both sell and buy a particular good.

c_i is value of consumption;
 q_i is value of production;
 x_i is value of inputs;
 m_i is marketing (sales); $m_i > 0$ if there is a sale of good i and $m_i < 0$ if good i is purchased;
 p_i^m is market price;
 A_i is endowment of good i ;
 z_u and z_q are 'exogenous shifters' in utility and production respectively;
 G is production technology;
 T is exogenous transfers and incomes (nonfarm).

z_u and z_q are vectors of household specific characteristics related to consumption (c) and production (q) including: labour, other inputs, assets, information, capital, household characteristics/ demographics, location, external factors including government support, prices, alternative economic opportunities. They can also include 'non-quantifiable factors' such as entrepreneurial spirit, social networks, transactions costs.

Marketing is thus linked with production depending on consumption, as well as command over labour and other inputs and access to information. For the household, production minus consumption gives marketed surplus. Thus, in the absence of transactions costs, market participation takes place in response to observable price signals – it is endogenously determined by prices and exogenously by household characteristics.

However, market participation is not cost free, and the decision price (shadow price at which a smallholder farmer decides whether or not to be a net buyer, seller or autarkic) faced by the smallholder household may be different from observed (market) prices because of transactions costs. Transactions costs, for example unobservable transportation and marketing costs, increase the price paid by the buyer and reduce the price received by producers/ sellers, effectively widening the price band between buyers and sellers. Thus, the household problem can be expressed under fixed transaction costs (t_{fi}) and variable transaction costs (t_{vi}).

Expressing variable transaction costs (VTCs) in monetary terms:

$$\sum_{i=1}^N \left[(p_i^m - t_{vi}^s(z_t^s)) \delta_i^s + (p_i^m - t_{vi}^b(z_t^b)) \delta_i^b \right] m + T = 0 \quad (G.6)$$

The price effectively received by sellers is:

$$p_i^m - t_{vi}^s \quad (G.7)$$

The price effectively paid by buyers is;

$$p_i^m - t_{vi}^b$$

(G.8)

Where the superscripts b and s denote buyers and sellers respectively; t_{vi} are variable transactions costs associated with selling good i in the market and are expressed as a function of observable exogenous household characteristics z_t^s and z_t^b , affecting these costs when selling and buying respectively.

Rewriting the budget constraint to incorporate both fixed and variable transactions costs:

$$\sum_{i=1}^N \left[(p_i^m - t_{vi}^s(z_t^s)) \delta_i^s + (p_i^m - t_{vi}^b(z_t^b)) \delta_i^b \right] m_i - t_{fi}^s(z_t^s) \delta_i^s - t_{fi}^b(z_t^b) \delta_i^b + T = 0 \quad (G.9)$$

Where the household pays fixed transaction costs t_{fi}^s only if it sells good i and pays t_{fi}^b only if it buys it.

From here it is possible to derive supply demand and supply equations conditional on market participation of households facing FTCs and VTCs using the Langrangian:

$$L = u(c; z_u) + \sum_{i=1}^N \mu_i (q_i - x_i + A_i - m_i - c_i) + \phi G(q, x; z_q) + \lambda \left[\sum_{i=1}^N [(p_i^m - t_{vi}^s) \delta_i^s + (p_i^m - t_{vi}^b) \delta_i^b] m_i - t_{fi}^s \delta_i^s - t_{fi}^b \delta_i^b + T \right] \quad (G.10)$$

The Lagrangian multipliers associated with the resource balance, the technology constraint and the cash constraint are denoted by μ , ϕ and λ respectively.

Because the FTCs create discontinuities in the Lagrangian, the optimal solution follows two steps: First, solve for the optimal solution based on the form of market participation (sellers; buyers; autarkic), and then choose the form of market participation regime leading to the highest level of utility. Conditional optimal supply and demand are obtained by solving for FOCs, under usual utility and technology assumptions:

FOCs for consumption goods:

$$\frac{\partial u}{\partial c_i} - \mu_i = 0, i \in \{i | c_i > 0\} \quad (G.11)$$

For outputs:

$$\mu_i + \phi \frac{\partial G}{\partial q_i} = 0, i \in \{i | q_i > 0\} \quad (G.12)$$

For inputs:

$$-\mu_i + \phi \frac{\partial G}{\partial x_i} = 0, i \in \{i | x_i > 0\} \quad (G.13)$$

For traded goods:

$$-\mu_i + \lambda[(p_i^m - t_{vi}^s)\delta_i^s + (p_i^m - t_{vi}^b)\delta_i^b] = 0, i \in \{i | m_i \neq 0\} \quad (G.14)$$

Following de Janvry, Fafchamps and Sadoulet (1991), the decision price can be defined as:

$$p_i = \begin{cases} p_i^m - t_{vi}^s \text{ if } m_i > 0, & \text{seller} \\ p_i^m + t_{vi}^s \text{ if } m_i < 0, & \text{buyer} \\ \tilde{p}_i = \frac{\mu_i}{\lambda} \text{ if } m_i = 0, & \text{autarkic} \end{cases}$$

Profit maximisation subject to the technological constraint given by (4) gives a system of output supply equations $q(p, z_q)$ and input demand equations $x(p, z_q)$.

Utility maximisation subject to the income constraint is:

$$\sum_{i=1}^N p_i c_i = y = \sum_{i=1}^N [p_i(q_i - x_i + A_i) - t_{fi}^s \delta_i^s - t_{fi}^b \delta_i^b] + T \quad (G.15)$$

The selling and buying production thresholds \underline{q}^s and \underline{q}^b can be defined “as the production level when the household enters the market as a seller or buyer respectively; i.e. when decision prices are at their thresholds \underline{p}^s and \underline{p}^b ”:

$$\underline{q}^s \equiv q(\underline{p}^s, z_q) \quad (G.16)$$

$$\underline{q}^b \equiv q(\underline{p}^b, z_q) \quad (G.17)$$

The following inequalities give participation (supply) decisions in relation to price bands for seller, autarkic and buyer households. Price bands are larger the greater are transactions costs (including: transport costs; search, recruitment and supervision costs and other opportunity

costs of selling and buying; mark-ups by traders; costs associated with risk and lack of competitions – see de Janvry et al, 1991)⁵⁴.

$$\text{If } q(p^m - t_v^s, z_q) > \underline{q}^s(t_f^s, z_q, z_u, T, A),$$

That is, the shadow price is below the price band, then the household sells, and supply is given by:

$$q^s = q(p^m - t_v^s, z_q)$$

However, if:

$$\begin{aligned} q(p^m - t_v^s, z_q) &\leq \underline{q}^s(t_f^s, z_q, z_u, T, A), \text{ and} \\ q(p^m + t_v^b, z_q) &\geq \underline{q}^b(t_f^b, z_q, z_u, T, A), \end{aligned} \tag{G.18}$$

The shadow price falls within the price band, then the household is autarkic, and supply is given by:

$$q^a = (\tilde{p}, z_q), \text{ with } \tilde{p} \text{ defined by } c + x - A = q$$

If :

$$q(p^m + t_v^b, z_q) < \underline{q}^b(t_f^b, z_q, z_u, T, A),$$

The shadow price is above the price band and the household purchases, and supply is given by:

$$q^b = q(p^m + t_v^b, z_q)$$

The theoretical model thus implies that transactions costs, as well as household characteristics, affect decisions on buying, selling and autarky (self-sufficiency), and helps to distinguish between the roles played by fixed and variable transactions costs in household marketing decisions:

- The supply equation is a function of fixed transaction costs when the household makes the decision whether or not to participate in the market. Fixed transactions costs thus shift the threshold at which market participation can take place. When the household effectively participates, the amount sold depends on both fixed and variable transactions costs.
- For sellers, high transactions costs, especially fixed transactions costs, may push down their decision price to the extent it may not be worthwhile participating in the market. Further, because transactions costs effectively raise purchase prices, and increase the

⁵⁴ The subscript on the transaction cost variable t in the \underline{q}^b function should be f and not s as given in Key et al's original paper. The notation presented here corrects for their typo.

costs of selling, more own-production is likely to be consumed under transactions costs than without transactions costs.

- Social networks enter into the model as household characteristics (z_q and z_c) that affect a household's ability to deal with the constraints of missing markets and transactions costs. While they enter the model as exogenous variables, in reality social networks are likely to be endogenous and this will be explored later.

Appendix H GMM Results: Rotating the Instrumental Variables

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	logcsold	logcsold	logcsold	logcsold	logcsold	logcsold
propland	-1.65 (1.132)	-1.55 (1.018)	-1.67 (1.155)	-1.78 (1.286)	-1.77 (1.205)	-1.66 (1.142)
logsize	-1.53* (0.879)	-1.45* (0.757)	-1.51 (0.970)	-1.71 (1.051)	-1.70* (0.887)	-1.57** (0.755)
logsizeed2	1.57* (0.933)	1.69** (0.831)	1.53 (1.141)	1.70 (1.089)	1.61 (1.034)	1.59* (0.927)
logsizeerad	0.35 (0.769)	0.41 (0.656)	0.37 (0.834)	0.20 (0.894)	0.36 (0.844)	0.34 (0.777)
affil	0.72 (0.639)	0.52 (0.600)	0.70 (0.697)	0.99 (0.809)	0.86 (0.634)	0.75 (0.523)
depend	-0.46 (0.706)	-0.52 (0.632)	-0.47 (0.715)	-0.18 (0.878)	-0.39 (0.755)	-0.44 (0.685)
loghh_age	-0.12 (0.447)	-0.24 (0.439)	-0.12 (0.445)	-0.24 (0.523)	-0.12 (0.485)	-0.12 (0.451)
propkin	1.18** (0.510)	1.18** (0.478)	1.18** (0.516)	1.21** (0.567)	1.19** (0.543)	1.18** (0.513)
_lcomid_2	1.16* (0.610)	1.32** (0.574)	1.18* (0.668)	1.04 (0.693)	1.06* (0.630)	1.14** (0.562)
_lcomid_3	-0.79 (0.926)	-0.53 (0.885)	-0.77 (0.972)	-1.20 (1.172)	-0.99 (0.888)	-0.83 (0.781)
kininf	0.48 (0.336)	0.50* (0.299)	0.48 (0.337)	0.38 (0.401)	0.44 (0.356)	0.47 (0.336)
education	-3.30* (1.995)	-3.55** (1.781)	-3.23 (2.393)	-3.60 (2.336)	-3.42 (2.204)	-3.34* (1.975)
store	0.30 (0.473)	0.40 (0.444)	0.31 (0.481)	0.16 (0.565)	0.28 (0.512)	0.29 (0.463)
radio316	0.15 (1.742)	-0.00 (1.490)	0.10 (1.886)	0.49 (2.023)	0.12 (1.911)	0.16 (1.762)
nfy2	0.00 (0.001)	0.00 (0.001)	0.00 (0.002)	0.00 (0.002)	0.00 (0.002)	0.00 (0.002)
propmult	-0.90 (1.130)	-0.72 (1.078)	-0.90 (1.125)	-1.12 (1.292)	-0.90 (1.212)	-0.92 (1.109)
agehet	-0.02 (0.028)	-0.02 (0.025)	-0.02 (0.029)	-0.03 (0.034)	-0.02 (0.031)	-0.02 (0.028)
Constant	7.60*** (2.002)	8.10*** (1.871)	7.59*** (1.997)	8.10*** (2.412)	7.77*** (2.153)	7.63*** (1.983)
Observations	112	112	112	112	112	112
R-squared	0.331	0.447	0.340	0.097	0.215	0.309

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Appendix I Diagnostic Tests, Rotating the Instrumental Variables

Diagnostic Test Results for instrument relevance and validity Equation 1 are presented in Table 6.6 in Chapter 6.

Table I.1 Summary results of diagnostic tests for instrument relevance and validity: Equation 2, Without *tithe*

	First-Stage F-test		(Underid)		(Weak id)
Variable	F(6, 93)	P-val	AP Chi-sq(2)	P-val	AP F(2, 93)
Proportion of land cultivated (propland)	2.89	0.0125	14.42	0.0007	5.99
Log network size (logsize)	50.07	0.0000	23.41	0.0000	9.72
Log network size*education	15.91	0.0000	13.72	0.0010	5.70
Log network size*radio	22.12	0.0000	56.37	0.0000	23.41
Group affiliation (affil)	3.57	0.0031	2.33	0.3126	0.97
Second Stage Estimation: GMM					
F(17, 94)					
Sargan-Hansen J statistic Chi-sq(1) (overidentification test of all instruments):		0.27			
Instrumented: propland logsize logsizeed2 logsizead affil					
Included instruments: depend loghh_age propkin _lcomid_2 _lcomid_3 kininf education store radio316 nfy2 propmult agehet					
Excluded instruments: areatot IQVst2 logmeansize logyearcomm iqvr ad iqved					

Table I.2 Summary results of diagnostic tests for instrument relevance and validity: Equation 3, Without *logyearcomm*

	First-Stage F-Test		(Underid)		(Weak id)
Variable	F(6, 93)	P-val	AP Chi-sq(2)	P-val	AP F(2, 93)
Proportion of land cultivated (propland)	2.51	0.0267	14.46	0.0007	6.00
Log network size (logsize)	47.02	0.0000	19.93	0.0000	8.27
Log network size*education	16.01	0.0000	10.78	0.0046	4.48
Log network size*radio	20.09	0.0000	64.56	0.0000	26.81
Group affiliation (affil)	2.91	0.0121	2.49	0.2877	1.03
Second Stage Estimation: GMM					
F(17, 94)					
Sargan-Hansen J statistic Chi-sq(1) (overidentification test of all instruments):		0.644			
Instrumented: propland logsize logsizeed2 logsizead affil					
Included instruments: depend loghh_age propkin _lcomid_2 _lcomid_3 kininf education store radio316 nfy2 propmult agehet					
Excluded instruments: areatot IQVst2 logmeansize tithe iqvr ad iqved					

Table I.3 Summary results of diagnostic tests for instrument relevance and validity: Equation 4, Without *logmeansize*

	First Stage F-test		(Underid)		(Weak id)
Variable	F(6, 93)	P-val	AP Chi-sq(2)	P-val	AP F(2, 93)
Proportion of land cultivated (propland)	2.63	0.0210	14.53	0.0007	6.03
Log network size (logsize)	22.74	0.0000	34.02	0.0000	14.12
Log network size*education	14.68	0.0000	14.18	0.0008	5.89
Log network size*radio	10.16	0.0000	33.32	0.0000	13.83
Group affiliation (affil)	2.04	0.0675	2.85	0.2410	1.18
Second Stage Estimation: GMM					
F(17, 94)					
Sargan-Hansen J statistic Chi-sq(1) (overidentification test of all instruments):		0.002			
Instrumented: propland logsize logsizeed2 logsizead affil					
Included instruments: depend loghh_age propkin _lcomid_2 _lcomid_3 kininf education store radio316 nfy2 propmult agehet					
Excluded instruments: areatot IQVst2 logyearcomm tithe iqvradiqved					

Table I.4 Summary results of diagnostic tests for instrument relevance and validity: Equation 5, With *domtime*

	First-Stage F-Test		(Underid)		(Weak id)
Variable	F(8, 91)	P-val	AP Chi-sq(4)	P-val	AP F(4, 91)
Proportion of land cultivated (propland)	1.94	0.0640	15.74	0.0034	3.20
Log network size (logsize)	17.58	0.0000	134.73	0.0000	27.37
Log network size* education	12.16	0.0000	31.16	0.0000	6.33
Log network size*radio	9.14	0.0000	95.72	0.0000	19.44
Group affiliation (affil)	1.55	0.1512	13.37	0.0096	2.72
Second Stage Estimation: GMM					
F(17, 94)					
Sargan-Hansen J statistic Chi-sq(1) (overidentification test of all instruments):		0.697			
Instrumented: propland logsize logsizeed2 logsizead affil					
Included instruments: depend loghh_age propkin _lcomid_2 _lcomid_3 kininf education store radio316 nfy2 propmult agehet					
Excluded instruments: areatot domtime IQVst2 logmeansize iqvradiqved logyearcomm tithe					

Table I.5 Summary results of diagnostic tests for instrument relevance and validity: Equation 6, With *firewood* variable

Variable	First-Stage F-test		(Underid)		(Weak id)
	F(8, 91))	P-val	AP Chi-sq(4)	P-val	AP F(4, 91)
Proportion of land cultivated (propland)	1.91	0.0681	15.98	0.0030	3.25
Log network size (logsize)	18.07	0.0000	131.06	0.0000	26.62
Log network size* education	12.44	0.0000	32.22	0.0000	6.54
Log network size*radio	9.03	0.0000	91.09	0.0000	18.50
Group affiliation (affil)	1.70	0.1099	16.75	0.0022	3.40

Second Stage Estimation: GMM

F(17, 94)	
Sargan-Hansen J statistic Chi-sq(1) (overidentification test of all instruments):	0.635

Instrumented: propland logsize logsizeed2 logsizead affil

Included instruments: depend loghh_age propkin _lcomid_2 _lcomid_3 kininf education store radio316 nfy propmult agehet

Excluded instruments: areatot fire IQVst2 logmeansize iqvradiqved logyearcommTithe

Appendix J OLS Results

VARIABLES	(1) logcsold	(2) logcsold
depend	-0.691 (0.482)	-0.533 (0.550)
loghh_age	-0.265 (0.358)	-0.389 (0.378)
propkin	0.880** (0.422)	
2.comid	1.583*** (0.275)	1.465*** (0.320)
3.comid	0.230 (0.360)	0.209 (0.347)
kininf	0.620** (0.244)	
education	-1.817** (0.872)	0.156 (0.265)
store	0.432* (0.247)	0.168 (0.275)
radio316	-0.431 (0.844)	1.094*** (0.226)
nfy	0.000** (0.000)	0.000 (0.000)
propmult	-1.064 (0.842)	
agehet	-0.016 (0.021)	
propland	-0.404 (0.298)	-0.623* (0.358)
logsize	-0.844** (0.374)	
1.education#c.logsize	0.938** (0.384)	
1.radio316#c.logsize	0.622* (0.356)	
affil	0.145 (0.095)	0.165 (0.104)
Constant	7.002*** (1.380)	6.110*** (1.355)
Observations	112	113
R-squared	0.595	0.490

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Appendix K Heckman Selection and Censored Tobit Models of smallholder market participation with transactions costs

K.1 Heckman Selection Model

The model is estimated using the two-stage procedure suggested by Heckman (1979). Consider a household's crop marketing decision separated into two stages – whether or not to sell; then, given that they do sell, how much. This can be shown by the following:

Outcome equation:

$$y_i = \mathbf{X}_i\beta + u_i \quad \text{if } y_i^* > 0 \quad (\text{K.1})$$

y_i = not observed otherwise

Whether or not the value of the outcome equation (K.2) is observed can be denoted by a dummy D_i . Observation of the dependent variable y_i is a function of a value of another equation – the selection equation (6.3), which relates a latent variable y_i^* (propensity to be included in the sample) to some observed characteristics \mathbf{Z}_i . The variables in \mathbf{X}_i and \mathbf{Z}_i may overlap, but if they are identical the model may not be identified.⁵⁵

Selection equation:

$$y_i^* = \mathbf{Z}_i\gamma + e_i \quad (\text{K.2})$$

$$\begin{aligned} D_i &= 1 && \text{if } y_i^* > 0 \\ D_i &= 0 && \text{otherwise} \end{aligned}$$

Where \mathbf{Z}_i is a vector of observed variables that determine whether or not to participate (but not by how much), and u_i and e_i are jointly normally distributed.

⁵⁵ Further, if \mathbf{X}_i and \mathbf{Z}_i are not completely independently distributed (ie have variables in common or are correlated) the estimated slope coefficients will be biased because omitted variable in the regression that is correlated with the \mathbf{X}_i . Further, even if the \mathbf{X}_i and \mathbf{Z}_i are independent, the data is non-randomly missing, which introduces heteroscedasticity into error term and OLS is not efficient.

To retain sampling weights in the estimation it is necessary to estimate the model ‘by hand’ in two separate steps rather than simultaneously using maximum likelihood. However, this means that the standard errors in the second stage equation need to be corrected by modifying the variance-covariance matrix to correct for bias induced by the selection stage of the model (see Greene, 2003; Main and Reilly, 1993; Bellemare and Barrett, 2006). However, using the `svy:` prefix in Stata 11.2 means that it is impossible to recover the standard error of the estimation (root mean square error) and thus make the necessary modification. Standard errors may be estimated using Jack-knife (non-parametric) method in compromise, which is easier to apply than bootstrap when data are collected using complex sampling structure, such as the multistage sampling and varying sampling weights employed in this research. Jack-knife takes subsamples of the data, systematically re-estimating the model leaving out one observation at a time.

In estimating the Heckman model, in order to identify the model only capital variables that proxy for variable transaction costs are used to explain the amount sold, whereas both fixed and variable transactions costs variables are used to explain the decision to enter the market.

IMR is insignificant in second stage regressions. Just 21 out of 136 households reported no crop sales over the survey period, which may explain this. Further, it is not possible to achieve convergence when estimating the Heckman model with the data using survey weights and instrumental variables. Thus the model can only be estimated under the assumption of exogeneity. Results are given below in Table K.1.

K.2 Censored Tobit

The underlying theoretical model is similar to the probit threshold model set out in equation (K.1) above, where y^*_i is the latent variable:

$$y^*_i = \mathbf{X}_i \boldsymbol{\beta}' + e_i$$

(K.3)

where $e_i \sim N(0, \sigma^2)$ and $y_i^* \sim N(\mathbf{x}_i \boldsymbol{\beta}, \sigma^2)$ and $i=1, \dots, n$

if $y_i^* > 0$, then $y_i = y_i^*$

if $y_i^* \leq 0$, then $y_i = 0$

The empirical model is therefore:

$$y_i = \mathbf{X}_i \boldsymbol{\beta} + u_i$$

(K.4)

This is estimated as an interval regression, left-censored at zero, both with and without social networks variables. Parameter estimates are identical to those obtained from a censored Tobit but allows us to accommodate complex survey design using the svy: prefix. Once again, it was not possible to achieve convergence in estimating the model with instrumental variables. Results are given alongside Heckman Selection results in Table K.1 below.

Table K.1 Heckman Selection Model and Censored Tobit results – assuming regressor exogeneity

EQUATION	VARIABLES	Heckman Selection Model		Censored Tobit
		selection	outcome	
		(1) market	(2) logcsold	(3) model
	depend	2.057 (1.333)	-0.528 (0.654)	0.426 (0.928)
	loghh_age	-4.379*** (0.979)	-0.347 (0.462)	-2.252*** (0.689)
	propkin	0.283 (1.031)	0.801* (0.471)	1.278 (0.990)
	_lcomid_2	-0.309 (0.627)	1.538*** (0.348)	1.245** (0.611)
	_lcomid_3	-0.580 (0.595)	0.083 (0.399)	-0.075 (0.700)
	kininf	0.446 (0.419)	0.541* (0.309)	1.219** (0.486)
	education	1.413 (1.390)	-1.609 (1.118)	-0.334 (1.956)
	store	0.395 (0.577)		0.723* (0.404)
	radio316	6.027*** (2.038)		3.412* (1.971)
	nfy	-0.000 (0.000)	0.000 (0.000)	0.000** (0.000)
	propmult	-2.201 (1.858)	-1.167 (1.001)	-2.556* (1.323)
	agehet	-0.004 (0.044)	-0.021 (0.031)	-0.071 (0.049)
	propland	3.189** (1.225)	-0.197 (0.428)	1.569** (0.764)
	logsize	1.113** (0.552)	-0.545 (0.422)	0.345 (0.696)
	logsizeerad	-1.855** (0.791)	0.483*** (0.124)	-0.720 (0.831)
	logsizeed2	-0.027 (0.662)	0.854* (0.498)	0.647 (0.809)
	affil	0.299* (0.168)	0.190* (0.109)	0.353** (0.156)
	Constant	12.383*** (3.287)	6.600*** (1.909)	8.770** (3.528)
	mills_3		0.527 (0.730)	
	Constant			0.649*** (0.089)
	Observations	133	112	133
	R-squared		0.586	

Appendix L: Diagnostic Tests for GMM estimation without social networks

For comparison with the GMM estimation in Table 6.7 in Section 6.6.1 of Chapter 6, the model was re-estimated omitting all social networks variables (network size, proportion of kin, proportion of multiplex ties, influential kin, age heterogeneity of alters, interactions between network size and education and radio ownership) but retaining the group affiliation variable, capturing household membership of community level groups and societies in order to test if it is important to know more about the network than simply whether a person is a group member or not. Diagnostic test results are given in Table L.1. The summary results for first-stage regression show that we fail to reject the partial R-squared of excluded instruments and thus we have not identified the model and will not learn anything from the IV results.

Table L.1 Summary results of diagnostic tests for instrument relevance and validity:

GMM without social networks variables

First-stage regressions	Proportion of land cultivated	Group affiliation
Partial R-squared of excluded instruments	0.28	0.27
F-test of excluded instruments ~F(3, 101)	5.61	1.49
Under-id Angrist-Pischke Chi-sq(3)	7.08	1.93
Weak id Angrist-Pischke F(2, 101)	3.17	0.86
Underidentification tests		
Kleibergen-Paap rk LM statistic Chi-sq(2)	Chi-sq(2)=1.40	
Cragg-Donald N*minEval statistic Chi-sq(2)		
Weak identification statistics		
Cragg-Donald Wald F statistic H0: equation weakly identified	0.46	
Kleibergen-Paap Wald rk F statistic	0.46	
Anderson-Rubin Wald test of joint significance of endogenous regressors B1 in main equation. Ho: B1=0 and orthogonality conditions are valid:		
F-statistic ~ F(3,101)	3.47	
Chi-sq(3)	11.64	
Stock-Wright LM S statistic Ch—sq(3)	7.69	
Number of observations N	113	
Number of regressors K	11	
Number of endogenous regressors K1	2	
Number of instruments L	12	
Number of excluded instruments L1	3	
Second Stage Estimation: GMM		
F(10,102)	5.50	
Underidentification test (Kleibergen-Paap rk LM statistic)	1.40	
Weak identification test (Cragg-Donald Wald F statistic)	0.46	
Hansen J statistic (overidentification test of all instruments): Chi-sq(1)	0.04	
Instrumented: propland affil		
Included instruments: depend loghh_age _lcomid_2 _lcomid_3 education store radio316 nfy2		
Excluded instruments: areatot logyearcomm tithe		