Social networks, community-based development and empirical methodologies



Bet Caeyers

University College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Economics

Trinity Term 2013

Social networks, community-based development and empirical methodologies

Bet Caeyers

University College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Economics

Trinity Term 2013

Abstract

This thesis consists of two parts: Part I (Chapters 2 and 3) critically assesses a set of methodological tools that are widely used in the literature and that are applied to the empirical analysis in Part II (Chapters 4 and 5).

Using a randomised experiment, the first chapter compares pen-and-paper interviewing (PAPI) with computer-assisted personal interviewing (CAPI). We observe a large error count in PAPI, which is likely to introduce sample bias. We examine the effect of PAPI consumption measurement error on poverty analysis and compare both applications in terms of interview length, costs and respondents' perceptions.

Next, we formalise an unproven source of ordinary least squares estimation bias in standard linear-in-means peer effects models. Deriving a formula for the magnitude of the bias, we discuss its underlying parameters. We show when the bias is aggravated in models adding cluster fixed effects and how it affects inference and interpretation of estimation results. We reveal that two-stage least squares (2SLS) estimation strategies eliminate the bias and provide illustrative simulations. The results may explain some counter-intuitive findings in the social interaction literature.

We then use the linear-in-means model to estimate endogenous peer effects on the awareness of a community-based development programme of vulnerable groups in rural Tanzania. We denote the geographically nearest neighbours set as the relevant peer group in this context and employ a popular 2SLS estimation strategy on a unique spatial household dataset, collected using CAPI, to identify significant average and heterogeneous endogenous peer effects.

The final chapter investigates social network effects in decentralised food aid (free food and food for work) allocation processes in Ethiopia, in the aftermath of a serious drought. We find that food aid is responsive to need, as well as being targeted at households with less access to informal support. However, we also find strong correlations with political connections, especially in the immediate aftermath of the drought.

Acknowledgements

I am greatly indebted to a number of colleagues, friends and family members who each have contributed to this thesis in various different ways. I am most grateful to my thesis supervisor Professor Stefan Dercon, whose encouragement and support - from the MSc until the final stages of my DPhil - enabled me to complete this thesis in all sanity. He seems to have some sort of invisible sensor telling him exactly when to push one forward whenever one is stuck and when to give one the space to focus when things are coming along. I also thank my college advisor, Professor Sujoy Mukerji, who has been a great source of inspiration during my tutoring experience at University College, and Professor Oriana Bandiera and Professor Marcel Fafchamps for their willingness to read and assess this thesis. Their thoughtful comments and suggestions were much appreciated and will certainly be used to prepare this work for further publication.

I would like to thank my numerous colleagues and friends from the Economics Department. In particular, Simon and Matt, for being among the very (if not only) few people who took time to carefully go through what is no doubt the most boring chapter of this thesis *and* to provide useful comments on it. Matt, Naureen and Sebastian, for keeping up with (babbly) me as your neighbour in the department and – most importantly - for being such wonderful friends. I am also grateful to University College, not only for offering me the Senior Scholarship but also for providing me with friends for life. Thank you, George, for a great deal of laughter, lovely music, beautiful beaches and your healing raki. Katri, for those cosy late evening social gatherings, those early morning physical exercises, and for your exceptional kindness throughout. Ricarda, I don't know where to start...For being the best house mate one could wish for, for taking care of me whenever I was ill or whenever I needed some emotional steer. And thanks to all my other friends (you all know who you are) who made the otherwise stressful Oxford experience so much more enjoyable. Massive thanks go to Sonya, James and little Neave, who gave me a warm home and made me feel part of their lovely family. Sonya, you're the best. James, you make me laugh. Neave, you're my little star.

My thanks extend to my friends from outside of Oxford. In particular, I owe my deepest gratitude to Kurt and his family, for their emotional support, their patience and understanding. To Catharina and Verena, my personal mentors since the age of 17, who certainly did not realise what they signed up for when becoming my friends, but who have far exceeded their terms of reference during my DPhil period. I take this opportunity to thank my friend and former supervisor Professor Wilfried Pauwels. I am extremely grateful to him for the incredible support and advice he has given me at the most crucial moments of my study career. This thesis would never have materialised without his mentorship and his unconditional faith in my abilities.

Special thanks also go to Andrew, whose genuine love, support and delicious cooking (!) kept me sane and healthy, especially during the final most stressful stages of my thesis. Finally, I would like to thank my parents, my three sisters and their families for their warm-hearted support over the course of my studies. I would not have been able to complete this thesis without their continuous affection and encouragement. Special thanks go to my twin sister, who not only provided me with incredibly valuable inputs during my thesis field work in Tanzania, but who has accompanied me throughout my entire thesis journey with her great sense of humour and cheerfulness. Thanks, Liesje, I owe you loads.

Word count

Excluding the references, this thesis consists of approximately 66,220 words (calculated using the number of words on page 5 (308) multiplied by the number of pages (215).

Declaration

Two chapters (out of four core chapters) from this thesis (Chapters 2 and 5) are based on two jointly authored published articles. Chapter 2, previously published in *Journal of Development Economics*, was co-authored with Neil Chalmers (independent IT consultant) and Joachim De Weerdt (Economic Development Initiatives Ltd). Neil Chalmers was in charge of the development of the computer-assisted personal interviewing (CAPI) survey software that was used in the field to collect the data from our randomised field experiment. I led the survey design of this project. Joachim De Weerdt and I jointly developed the published article, for which I carried out a large part of the analytical work. Chapter 5 was previously published in *Economic Development and Cultural Change*, in co-authorship with Stefan Dercon. Given that this article built on my single-authored extended essay for the degree of MSc in Economics for Development, my contribution to the core ideas and empirical analysis of this chapter was substantial.

Contents

1. Introduction 1

1.1. Overview 1

1.2. Thesis outline and contributions 4

1.2.1. Part I: Empirical methodologies 4

1.2.2. Part II: Empirical analysis 7

PART I: EMPIRICAL METHODOLOGIES

2. Improving consumption measurement and other survey data through CAPI: evidence from a randomised experiment 11

2.1. Introduction 11

2.2. Experimental set-up and hypothesised effects 14

2.2.1. Set-up 14

2.2.2. Experiment 1: the effect of validation checks 16

2.2.3. Experiment 2: bundle of other CAPI features 18

2.2.4. Implications for sample bias and analysis 21

2.2.5. Interviewer effects 21

2.3. Descriptive statistics 22

2.4. Errors and sample size reduction 25

2.4.1. Methods of analysis 25

2.4.2. Routing errors 26

2.4.3. Unlikely and impossible entries 28

2.4.4. Implications for sample size reduction and sample bias 29

2.4.5. Interaction with interviewer characteristics and survey period..... 32

2.5. Measuring consumption 35

2.5.1. Food consumption 35

2.5.2. Total consumption, poverty and inequality 40

2.5.3. Testing for classical measurement error in consumption 43

2.6. Further dimensions of comparison..... 49

2.7. Conclusion 51

2.A. Chapter 2 Appendix..... 55

2.A.1. Screen shots of the CAPI application used in the experiment 55

2.A.2. Validation checks in full CAPI 56

2.A.3. Respondent's perception of the interview..... 59

2.A.4. More examples of pictures used in the CAPI consumption section..... 60

3. Exclusion bias in empirical social interaction models:

causes, consequences and solutions 61

3.1. Introduction..... 61

3.2. Basic model..... 65

3.3. Model allowing for cluster sampling 70

3.3.1. Motivation..... 70

3.3.2. Two empirical scenarios 71

3.3.3. Model..... 73

3.3.4. Exclusion bias in models adding cluster fixed effects..... 74

3.3.5. Exclusion bias in models omitting cluster fixed effects 76

3.3.6. Comparison results of models with and without cluster fixed effects..... 79

3.4. Solutions to the bias 80

3.4.1. Method suggested by GKN (2009)..... 80

3.4.2. 2SLS estimation strategies 82

3.5. Simulation results 84

3.6. Extensions of the basic model..... 93

3.7. Conclusion 97

3.A. Chapter 3 Appendix..... 100

3.A.1. Relationship between σ_u^2 and σ_ε^2 100

3.A.2. Deriving an expression for $\text{var}(\bar{y}_{k,-i})$ 100

3.A.3. Exclusion bias in models adding cluster fixed effects

(equal cluster sizes) 101

3.A.4. Exclusion bias in the between estimation model..... 105

3.A.5. Derivation of $p\lim(\eta^2)$ 106

3.A.6. Exclusion bias in models adding cluster fixed effects,

with varying cluster sizes..... 109

PART II: EMPIRICAL ANALYSIS

4. Peer effects in development programme awareness of vulnerable groups in rural Tanzania 111

4.1. Introduction..... 111

4.2. Background of Tanzania Social Action Fund II 116

4.2.1. Programme objectives and resource allocation process 116

4.2.2. Official programme awareness strategies and the scope for peer effects ... 117

4.3. Data 120

4.3.1. Data source 120

4.3.2. Peer group measure 122

4.3.3. Data description 127

4.4. Empirical strategy 130

4.4.1. Empirical specification 130

4.4.2. Empirical challenges and identification strategy 134

4.5. Results 141

4.5.1. Average peer effects 141

4.5.2. Varying the size of the neighbours set 147

4.5.3. Heterogeneous peer effects 148

4.6. Robustness tests 154

4.6.1. Correcting for spatial dependence of standard errors within hamlets 154

4.6.2. Other identification challenges 155

4.7. Conclusion 158

4.A. Chapter 4 Appendix 161

4.A.1. Population and sample information 161

4.A.2. Interaction between peer effect and geographical distance of peers 162

4.A.3. Evidence of exclusion bias in OLS estimation 163

5. Political connections and social networks in targeted transfer programmes: evidence from rural Ethiopia 167

5.1. Introduction 167

5.2. Food aid targeting in Ethiopia 171

5.3. The role of social and political networks in food transfers 174

5.4. Data source and descriptive statistics 178

5.5. Empirical analysis 185

5.5.1. Model specification 185

5.5.2. Free food distribution targeting results (OLS) 190

5.5.3. Food for work targeting results (OLS) 194

5.5.4. Identification concerns and 2SLS 199

5.6. Channels: improved information or favouritism? 206

5.7. Conclusion 208

6. Conclusion 212

References 216

List of Figures

2.1 Picture of pieces of cassava (used in CAPI consumption section) 36

2.2 Example of a consumption data summary report 37

2.3 CDF of annual consumption per adult equivalent unit (AEU) 42

2.4 Screen shot of section 'household head information' 55

2.5 Screen shot of section 'roster' 55

2.6 Bunches of bananas (with a bike and chair as a reference) 60

2.7 Fish (with a UMPC unit as a reference) 60

3.1 Rejection rates $H_0 : \beta_1 = 0$ (basic model OLS: $N = 50 ; k = 10$) 87

3.2 Rejection rates $H_0 : \beta_1 = 0$ (case 2 – FE: $N = 200 ; L = 20 ; k = 10$) 91

3.3 Rejection rates $H_0 : \beta_1 = 0$ (case 2 – 2SLS: $N = 200 ; L = 20 ; k = 10$) 93

4.1 Sampled districts (each containing 20 sampled villages) 121

List of Tables

2.1 Summary statistics on interviewer and survey characteristics 23

2.2 Effect of CAPI and checks on data quality..... 27

2.3 Interaction with household characteristics 30

2.4 Interaction with interviewer characteristics..... 33

2.5 Interaction with survey period 34

2.6 Summary statistics on calorific intake..... 39

2.7 Consumption, poverty and inequality 41

2.8 Implications of measurement error in the explanatory variable 45

2.9 Implications of measurement error in the dependent variable..... 48

2.10 Subset of a total of 366 validation checks used in full CAPI 56

2.11 Fifteen most common missing values outside the consumption section in pen-and-paper interviewing (PAPI)..... 57

2.12 Ten most common missing values in consumption section in pen-and-paper interviewing (PAPI) 58

2.13 Respondent's perception by survey method (%) 59

3.1 Simulation results $\hat{\beta}_1$: basic model..... 86

3.2 Simulation results $\hat{\beta}_1$: case 1 ($N_P = N = 200$) 88

3.3 Simulation results $\hat{\beta}_1$: case 2 ($N = 200$; $N_P = L$)..... 90

3.4 Simulation results on $\hat{\beta}_1$: 2SLS 92

4.1 Elderly/disabled versus other households: a comparison in characteristics..... 128

4.2 Characteristics of informed and uninformed elderly/disabled households 131

4.3 Determinants of being informed: 10 nearest neighbours specification (OLS).... 142

4.4 Two-stage least squares first-stage regression..... 144

4.6 Endogenous peer effect for varying neighbours set sizes 148

4.7 Heterogeneous peer effects in the probability of being informed (2SLS) 150

4.7 *Continued*..... 151

4.8 Robustness to Conley's (1999) correction of standard errors 154

4.9 Samples restricted to the least mobile households (HHs) (2SLS) 156

4.10 Evidence of exclusion bias in ordinary least squares (OLS) estimation 166

5.1 Characteristics of free food distribution (FFD)/food for work (FFW) beneficiary households (2002–4) and non-FFD/FFW beneficiary households: based on peasant association criteria (1999) 180

5.2 Characteristics of free-food-receiving (2002–4) and non-free-food-receiving households: based on social networks[a] 183

5.3 Determinants of free food distribution (FFD) 191

5.4 Determinants of food for work (FFW) participation..... 196

5.5 Determinants of food for work (FFW) rewards among FFW-participating households (OLS) 198

5.6 Instrumental variable (IV) estimations..... 203

5.7 Change in informal social safety net (ISSN), 1999–2004..... 205

Chapter 1

Introduction

1.1. Overview

In the developing world, millions of people rely on welfare programmes for survival. Food insecure households rely on food aid, especially during crisis situations such as droughts (World Food Program, 2012), and the physically and mentally challenged rely on various other forms of social protection for the improvement of their living standards (World Bank, 2012). However, the effectiveness of these initiatives to reduce poverty crucially depends on how well they reach the intended beneficiaries. Who is aware of these policies? Who has access to their benefits? This thesis demonstrates the importance of *social networks* in accessing welfare benefits in sub-Saharan Africa. In particular, this thesis sheds light on the interaction between social networks and community-based development (CBD) initiatives, that is, programmes that give control over project design, implementation and/or investment resources to local communities and governments. In addition, this thesis addresses some methodological concerns that are related to the empirical analysis.

Since the seminal work by Granovetter (1975), Coleman (1988) and Putnam et al. (1993), an extensive literature has been developed on the merit of social networks for the well-being of the poor in the presence of market failures (for a review, see Durlauf and Fafchamps, 2008). Informal risk-sharing and credit arrangements with relatives and friends may compensate for a household's restricted access to formal insurance and credit markets in developing countries (Townsend, 1995; Dercon, 2002; Fafchamps and Lund, 2003; De Weerdt and Dercon, 2006). Furthermore, in a world of imperfect information, social networks may reduce search costs through information sharing and learning. As such, social networks have been shown to create job opportunities

(Granovetter, 1975), increase trader profitability (Fafchamps and Minten, 2002) and enhance welfare-improving technology adoption (Bandiera and Rasul, 2006; Conley and Udry, 2010).

However, the merit of social networks in terms of access to welfare benefits is less well understood. There is evidence from developed countries that low uptake of public welfare programmes may be a result of information constraints (Daponte et al., 1999; Hernanz et al., 2004; Aizer, 2007). Keefer and Khemani (2004) and Berg et al. (2013) show how insufficient programme awareness forms a root cause for inadequate social services in India. Using a randomised experiment on US university employees, Duflo and Saez (2003) highlight the importance of information sharing among peers for enrolment in a retirement benefit programme. However, there is currently little evidence from the developing world on social learning in development programme awareness. This thesis emphasises the role of information sharing and other attributes of social networks in CBD initiatives in sub-Saharan Africa.

Analysing these processes within the framework of CBD is important. Over the last decade, international aid organisations have increasingly favoured bottom-up approaches. In 2006 (i.e. at the time when the ideas for this thesis originated) the share of CBD projects in the World Bank's total lending portfolio had reached over 9% (World Bank, 2006). There are growing concerns, however, about the targeting effectiveness of these rapidly expanding initiatives, particularly in sub-Saharan Africa (for a recent review, see Mansuri and Rao, 2013).

When central government agents lack detailed information about the individual profiles of their citizens, there is much scope, indeed, for decentralised community-based systems to improve the service delivery: local communities are likely to be better informed about the current needs of their community members and, therefore, are possibly in a better position to design development projects and assign beneficiaries.

Nevertheless, it should not be presumed that community-based institutions automatically reflect the preferences of the poor (Conning and Kevane, 2002).

A necessary, although insufficient, condition for giving voice to the most vulnerable people is widespread programme awareness. Various grassroots programmes rely on village meetings to enhance village participation and monitoring (Goetz and Jenkins, 2001; Bjorkman and Svensson, 2007; Khemani, 2007; Olken, 2007; Banerjee et al., 2010; Shankar et al., 2010). However, as a result of their physical and socioeconomic conditions, vulnerable groups (which are usually targeted by social protection programmes) may have only limited access to such primary sources of information. The importance of social learning through peers, for effective CBD outreach, has not been recognised in the literature to date. Using household survey data from rural Tanzania, Chapter 4 of this thesis is devoted entirely to this empirical question.

Moreover, the fact that an individual knows some people may not be sufficient to gain access to CBD programmes. *Who* the individual knows may be more important. Poor individuals may be well aware of a programme, but they may not know the right people to secure them access to it. The literature on social networks distinguishes between horizontal and vertical social ties (Putnam et al., 1993). Whereas the former refer to contacts with similar degrees of power, the latter refer to links with the political elite. There is a large literature on the local political economy of decentralised service delivery, particularly from South Asia (Bardhan and Mookherjee, 2000, 2006; Conning and Kevane, 2002; Platteau and Gaspart, 2003; Araujo et al., 2008). The evidence from sub-Saharan Africa is limited, however. Chapter 5 sheds light on the local political economy of targeted transfers in the context of rural Ethiopia.

To fill these gaps in the empirical literature, my analysis relies on a set of methodological tools. Whereas the validity of the majority of these strategies is well documented in the literature, a few methodological concerns remain, as follows. (i) The

household data that I use in one of the empirical chapters (Chapter 4) were collected using computer-assisted personal interviewing (CAPI) techniques. Although such advanced survey tools are expected to improve the quality of data collection, it is unclear to what extent they actually do so and what effect this has on analysis. (ii) My empirical analysis makes use of the linear-in-means peer effects model. Although this model is standard in empirical social interaction studies (for a review, see Durlauf and Ioannides, 2010), Guryan et al. (2009) have argued that it may produce biased results. It is not clear, however, to what extent, and under which conditions, this is the case. The methodological contribution of this thesis is to address both these concerns, which have largely been ignored in the literature to date.

1.2. Thesis outline and contributions

This thesis consists of two parts: Part I comprises two chapters (Chapters 2 and 3) that deal with the methodological issues set out in the previous section, and Part II includes two applied chapters (Chapters 4 and 5) that address the empirical questions that were postulated above.

1.2.1. Part I: Empirical methodologies

Chapter 2 (co-authored) reports on a randomised survey experiment among 1840 households in Pemba, Tanzania, designed to compare pen-and-paper interviewing (PAPI) with CAPI. The specific CAPI application used in the experiment was also used to collect the household data that are analysed in Chapter 4 to investigate the role of social networks in programme awareness. Chapter 2 puts these data into context. Apart from a study by Fafchamps et al. (2012), I am not aware of any other systematic attempt to study the effect of CAPI on data quality and analysis. The main contribution of Chapter 2, therefore, lies in informing researchers on the costs and benefits of switching from PAPI to CAPI. More generally, cautioning against the limitations of PAPI survey

data is important, in view of its still dominant position in current research in development economics.

Chapter 2 begins by comparing CAPI with PAPI in terms of the number of errors encountered in the completed questionnaires and investigates whether these errors are randomly distributed across the sample. If these error counts are correlated with household characteristics, they may introduce sample bias when dubious observations need to be dropped. We find that PAPI questionnaires have, on average, a higher error count than CAPI questionnaires and that these error counts are correlated with respondent characteristics such as gender of the household head and household size. We also explore heterogeneity in error counts across interviewer characteristics and learning effects over time. Chapter 2 reveals that, whereas interviewers with less education and less survey experience tend to make relatively more survey mistakes on paper, especially at the beginning of the field work, the differences in the data collected by enumerators with varying skills are not statistically significant in CAPI surveys. Given the importance of consumption data in poverty analysis, this chapter has special focus on the differences between CAPI and PAPI in terms of consumption measurement. This includes an examination of the effects of measurement error on poverty analysis. Probably one of the most important practical findings of this chapter is that poverty analysis results depend on the type of survey methodologies that are used. For instance, data collected through PAPI yield relatively lower poverty estimates and relatively higher inequality estimates than data collected using CAPI survey techniques. Finally, we compare CAPI with PAPI in terms of interview length, costs and respondents' perceptions.

Chapter 3 continues the methodological discourse by formalising a cited, but so far unproven, source of estimation bias in the standard linear-in-means peer effects model, which is used in the empirical analysis in Chapter 4. Guryan et al. (2009) have

intuitively argued that, in such models, ordinary least squares estimates of endogenous peer effects are downward biased and inconsistent, if individuals are systematically excluded from the empirical construction of their peer groups. For instance, the bias is expected to appear in models analysing peer effects in educational achievement, if a student's peers are defined by all her classmates, *excluding* herself. However, the literature to date lacks a formal treatment of this bias. This is surprising, given that, as illustrated by simulations (Guryan et al, 2009; Wang, 2010), the bias may significantly affect the estimation results of empirical social interaction models. Many questions remain. Which specific parameters are driving the bias? Does the bias become negligent with large sample sizes? Which empirical specifications are most likely to be affected by this? Does it disturb inference? To what extent does it confound the interpretation of the estimation results? And, most importantly, how can we avoid the problem? These and various other questions are dealt with in Chapter 3.

In addition to deriving an expression for the magnitude of the bias – hence identifying the parameters that are underlying it – an important contribution of Chapter 3 lies in revealing that the bias becomes significantly more severe in models adding cluster fixed effects (such as peer effects models adding classroom dummies when peer groups are formed within the classroom). Importantly, this chapter highlights that in the latter models the bias does not decrease as the sample size tends to infinity. In addition, Chapter 3 discusses various limitations associated with the solution suggested by Guryan et al (2009) and shows how two-stage least squares (2SLS) estimation techniques are generally more effective in eliminating the bias. The latter result may explain the counter-intuitive, yet common, observation in the social interaction literature of obtaining 2SLS estimates of endogenous peer effects that are larger than their OLS counterparts.

1.2.2. Part II: Empirical analysis

Chapter 4 is the first applied chapter of this thesis. It focuses on the potential of social networks to improve the CBD programme awareness of the elderly and disabled in rural Tanzania. Given their physical and socioeconomic conditions, these vulnerable groups are expected to face the greatest information constraints and, hence, constitute an especially relevant sample. The programme under case study concerns the Tanzania Social Action Fund II (TASAF II), Tanzania's flagship community-driven development programme. Baird et al. (2011) find that less than 50% of the eligible population was informed about TASAF II, after the beneficiaries had been selected. These authors examine the role of the household's *direct* access to information as determinants of being informed, such as village meeting attendance, newspapers and the ownership of radios and phones. In rural Tanzania, however, as a result of physical and socioeconomic constraints, the elderly and disabled have only limited access to such primary sources of information. Using a rich spatial dataset, Chapter 4 contributes to this literature by analysing the extent to which social interactions with geographically nearest neighbours may compensate for these information constraints.

The analysis makes use of a unique dataset on the households of *all* elderly and disabled citizens in 100 Tanzanian villages. These data were collected in 2008, under my supervision, at baseline of a World Bank impact evaluation of TASAF II. Given that the surveys were conducted using CAPI technology, global positioning system coordinates were automatically captured, ensuring high-quality data, as will be demonstrated in Chapter 2. The richness of the data allows me to control for 'hamlet' fixed effects, something that is rarely feasible in practice (especially when focusing on population subgroups such as the elderly and disabled).[1]

[1] The hamlet is the smallest administrative unit in rural Tanzania.

This chapter uses results from the social interaction literature (Manski, 1993; Brock and Durlauf, 2001; Moffitt, 2001; Bramouille et al., 2009; De Giorgi et al., 2010) together with the methodological insights of Chapter 3 to deliver consistent and unbiased estimates of the endogenous peer effect in programme awareness. My strategy to estimate the linear-in-means model employs a popular two-stage least squares (2SLS) method, using characteristics of 'excluded neighbours' as instruments. As will be shown in Chapter 3, using this particular 2SLS estimation strategy has the additional advantage of eliminating the exclusion bias. I find that every additional informed person in the set of ten nearest sampled neighbours improves a vulnerable household's probability of being informed by 8 percentage points. This Chapter 4 further contributes to the literature by taking the theoretical results on the exclusion bias (derived in Chapter 3) to practice. It demonstrates the empirical importance of the consequences of the bias and uses simulation results to correct for this.

The unique size of the dataset also enables me to conduct a heterogeneity analysis of peer effects across different socioeconomic groups. This allows me to gauge the profile of (i) citizens who are most and least sensitive to information in the neighbourhood (e.g. whereas households that have blood relatives in the village are more responsive to information in the neighbourhood, households that do not belong to the major religious group of the village are less likely to benefit from it) and (ii) those who are most effective in transmitting information to the elderly and disabled (e.g. households that hold political positions in the village). The empirical findings of this chapter are of interest to anyone involved in the design, implementation and monitoring of information dissemination campaigns as part of development programmes.

Chapter 5 (co-authored) moves the geographic focus to Ethiopia and considers the role of social networks in this country's decentralised food aid allocation process in the aftermath of a serious drought (2002–2003). This geographic focus is important, given

that Ethiopia is one of the poorest settings in the world, experiencing famine and food insecurity for over a century (Pankhurst, 1986; World Food Program, 2006). As a result, it has been the largest recipient of food aid in the last decade, after North Korea (World Food Program, 2006). Whereas existing evidence points to significant biases against the poor in Ethiopia's food distribution process (Sharp, 1997; Clay et al., 1999; Jayne et al., 2001, 2002; Coady et al., 2004; Asfaw, 2006; Gilligan and Hoddinott, 2007), it is unclear how transfers are actually reaching the final recipients. Chapter 5 contributes to this literature by analysing the local political economy of these public service deliveries, focusing on the role of social and political networks in food aid access (considering both free-food and food-for-work programmes).

Whereas Chapter 5 is, in terms of theory and methodology, relatively straightforward, its value added clearly lies in its findings. Although we find evidence that food aid is somehow responsive to need, as well as targeted to households with less access to informal support, we also reveal, probably for the first time in African contexts, a strong relationship between access to food aid and political affiliation.

PART I

Empirical methodologies

Chapter 2

Improving consumption measurement and other survey data through CAPI: evidence from a randomised experiment[2,3,4]

2.1. Introduction

Whereas the analysis of survey data has benefited from the information technology revolution, most data collection in developing countries still uses traditional pen-and-paper interviewing (PAPI). In computer-assisted personal interviewing (CAPI) the interviewer reads questions from the screen of a handheld device, preloaded with the questionnaire, to the respondent. The respondent's answers are immediately entered into the device, which eliminates the need for manual rekeying of the data. The computer also automates the routing through the questionnaire and enables the interviewer to run a set of consistency checks during the interview, so that anomalies can be resolved with the respondent. These and numerous other features are believed to improve data quality, but it is unclear to what extent they actually do so and what effect this has on analysis. Furthermore, there is currently no empirical evidence from the developing world on how a switch from PAPI to CAPI would influence the length of the interview, respondents' perceptions, the cost of the survey, requirements on the level of education of interviewers and so forth. This chapter reports on a formal experiment, designed specifically to compare CAPI and PAPI along these and other lines.

---

[2] Previously published as: Caeyers, B., Chalmers, N. and De Weerdt J., 2012. Improving Consumption Measurement and other survey data through CAPI: Evidence from a randomized experiment. *Journal of Development Economics* 98, 19–33.

[3] Co-authors: Neil Chalmers (independent consultant, who led the development of the CAPI application used in our experiment); Joachim De Weerdt (Economic Development Initiatives).

[4] We gratefully acknowledge financial support from the World Bank's multi-year research agenda in survey methodology (LSMS Phase IV). We appreciate permission from the Millennium Challenge Corporation to build on its existing survey in Pemba. We thank Kathleen Beegle, David McKenzie, Kinnon Scott and participants at the Conference on Survey Design and Measurement in Washington, DC, for feedback on the experiment's design and on an earlier draft of this chapter. The chapter was substantially improved after incorporating suggestions made by the editors of *Journal of Development Economics* and two anonymous referees. A previous version of the chapter was circulated under the title "A Comparison of CAPI and PAPI through a Randomized Field Experiment". Leonard Kyaruzi, Deogratias Mitti and Mujobu Moyo led the field teams, while Alessandro Romeo and Thaddaeus Rweyemamu took care of data entry of the paper questionnaires.

The study was built on an existing LSMS-style CAPI survey of 1200 households on the Island of Pemba in Tanzania.[5] The experiment consisted of randomly sampling, within the same enumeration areas, 320 additional households to be interviewed using restricted CAPI (with disabled consistency checks) and 320 using PAPI. This design allows for a detailed comparison of errors, outliers, interview times, respondents' perceptions, interviewer effects and costs across the three methodologies. Special focus was given to improving the collection of consumption data, which utilise many of the powerful features of the computer, including complex validity checks and the ability to show pictures on the screen. The experiment lends itself to comparing simple poverty and inequality measures across the experiments.

While CAPI survey applications are still relatively rare in international development, their use is no longer a novelty in developed country settings. Although the first computer-assisted telephone interviews were conducted by a US marketing firm in 1971, the first nationwide CAPI survey occurred only in 1987 in the Netherlands (Nicholls and de Leeuw, 1996). As CAPI became more popular for large-scale face-to-face surveys in western countries, researchers became more aware of its impact on the survey process and outcomes. It was found that interviewers and respondents reacted favourably to the technology (Couper and Burt, 1994; de Leeuw and Nicholls, 1996). Taylor (1998) shows that this remains true for respondents with, presumably, less exposure to modern technology, such as the elderly (i.e. those over 70 years old). Banks and Laurie (2000) report that the attrition rate in the British Household Panel Survey was not affected when it switched from PAPI to CAPI in 1998. The literature also indicates the potential of CAPI to reduce routing and other errors (de Leeuw, 2008). There have been a number of CAPI surveys in the developing world, an enumeration of which is beyond the scope of this chapter. IRIS (2011) provides an evaluation of various

[5] LSMS: Living Standards Measurement Study (World Bank).

CAPI software packages in terms of their functional performance in supporting complex household surveys. Apart from the paper by Fafchamps et al. (2012), however, we are not aware of any systematic attempt to study the effect on data quality and analysis.

The lack of evidence on how to reduce errors in surveys in developing countries stands in stark contrast to how much is known about the effects of measurement error in analysis (Bound et al., 2001; Chesher and Schluter, 2002). Classical measurement error is defined by Bound et al. (2001) as an error in the measurement of a particular variable which is uncorrelated with the true value of that variable, the true values of other variables in the model and any errors in measuring those variables. As we do not have independent, validation data in this experiment, we cannot directly measure the error to analyse its nature. We are, however, able to set up two testable hypotheses that should hold if measurement error is classical: in regression analysis, classical measurement error causes no bias when only the dependent variable has error, but attenuates the estimated coefficient on a single error-ridden explanatory variable.

We fail to reject the hypothesis that the introduced measurement error is classical, at least for consumption measurements and based on these two tests. There is some consolation in this finding, as non-random, mean-reverting errors negatively correlated with true values bias regression coefficients even when only the dependent variable has error. When an explanatory variable has such error, its coefficient may be biased either towards or away from zero (Gibson and Kim, 2007). Moreover, the main correction for measurement error bias – instrumental variables – is inconsistent when errors are correlated with true values (Black et al., 2000).

The next section describes the design of the experiment and the differences we hypothesise to exist between CAPI and PAPI. Section 2.3 provides descriptive statistics, whereas section 2.4 discusses estimation results pertaining to errors and sample size reduction. It shows that CAPI significantly reduces the number of inconsistencies per

survey. Some of these errors may require observations to be omitted from the analysis, which could bias the sample because missing variables are not randomly distributed. Section 2.5 analyses the nature of measurement error in consumption aggregates. It first compares nutrition, consumption, poverty and inequality data across the three experiments. It then hypothesises that error is introduced through PAPI and sets up two testable predictions to verify whether this measurement error is classical. The first is that regression coefficients on consumption as an independent variable should be attenuated. The second is that there is no bias in a model where the error-ridden variable is used as a dependent variable. We find that, despite the fact that error counts are higher in certain types of households, we cannot reject that (after cleaning) the introduced measurement error is classical. Section 2.6 looks at other dimensions of comparison, such as cost, length of the interview and respondents' perceptions. Section 2.7 makes some concluding observations.

2.2. Experimental set-up and hypothesised effects

2.2.1. Set-up

The experiment was run alongside an existing household survey on Pemba Island (which is part of Zanzibar, Tanzania). The main survey was conducted in July and August 2009 on behalf of the Millennium Challenge Account Tanzania (MCA-T) as a baseline to evaluate its rural roads upgrade programme. In total, 1200 households were interviewed – 15 in each of the 80 enumeration areas (EAs). All households were administered a full CAPI questionnaire using an ultra-mobile personal computer (UMPC), which is a handheld device with an 18-cm touch screen (a screen smaller than that of a laptop, but larger than that of a PDA). The CAPI software package used for this experiment, CWEST (Capture With Enhanced Survey Technology), has frequently

been used in the field, for instance by the LSMS team within the Poverty and Inequality Group of the Development Research Group at the World Bank.[6]

In the first experiment, we randomly selected four additional households per EA (320 in total) and interviewed these households using the same CAPI questionnaire, but with one important CAPI feature disabled: the system of consistency checks. The purpose of this experiment was to isolate the effect of consistency checks, which are believed to have an important impact on data quality, especially in the consumption data. In the remainder of this chapter, we will refer to this application as 'restricted CAPI', in order to distinguish it from the unrestricted 'full CAPI' application that included the system of consistency checks.

To investigate all other CAPI effects, as a bundle, the second experiment randomly selected another four households per EA to be interviewed using PAPI. The PAPI data were transferred to a computer using two-pass verification to minimise keystroke errors. Each of the four interviewers in a team conducted one restricted CAPI, one PAPI and three or four full CAPI interviews per cluster. For the restricted CAPI and PAPI, interviewers were allocated a specific household to interview at a specific time within the team's 2-day visit to the EA. This was done to ensure that questionnaires were not clustered by interviewer or by time.

All experimental questionnaires were conducted by the same 20 interviewers working on the main MCA survey. This increased the likelihood of contamination within the experiment, though it is hard to know the direction of the bias a priori. On the one hand, interviewers could learn about the types of checks that CAPI implements (something they may not have done had the questionnaire been purely on paper), but, on the other hand, interviewers could unlearn the practice of carefully verifying a

[6] CWEST is a CAPI application developed by the survey research firm EDI (Economic Development Initiatives). It has now transitioned into a package called Surveybe.

questionnaire at the end of the day as they became accustomed to the computer doing it for them. We tackle this contamination bias in two different ways. First, during training and fieldwork, interviewers were repeatedly instructed to check questionnaires at the end of the interview and again before submitting them to the supervisor. The supervisor, in turn, would check the questionnaires for errors. Questionnaires with errors that could not be resolved at base camp were returned to the interviewer, who was then required to revisit the household. Second, we have data to control for the number of months of experience that interviewers had using paper questionnaires and using electronic questionnaires.

The experimental questionnaire took, on average, 84 minutes to administer and included the following sections: control data, global positioning system (GPS) coordinates, household head details, household member roster, demographics, education, health, amenities, assets, livestock, agriculture and consumption.[7]

A few days after the questionnaires were conducted, a separate team of locally recruited interviewers returned to the four households per experiment to ask 13 simple questions about the experience of the respondent in participating in the survey.

2.2.2. Experiment 1: the effect of validation checks

The full electronic questionnaire included a comprehensive system of internal validation checks. The first experiment was set up to isolate the effects of these checks by comparing full CAPI with restricted CAPI. The checks are believed to have led to more accurate data capture, because they were run during the interview, at a time when they could still be resolved with the respondent. The check procedure did not run automatically, but was activated by the interviewer by manually clicking check buttons.

[7] The main questionnaire, implemented on behalf of MCA-T, included some additional sections on prices, transfers, shocks, credit, self-help groups and the like. To avoid these sections interfering with the experiment they were placed at the end of the main questionnaire. The full questionnaire is available from the authors upon request.

Some examples of screen shots of the electronic questionnaire, including check buttons, are available in Appendix 2.A.1. These checks were run at various stages during the interview, typically after completing all the questions on one screen. A final, global check was run at the end of the interview. The checking procedure was repeated by the supervisor at the end of each survey day, and once more by the data-processing team at headquarters after data transfer (usually the day after data collection). The full CAPI application contained 366 consistency checks, a subset of which is provided in Table 2.10 in Appendix 2.A.2 as an example. These checks fall into three broad categories, depending on whether they were designed to detect routing errors (248 checks), unlikely entries (61 checks) or impossible entries (57 checks). We will discuss each of these checks in turn.

Over two-thirds of the checks were aimed at detecting violations of the questionnaire's routing scheme. Routing errors occur by answering a question that is supposed to be skipped, or by skipping a question that is supposed to be answered. The questionnaire had a total of 152 variables, out of which 100 were dependent on previous answers and 52 were unconditional. Each unconditional question had a single check to detect a missing entry, whereas each conditional question had two checks: one detecting a missing entry and one detecting an entry made in a disabled field. Four routing checks turned out to have malfunctioned, leading to a total of 248 routing checks. Answers such as 'don't know' or 'refused' were not recorded as missing, but had their own codes.

Another 16% of the checks constituted checks detecting impossible entries or impossible combinations of entries. Some were simple range checks on a single variable; for example, verifying that the number of days for which a person reported illness in the past 4 weeks did not exceed 28 or ensuring that the value for a consumed quantity was not negative. Others checked consistency across variables, highlighting,

for example, situations in which the age someone started school exceeded his current age, or a member's relation to the head of the household was 'spouse' but the head's marital status was 'never married', or a male person had pregnancy-related problems. Some of these checks could have been avoided by restricting the range of permissible responses in the first place (see below).

The remaining 17% of the checks constituted checks detecting possible, but unlikely, entries, such as an uncommon number of cows or an uncommon expenditure value. Verifications for unlikely combinations of entries could trigger warning messages such as 'Nobody in the household is older than 15 years', 'The main activity of person is full-time student but person is not currently in school', or 'A house with a thatched roof is unlikely to have electricity, please verify'. If an unlikely entry was detected, the interviewer was obliged to verify with the respondent, and, if the unlikely entry turned out to be correct, to comment on the situation to reassure the analyst that the datum point was indeed correct.

Besides the system of 366 consistency checks, the full electronic questionnaire also included a report summarising the total calorific intake and its sources, as implied by the entries in the consumption section, allowing the interviewer to verify the plausibility of the consumption data. This consumption report was omitted in the restricted CAPI. It will be more completely discussed in section 2.5.

Finally, as respondents partake in resolving errors and inconsistencies, one could hypothesise that attitudinal factors, such as belief in the accuracy and usefulness of the survey, are affected by consistency checks.

2.2.3. Experiment 2: bundle of other CAPI features

Experiment 2 consisted of adding a further 320 PAPI questionnaires to the sample. Because of the random nature of the questionnaire allocation, any difference in data

quality between restricted CAPI and PAPI can only be due to the bundled effect of all CAPI features, excluding checks.

In line with most CAPI applications, we incorporated automated routing. The literature stresses automated routing as one of the most important error-reducing features of CAPI. For example, Banks and Laurie (2000) note that reducing errors related to complex routing in a 45-minute questionnaire was the main justification for migrating the British Panel Household Survey to CAPI in 1999. Automated routing avoids asking a question that should have been skipped (which may decrease the length of the interview), avoids asking irrelevant questions (which confuses respondents and may lower the regard they hold for the survey and its results) and decreases the time spent correcting data after the fieldwork. Automated routing also avoids the converse – skipping questions that should have been asked – and may therefore prevent dropping observations during analysis.

In this CAPI application, automated routing did not eliminate the need for routing checks. Unlike other existing CAPI surveys, our experiment displayed multiple questions and sections per screen and allowed the interviewer to continue the survey even if a required field/section was left blank. We made a conscious decision to set the program up in this way to allow the interviewer to return to a question later if, for example, the most knowledgeable person was not around. If an interviewer backtracked to change a response that determines subsequent routing, then an entry in a disabled field occurred. Again, we could have set it up so that the computer deletes entries in disabled fields automatically, but we were concerned that this could lead to unintended data loss, especially if gateway questions are accidentally changed after completing a section. The experiment allows us to disentangle the effects of checks from those of automated routing.

The data were stored in a relational database, using a record structure that eliminates redundancy. Key identifiers were used to link the various data tables in a manner that ensures the referential integrity of the complete dataset (this means, for example, that a household asset cannot exist without a related household, the identifier key being common to both data tables). Answers to most questions were selected from pre-coded drop-down menus or made use of radio-buttons. In some cases, drop-down menus were altered dynamically, depending on previous responses, so that the interviewer was never presented with an impossible response code. For example, when linking a woman to the ID of her husband the drop-down menu was restricted to married men within the household based on the previously entered marital status and sex variables.[8]

GPS coordinates and start and end times of the interview were captured automatically by the computer, eliminating any scope for interviewer error. In PAPI, the interviewer needed to copy the GPS coordinates from a GPS receiver and to record the start and end times of the interview in the appropriate fields.[9] Finally, PAPI had a data entry stage in which paper forms were rekeyed into the computer. There were numerous other smaller features that could all add up to a cleaner dataset. The experiment was not set up to isolate the effect of each of these features separately, so we can only identify them as a bundle of effects driving the difference between restricted CAPI and PAPI.

Just like the system of consistency checks, the bundle of other CAPI features may also contribute to the respondent's attitude towards the survey. For instance, noticing

[8] As pointed out by one of the referees for *Journal of Development Economics*, some of the checks could have been alternatively implemented by restricting answer options. The spouse drop-down and the item-specific unit list in the consumption section (described in section 5.1) are two examples of where we opted for this approach. In many places, however, we preferred checks because it could confuse an interviewer if he or she fails to locate an expected response option from the drop-down menu without any indication of which previous answer triggered its elimination from the list.

[9] Time data are notoriously difficult to collect in Tanzania, because Swahili time is counted differently; for example, 7 a.m. is considered to be the first hour of the day and is called '1 o'clock'. Time during the day is counted upwards from there until 6 p.m., which is called '12 o'clock' (the 12th hour of the day). After that, the first hour of the night is 7 p.m. and so forth. English and Swahili times are often mixed up in the same questionnaires.

that the interviewer is using a computer device instead of pen and paper may increase the respondent's perception of survey reliability.

2.2.4. Implications for sample bias and analysis

One likely consequence of the survey errors as described is that they generate missing variables and so reduce the effective sample size available for analysis. A questionnaire with missing or obviously erroneous data may lead the analyst to drop the observation entirely. If observations are randomly dropped, then one could simply increase the sample size of a PAPI survey to compensate. If, however, such mistakes are correlated with household characteristics of interest to the data user, then the analysis could be affected. We set up a formal test for this in section 2.4. Alternatively, an analyst may decide to make assumptions about the problematic observations in order to avoid dropping them from the sample. These assumptions may then introduce measurement error. Section 2.5 analyses the nature of that measurement error and its effects on analysis.

2.2.5. Interviewer effects

The quality of survey data depends to a large extent on both the technical capacity and the integrity of the interviewers. We expect education level and previous survey experience to improve the quality of survey data. In CAPI, the use of new survey technology might pose additional challenges to the interviewers, on the one hand. On the other hand, we expect some CAPI features, such as automatic routing and the elaborate system of validation checks, to assist the interviewers, possibly compensating for lower education and experience. In PAPI, it is likely that interviewers make fewer routing and consistency errors as the field work progresses because they receive feedback from their supervisors at the end of each survey day.

2.3. Descriptive statistics

Table 2.1 displays a set of summary statistics on survey and interviewer characteristics. Our measure for the number of routing errors is a simple count of the number of times an unconditional variable was missing or a conditional variable was mistakenly entered or missing (dependent on previous responses). It should be noted that a single error early on can sometimes have a cascading effect, creating a large number of routing errors throughout the questionnaire. Table 2.1 shows that PAPI contained an average of 10 routing errors per survey, restricted CAPI 0.6 and full CAPI 0.0 (out of a total of 152 fields; see section 2.2.2).

Whereas full CAPI contained an average of 0.0 impossible entries and 0.6 unlikely entries per questionnaire, these numbers were 0.2 and 1.1 for restricted CAPI and 0.5 and 1.4 for PAPI, respectively. Table 2.1 also shows differences in terms of the number of surveys with problematic time stamps and problematic GPS coordinates. A more extensive discussion, and an analysis of the statistical significance, of these differences among full CAPI, restricted CAPI and PAPI will be provided in section 2.4.

The share of questionnaires that have at least one impossible or missing entry potentially leading to missing values in our dataset amounts to 2%, 40% and 83% in, respectively, full CAPI, restricted CAPI and PAPI (not reported in tables). Table 2.11 in Appendix 2.A.2 lists the 15 most common missing values in any section in PAPI, excluding the consumption section (discussed separately below and in section 2.5).[10]

The most frequent errors are nonsensical survey durations, which occur in 24% of PAPI questionnaires but in virtually no CAPI questionnaires. The questionnaire was implemented in the context of a rural upgrade project and thus any questions on transport were especially important in the study.

[10] Note that we look at the 15 most common missing values in PAPI, as opposed to the 15 most common missing values over all three applications. The main purpose of Table 2.11 in Appendix 2.A.2 is to inform us of the types of errors made in PAPI, and not necessarily to compare the frequency of missing values across the three applications.

Table 2.1 Summary statistics on interviewer and survey characteristics

			Full CAPI (1)	Restricted CAPI (2)	PAPI (3)
Routing errors	Average number of routing errors per HH (total)		0.0 (0.1)	0.6 (1.1)	10.4 (11.9)
	Average number of entries in fields to be skipped per HH		0.0 (0.0)	0.1 (0.6)	6.3 (9.3)
	Average number of missing entries in required fields per HH		0.0 (0.1)	0.5 (0.9)	4.0 (5.6)
Impossible/ unlikely entries	Average number of impossible entries per HH		0.0 (0.0)	0.2 (0.4)	0.5 (1.1)
	Average number of unlikely entries per HH		0.6 (1.0)	1.1 (1.2)	1.4 (1.4)
Errors/unlikely entries per survey period quartile	Average number of routing errors + impossible entries + unlikely entries per survey period quartile (37 survey days in total)	1st	0.8 (1.1)	2.3 (2.2)	17.9 (16.6)
		2nd	0.8 (1.0)	2.1 (1.8)	12.2 (11.6)
		3rd	0.5 (0.9)	1.7 (1.6)	9.4 (10.2)
		4th	0.4 (0.7)	1.2 (1.2)	8.4 (7.4)
GPS data	% HHs >1 km from cluster centre (likely outliers given the small size of the EAs)		0.6	1.3	6.6
Time stamp data	% surveys with problematic time stamps		0.9	0.3	23.8
	% surveys conducted on day 1 of the cluster visit		45.6	50.0	49.1
	Average survey duration[a]		81 (24)	78 (23)	89 (25)
Interviewer characteristics	Average PAPI survey experience (months)		5.7 (8.9)	5.7 (8.9)	5.7 (8.9)
	Average CAPI survey experience (months)		7.4 (4.6)	7.3 (4.6)	7.4 (4.6)
	Average education (years)		13.3 (1.3)	13.6 (1.3)	13.4 1.3)
	Total number of interviewers		20	20	20
Sample size	Total number of clusters		80	80	80
	Total number of HHs per cluster		1200	320	320

Notes: Values in parentheses are standard deviations.
[a]For full CAPI, only the modules overlapping with the experiment were counted towards the survey duration (see section 2.2 for details).
CAPI, computer-assisted personal interviewing; PAPI, pen-and-paper interviewing; EA, enumeration area; GPS, global positioning system; HH, household.

Despite this, many PAPI questionnaires have problematic transport data. Table 2.11 shows that 9% of questionnaires miss the transport costs for at least one sold agricultural item, 7% miss data on the amount spent on transport to school for at least one household member, 6% on the one-way fare to school and 7% on the location at which crops sold fetched the highest price.

In practice, an analyst may assume that, by leaving the value blank, the interviewer may have wanted to indicate that these values were zero. Another analyst may decide that the interviewer made a mistake and place the value at the cluster or sample median. Neither will have much basis for that decision. Robustness analysis for these and hundreds of similar data-cleaning decisions that need to be made in a typical dataset is unlikely to be feasible. Assuming that a purist will drop any household that has any of the four transport-related questions missing would imply dropping 20% of observations. The other potentially missing variables listed in Table 2.11 in Appendix 2.A.2 occur in core variables, which are key to calculating statistics such as fertility rates, literacy rates, the number of people living with a disability and the number of landless households.

Table 2.12 in Appendix 2.A.2 lists the 10 most common consumption-related (potentially) missing values. In terms of the share of questionnaires in which the error occurs at least once, the most common consumption-related error concerns food items for which the three sources ('purchases', 'home production' and 'gifts') do not sum to the indicated total. This error occurred at least once in more than 17% of all PAPI surveys. In comparison, this error occurred in only 3% of restricted CAPI households and in almost none of the full CAPI households.[11] In terms of the average frequency per

[11] The most likely reason why this error did not occur as frequently in restricted CAPI as in PAPI is that CAPI displayed the total amount consumed coming from the three different sources on the screen, allowing the interviewer to check. If, despite this, the sum was still not correct, then a consistency check in full CAPI warned the interviewer of the mistake.

questionnaire, the top error concerns the question 'In the past 7 days did this household consume any [food item]?', which was missing for four food items on average (out of a total of 53 items per survey) in approximately 6% of all PAPI surveys, once on average in approximately 9% of all restricted CAPI surveys, and was not missing at all in full CAPI surveys. In section 2.5, we will determine whether these inconsistencies could lead to different analytical conclusions.

2.4. Errors and sample size reduction

2.4.1. Methods of analysis

To investigate the effects of CAPI on errors and potential sample size reduction more formally, we start by estimating Y_{ijc}, which is a count of the number of problematic variables (some of which may potentially have to be dropped in analysis) in the questionnaire of household i, interviewed by interviewer j in community c:

$$Y_{ijc} = \alpha + \beta C_{ijc} + \gamma V_{ijc} + \varepsilon_{ijc} \qquad (2.1)$$

where C_{ijc} indicates a CAPI questionnaire (a dummy equal to 1 for both full and restricted CAPI and to 0 for PAPI) and V_{ijc} is a dummy set to 1 if the interviewer had access to validation checks during the interview, which was the case in full CAPI, but not in restricted CAPI and PAPI. In equation (2.1), γ measures the effect of the validation checks on the dependent variable, whereas β is an estimate of the bundled effect of all other CAPI features that could influence the number of errors in a questionnaire.

If error counts depend on household characteristics of interest to the data user, then the dropping of observations with erroneous variables could introduce sample bias.[12] To investigate this, we check whether the number of problematic values in a questionnaire depends on a set of household characteristics X_{ijc} and whether CAPI can correct for this.

[12] Observations may not need to be dropped if cleaning assumptions are made. This may introduce measurement error, the nature of which is the subject of section 5.

We are particularly interested in the level effect of X_{ijc} as well as its interactions with C_{ijc} and V_{ijc}:

$$Y_{ijc} = \alpha + \beta C_{ijc} + \gamma V_{ijc} + \delta X_{ijc} + \phi X_{ijc}.C_{ijc} + \rho X_{ijc}.V_{ijc} + \varepsilon_{ijc} \qquad (2.2)$$

where Y_{ijc} is a count of the number of variables that potentially have to be dropped or cleaned in household *i*.

In a final specification we estimate the interaction effects of C_{ijc} and V_{ijc} with the interviewer's characteristics, such as months of experience with CAPI, months of experience with PAPI and years of education. This allows us to verify whether the measured effects differ by experience in either type of questionnaire as well as by education level. We will also look at interactions with the survey period, to determine whether error rates drop over time as a result of interviewer learning effects.

Although the set-up ensured that questionnaires were equally and randomly spread over interviewers, clusters and time, we also verified that all results were robust to the controls for additional factors that may influence the number of errors in an interview: characteristics of the respondent (age, sex, literacy, whether a head of household), characteristics of the interview (conducted on day 1 or 2 of the team's visit), the interviewer and the location. The last two effects are included as cluster and interviewer fixed effects. We find that all estimations are robust to these further controls, so will not report this further.

2.4.2. Routing errors

Columns 1 and 2 in panel A of Table 2.2 show that restricted CAPI significantly reduces the total number of routing errors by almost 10 per questionnaire compared with PAPI. Column 1 shows that there are, on average, four missing entries in required fields (the constant in the regression without controls), out of which 3.5 are eliminated through CAPI. The remaining 0.5 errors are eliminated by adding checks to CAPI.

Table 2.2 Effect of CAPI and checks on data quality

Panel A				
	Routing errors: missing entries in required fields	Routing errors: entries in fields that should have been skipped	Impossible entries	Unlikely entries
	(OLS)	(OLS)	(OLS)	(OLS)
	(1)	(2)	(3)	(4)
CAPI	–3.544***	–6.225***	–0.337***	–0.259***
	(0.187)	(0.306)	(0.040)	(0.087)
Checks	–0.490***	–0.117	–0.148***	–0.461***
	(0.149)	(0.244)	(0.032)	(0.069)
_cons	4.038***	6.344***	0.487***	1.347***
	(0.132)	(0.216)	(0.028)	(0.061)
Panel B				
	Potentially missing values in non-consumption sections	Potentially missing values in consumption sections	Time stamp problems	GPS problems
	(OLS)	(OLS)	(LPM)	(LPM)
	(1)	(2)	(3)	(4)
CAPI	–2.800***	–1.081***	–0.234***	–0.053***
	(0.142)	(0.121)	(0.015)	(0.010)
Checks	–0.287**	–0.351***	0.006	–0.007
	(0.113)	(0.096)	(0.012)	(0.008)
_cons	3.091***	1.434***	0.238***	0.066***
	(0.101)	(0.085)	(0.011)	(0.007)

Notes: $N = 1840$. Standard errors are shown in parentheses. All estimates are robust to cluster and interviewer fixed effects, controlling for respondent characteristics (age, head of household, sex, literacy), household characteristics (number of household members) and interview characteristics (day of team's visit). In 16 pen-and-paper interview households at least one of these variables was missing, in which case we included a dummy indicating this. Potentially missing values are measured as the sum of missing and impossible entries. The dependent variables in columns (3) and (4) in panel B are dummies = 1 if survey had problematic time stamp or GPS coordinates, respectively.
GPS, global positioning system; LPM, linear probability model; OLS, ordinary least squares.
***$p<0.01$, **$p<0.05$, *$p<0.1$.

All of the 6.3 entries made in fields that ought to have been skipped, on average in PAPI, are eliminated by CAPI, with no additional effect of the checks. The latter type of error is perhaps less problematic than the former one, but such ambiguity in the data is nevertheless best avoided and will, in any case, add time to the interview (see below).

Taken together, this shows that 94% of routing errors are avoided through the automated routing system and that the checks eliminate almost all those that remain. Appendix 2.A.3 shows that this does not lead to respondents reporting a smoother

survey experience. It is unlikely that this result stems from interviewers leaving 'Don't know' responses blank. First, there were specific codes for such a response and the interviewers were trained extensively on this matter. Second, a comparison of the occurrence of 'Don't know' answers across the three different experiments does not show any significant differences. CAPI lends itself to the use of unfolding brackets to reduce 'Don't know' answers, but this specific experiment did not make use of them.

2.4.3. Unlikely and impossible entries

Column 3 in panel A of Table 2.2 shows that restricted CAPI reduces the number of impossible entries by 0.34 per questionnaire compared with PAPI, and adding checks further reduces this number by 0.15 to almost zero. This means that, in a dataset of 1200 households, moving from PAPI to full CAPI would reduce the number of impossible entries by 588 in total. The bundled effect of 'all other CAPI features' on the occurrence of impossible entries, as discussed in section 2.2, seems to be larger than that of the checks.

Column 4 in panel A of Table 2.2 shows that CAPI significantly reduces the number of unlikely entries by 0.26 per household survey. This effect is even greater when checks are available, with the number of unlikely entries falling from 1.35 in PAPI to 0.63 in full CAPI. This result suggests that, although some unlikely entries remain (once confirmed to be correct by the interviewer), full CAPI successfully assists the interviewer in detecting unusual entries that turn out to be incorrect after confirmation. Furthermore, because the program flags these entries and reminds the interviewer to comment, the analyst is reassured that the datum point is indeed correct.

Appendix 2.A.3 further shows that the techniques introduced by CAPI to avoid these errors do not increase the credibility or usefulness of the results in the eyes of the respondent.

An unintended natural experiment occurred within the experiment. It was realised, during analysis for this chapter, that 13 validation rules had been erroneously omitted from the program. Tabulating the number of times each of these malfunctioning checks was violated in the resulting dataset, we find no significant differences across the three types of questionnaires. This suggests that CAPI is only as good as the features that are built into it. Without checks or other error-reducing features, CAPI has no impact on impossible entries.

Column 3 in panel B in Table 2.2 shows that there are 24% of households that had problematic interview duration calculations in PAPI, but CAPI reduces this to virtually 0. Panel B also shows that PAPI has 6.6% problematic GPS locations, which are largely eliminated through CAPI's automatic GPS capture. EAs were very small in Pemba, and we can be confident that any location further than 1 km away from the cluster centre is problematic. One may argue that any analysis requiring the use of time stamps or GPS locations should simply increase its sample size to account for this. But, as will be shown next, these missing observations occur more frequently in certain types of households.

2.4.4. Implications for sample size reduction and sample bias

A missing or an impossible entry may cause an observation to be dropped, which may lead to biased estimates if the missing values are non-randomly distributed across the sample. To investigate this, Table 2.3 shows estimates of equation (2.2) for four different left-hand-side variables (the uninteracted results for these four regressions are shown in panel B in Table 2.2). The dependent variables in columns 1 and 2 are simple sums of the number of missing entries in the required fields and the number of impossible entries. This sum is first made for entries in any part of the questionnaire, excluding the consumption section (column 1) and then separately for entries in the consumption section (column 2).

Table 2.3 Interaction with household characteristics

	Missing/ impossible entries in non-consumption section (OLS)	Missing/ impossible entries in consumption section (OLS)	Time stamp problems (LPM)	GPS problems (LPM)
	(1)	(2)	(3)	(4)
CAPI	0.197	–1.369***	-0.163***	0.013
	(0.434)	(0.379)	(0.049)	(0.032)
Checks	–0.203	–0.446	0.005	–0.022
	(0.333)	(0.291)	(0.037)	(0.025)
Dummy = 1 if HH head is female	1.367***	–0.044	–0.010	0.003
	(0.271)	(0.237)	(0.030)	(0.020)
HH size	0.421***	–0.084**	0.014***	0.008***
	(0.040)	(0.035)	(0.005)	(0.003)
Dummy = 1 if head *not* employed in agriculture	1.149***	–0.408**	0.013	0.015
	(0.209)	(0.183)	(0.023)	(0.015)
Dummy = 1 if HH belongs to richest 25th percentile	–0.049	0.701***	–0.034	0.014
	(0.212)	(0.186)	(0.024)	(0.016)
CAPI interacted with:				
Dummy = 1 if HH head is female	–1.322***	–0.084	0.006	–0.020
	(0.370)	(0.323)	(0.041)	(0.027)
HH size	–0.410***	0.079	–0.014**	–0.010**
	(0.055)	(0.048)	(0.006)	(0.004)
Dummy = 1 if head *not* employed in agriculture	–1.069***	0.456*	–0.019	–0.007
	(0.297)	(0.259)	(0.033)	(0.022)
Dummy = 1 if HH belongs to richest 25th percentile	0.027	–0.882***	0.046	–0.005
	(0.310)	(0.271)	(0.035)	(0.023)
Checks interacted with:				
Dummy = 1 if HH head is female	–0.047	0.128	–0.002	0.027
	(0.284)	(0.249)	(0.032)	(0.021)
HH size	–0.009	0.005	0.001	0.002
	(0.043)	(0.037)	(0.005)	(0.003)
Dummy = 1 if head *not* employed in agriculture	–0.080	–0.044	0.004	–0.006
	(0.235)	(0.206)	(0.026)	(0.017)
Dummy = 1 if HH belongs to richest 25th percentile	0.030	0.184	–0.008	–0.006
	(0.260)	(0.227)	(0.029)	(0.019)
_cons	0.002	1.813***	0.166***	0.010
	(0.320)	(0.280)	(0.036)	(0.024)
N	1840	1840	1840	1840

Notes: Standard errors are shown in parentheses. All estimates are robust to cluster fixed effects, controlling for interview characteristics (interviewer ID, day of team's visit).
CAPI, computer-assisted personal interviewing; GPS, global positioning system; HH, household; OLS, ordinary least squares; LPM, linear probability model.
***p<0.01, **p<0.05, *p<0.1.

Both of these can lead to either dropping the observation in question or making an ad hoc data-cleaning decision about what is happening. The third dependent variable indicates whether or not there was a problem with the time stamps, and the fourth whether or not there was a problem with the GPS coordinates. We do not use the information on entries in disabled fields or unlikely observations as these two types of errors would probably not lead to dropping the observation. Unlikely observations may introduce error and affect the analysis, but that will be the subject of section 2.5.

Table 2.3 shows that the sum of the number of missing entries in required fields and the number of impossible entries (as picked up by the validation rules) are dependent on household characteristics. Column 1 shows how large, female-headed and non-farm households are more likely to have non-consumption-related entries that could cause the observation to be dropped or the entry to be altered by the analyst.

Column 2 shows a different pattern when focusing on the consumption section. We see here that rich households make more errors, possibly because of their more complicated consumption patterns. As expected, farming households have more problematic consumption data, as a larger share needs to be estimated from home production, often using subjective units of measurement. The coefficient of household size is now significantly negative. The effect is not large – increasing household size with five members would reduce the number of consumption errors in a questionnaire by an average of 0.4 – but is still significant and could be explained by the fact that larger households (more than nine members) have an average of only 1.8 more consumed items than small households (one to three members), whereas smaller households are 40% more likely to use decimals in their quantity estimation, which generally are more prone to erroneous entries. Furthermore, although there is no difference in the types of consumption items consumed by smaller households, the sources from where they obtain them are different: small households have more

consumption from gifts and may therefore be less familiar with objective units of measurement found in the market place.

The coefficient of female-headed household is no longer significant. Importantly, once the interaction with CAPI and any of the characteristics discussed above is made, the effects disappear: the sum of the level and interaction effects is never statistically significant (verified by the authors). Interactions with age and education of head were not found to be significant.

Interestingly, we find that even problems with time stamps and GPS locations are not independent of household characteristics. In particular, they occur more frequently in large households. This could be because large households have a much longer interview time, as the questionnaire contains many roster questions that are repeated for each member. Median interviewing times on paper are 53, 82 and 113 minutes for a one-, three- and eight-person household, respectively. It took 141 minutes to interview the one 13-member household in the survey. This increase in interviewer workload may reduce concentration when copying time or GPS coordinates.[13] We confirmed by a formal statistical test that CAPI undoes the negative effect of household size on problematic GPS and interview duration measurements.

2.4.5. Interaction with interviewer characteristics and survey period

Table 2.4 shows the interaction effects of our main variables of interest (CAPI and checks) with the total number of years of formal schooling and the number of months of CAPI and PAPI survey experience of the interviewer.

[13] Lengthy questionnaires have more non-consumption errors in general. This is confirmed by the results of a regression of the number of non-consumption-related missing/impossible entries on the duration of interview (not reported), which shows a significantly positive correlation between the two. Interestingly, the interview length does not influence the number of missing/impossible entries in the consumption section.

Table 2.4 Interaction with interviewer characteristics

	Total number of alerts	Total number of alerts
CAPI	–10.366***	–30.967***
	(0.424)	(4.865)
Checks	–1.217***	–5.048
	(0.337)	(3.872)
PAPI experience (months)		–0.315***
		(0.037)
CAPI experience (months)		0.199***
		(0.065)
Education (years)		–1.833***
		(0.252)
CAPI interacted with:		
PAPI experience (months)		0.284***
		(0.052)
CAPI experience (months)		–0.188**
		(0.092)
Education (years)		1.526***
		(0.356)
Checks interacted with:		
PAPI experience (months)		0.027
		(0.042)
CAPI experience (months)		0.003
		(0.073)
Education (years)		0.274
		(0.283)
_cons	12.216***	37.015***
	(0.300)	(3.444)
N	1840	1840

Notes: Standard errors are shown in parentheses. All estimates are robust to cluster fixed effects, controlling for respondent characteristics (age, head of household, sex, literacy), household characteristics (number of household members) and interview characteristics (day of team's visit). Total number of alerts is the sum of routing errors and impossible and unlikely entries.
CAPI, computer-assisted personal interviewing; PAPI, pen-and-paper interviewing.
***$p<0.01$, **$p<0.05$, *$p<0.1$.

We find that both PAPI survey experience and education significantly reduce the total number of alerts (routing + impossible + unlikely entries) in PAPI surveys. Interestingly, the number of years of CAPI survey experience seems to significantly increase the number of errors on paper, for a given number of years of PAPI experience. This suggests some unlearning of best practice PAPI skills once interviewers switch from paper to the computer. Both experience and education effects disappear once CAPI is used (confirmed by formal statistical tests). Banks and Laurie (2000) noted

Table 2.5 Interaction with survey period

	Total number of alerts	Total number of alerts
CAPI	–10.366***	–15.657***
	(0.424)	(0.782)
Checks	–1.217***	–1.528**
	(0.337)	(0.616)
Survey days 10–18		–5.723***
		(0.784)
Survey days 19–27		–8.509***
		(0.817)
Survey days 28–37		–9.522***
		(0.817)
CAPI interacted with:		
Survey days 10–18		5.568***
		(1.107)
Survey days 19–27		7.878***
		(1.149)
Survey days 28–37		8.433***
		(1.156)
Checks interacted with:		
Survey days 10–18		0.156
		(0.879)
Survey days 19–27		0.418
		(0.908)
Survey days 28–37		0.756
		(0.922)
_cons	12.216***	17.941***
	(0.300)	(0.558)
N	1840	1840

Notes: Standard errors are shown in parentheses. All estimates are robust to cluster and interviewer fixed effects, controlling for respondent characteristics (age, head of household, sex, literacy) and household characteristics (number of household members) and interview characteristics (interviewer ID). Total number of alerts is the sum of routing errors and impossible and unlikely entries.
CAPI, computer-assisted personal interviewing. ***p<0.01, **p<0.05, *p<0.1.

how PAPI interviewers can be easily retrained to conduct CAPI. This result suggests that CAPI can, to some extent, compensate for lower education and experience level of interviewers, mainly as a result of automated routing. The interaction effects of checks with education and experience are not significant.

Table 2.5 provides data on whether error rates drop as the survey progresses over time and, if so, whether the pattern is different for CAPI and PAPI. To do this, we split the 37 survey days into quartiles and include dummies for each period, both as levels

and as interactions with the CAPI and checks dummies. The results suggest that error rates do indeed drop for PAPI, but not for CAPI. Compared with the first quartile, the total number of alerts is significantly lower in subsequent survey quartiles for PAPI, with almost 10 alerts fewer per PAPI household survey in the last 9 days of the survey. Once the interaction effects with CAPI are added to the level coefficients, the effect of the quartile disappears (confirmed by formal statistical tests). In other words, there is no similar learning effect for CAPI. One reason for this could be that the average number of alerts in the first quartile of CAPI survey work is very low (0.8) compared with PAPI (18) and therefore there is much less scope for improvements under CAPI than under PAPI (Table 2.1). Interactions with checks are insignificant. Taken together, the results from Tables 2.4 and 2.5 indicate that CAPI is less dependent on interviewers' experience, education and learning over time.

2.5. Measuring consumption

2.5.1. Food consumption

Estimates of poverty and welfare in developing countries are frequently calculated using a consumption recall module in a household questionnaire. Although the largest share of consumption is related to food, it is exactly food consumption that is most problematic to measure accurately. The typical food recall module will have the interviewer go over a list of food items in two iterations. In the first iteration, each item consumed by the household over the recall period is flagged. A second iteration then goes through the list of flagged items and, for each, asks total household consumption and its decomposition into sources (purchases, home production and other sources).

Three important problems arise. First, quantities are expressed in imprecise units; households report consumption as 'pieces' of cassava, 'bundles' of spinach or 'bunches' of bananas (Capéau and Dercon, 2006). This leads to ambiguous item–unit combinations. Whereas the size of such units is subject to interpretation (large versus

small), the analyst needs a clear mapping onto metric units. Second, the list of units is uniform for each consumption item, even though some units in the list do not apply to all items. This causes conflicting item–unit combinations (e.g. 'litre' for 'potatoes'), usually detected only much later during data analysis. Third, a completed consumption module represents a rather unwieldy matrix, making it hard for an interviewer to maintain an overview of the consumption pattern of the household. Therefore, obvious errors and irregularities in the reported consumption are only highlighted several months later when researchers start analysing the data. At that point, the only solution is to either make an ad hoc assumption about what is meant or omit the observation from the sample.

In CAPI, the screen of the handheld device can be used to display pictures of vague units, such as 'bundle' or 'bunch', so that they can be more precisely mapped onto metric units. One example of such a picture is shown in Figure 2.1 and more examples are shown in Appendix 2.A.4.

Figure 2.1 Picture of pieces of cassava (used in CAPI consumption section)



The CAPI survey application can also tailor the list of units to be specific to the item, making it impossible to, for instance, express potato consumption in litres or cooking oil in bags. Finally, by mapping each item–unit combination to its calorific value, the computer can summarise, in a report, the calorific intake pattern of the household. Figure 2.2 shows an example of such a report.

Figure 2.2 Example of a food consumption data summary report

Food energy consumption data

540201 Bahareni

540201/0101 [redacted]

Item		Amount consumed over reporting period		Kcals/day	
0201	Cassava fresh	75	Small piece	4,457	38%
0102	Rice (husked)	4	Kilograms	2,040	17%
0106	Wheat flour	2	Kilograms	1,040	9%
0301	Sugar	1.8	Kilograms	964	8%
0504	Coconuts (mature)	7	Piece/Number	896	8%
0814	Fresh fish (excl. dagaa)	4	Medium fish	644	5%
0105	Maize (flour) Sembe	1	Kilograms	520	4%
0807	Chicken, other poultry and wild birds (incl. pigeons, kwarara, sabile)	2	Piece/Number	420	4%
0701	Ripe bananas	3	Hand	253	2%
0401	Dried beans	0.5	Kilograms	229	2%
0503	Coconuts (immature)	2	Piece/Number	84	1%
0110	Bofio	2	Piece/Number	77	1%
0811	Eggs	6	Piece/Number	68	1%
0818	Fresh dagaa	1	1/4 KG Tin	16	0%
0601	Tomatoes	6	Small piece	11	0%
0602	Onions	1	Small piece	1	0%
			Total Kcalories per day	**11,720**	**99%**
			Total number of Adult Equivalent Units (AEU) in the household	**4.72**	
			Total Kcalories consumed per AEU per day	**2,506**	

This feature allows the interviewer to carry out a report-based check during the interview to determine whether the total kilocalories (kcal) per adult equivalent unit (AEU) lies within reasonable boundaries and whether the sources of calories are sensible given the context in which the interviewer is conducting the work. We refer to this report as the ‘consumption report’ in what follows.

Additionally, the automated routing and the consistency checks, discussed in detail in section 2.2, are expected to improve data quality. Some of these features could, in principle, be implemented in paper questionnaires, although the logistics are more

complicated here. This is especially true for the automated routing, the consistency checks and the consumption report, which rely on complex matrix manipulations and look-up tables. In this experiment, full CAPI had all three features, restricted CAPI omitted the checks (which we mean to include both the validation checks and the consumption report on total food energy intake and its sources), whereas PAPI also omitted pictures and item-specific units (e.g. we would just have 'bunch', or 'litres', as a possible unit code for reporting banana consumption).

Table 2.6 shows that, in PAPI, in 1% of cases the item–unit combination made no sense; in these cases, the calorie value was replaced by the median EA-level value in the subsequent analysis. A further 42% of the item–unit combinations in PAPI were ambiguous (pieces, bunches, bundles, heaps, etc.) and, in order to obtain a precise conversion to kcal values, an assumption about the exact size of the ambiguous unit needed to be made. We used lower and upper bound estimates of the unit conversion rates, as well as a mid-range value (a typical user would have probably used this mid-range estimate). Although upper and lower bound conversion rates were quite reasonably set,[14] Table 2.6 shows that changing the assumptions on unit conversions from lower to upper bound estimates raises calorific intake per AEU per day from 2478 to 4362. There is also a substantial increase in the standard deviation as one goes from full CAPI (655) to restricted CAPI (1177) and PAPI (1644–3379). The number of outlier observations, with values over 4000 kcal per adult per day, is 1% for full CAPI, 8% for restricted CAPI and 7%, 20% and 35% for PAPI, depending on the conversion assumption.

[14] In many cases the units could be matched to the CAPI pictures and the lower bound was taken as the size of the smallest depicted unit and the upper bound as the size of the largest unit.

Table 2.6 Summary statistics on calorific intake

		Full CAPI	Restricted CAPI	PAPI		
Item–unit combinations in food consumption section (%)	No problem	98.4	98.0	56.8		
	Ambiguous	1.6	2.0	42.0		
	Non-valid	0.0	0.0	1.2		
Unit size assumption		N/A	N/A	Small	Medium	Large
Calorific intake per AEU per day (kcal)[a]	Mean	2297 (655)	2471 (1177)	2478 (1644)	3145 (2038)	4362 (3379)
	Minimum	413	235	232	232	232
	Maximum	5117	10,528	18,969	20,181	23,057
Outliers in the distribution of calorific intake (kcal) per AEU per day (%)	<1000	1.8	3.4	3.8	1.9	1.3
	>4000	0.8	7.8	7.2	19.7	35.4
Malnutrition: % of households under threshold of calorific intake (kcal) per AEU per day[b]	<1500	8.3***	14.4**	20.6		
	<1800	23.1**	29.4	29.4		
	<2200	47.5	50.3	51.9		
N		1200	320	318	318	318

Notes: Standard errors are shown in parentheses. Calorific intake values include meals taken outside of the household by household members. AEU positively adjusted for number of meals taken by guests in the household.

[a]Two PAPI observations with calorific intake values based on a small unit assumption >23,570 kcal per AEU per day were excluded from the analysis to avoid these outliers driving the results.

[b]Asterisks indicate *p*-values for a one-sided *t*-test whether the value is significantly different from PAPI.

CAPI, computer-assisted personal interviewing; PAPI, pen-and-paper interviewing; Full CAPI, CAPI with checks; Restricted CAPI, CAPI with checks disabled; N/A, not applicable.

***$p<0.01$, **$p<0.05$, *$p<0.1$.

These results indicate that the effect of the 'other CAPI features', probably pictures and item-specific units in this case, depends on the difference between the ad hoc assumptions on unit size in PAPI and reality, whereas the effect of the checks is independent of this. In fact, we do see that in CAPI the pictures of the smaller units were 14 times more likely to be chosen than those of large units and nearly 2.5 times more likely than mid-range units. Equipped with this knowledge we can adapt the unit conversions, but it is fair to say that most similar surveys would base their unit conversions on much thinner data. Because we know that the small unit assumptions are

closest to the truth, we will use these in the remainder of the text. In this way we expect any differences between PAPI and CAPI to be lower bound estimates. Finally, there are a number of other small data-cleaning decisions that needed to be made with regard to all the violated consistency checks.[15]

Would our assessment of the food situation have changed, depending on whether we did a survey on paper or electronically? The answer depends on the calorific intake threshold we consider when defining malnutrition. Had we done the survey on paper, we would have concluded that 21% of households live on less than 1500 kcal/AEU/day. Had the same survey been conducted in full CAPI then the conclusion would have been that 8% of households live below this threshold. This difference is statistically significant at well below 0.01%. Table 2.6 further shows that restricted CAPI puts the same figure at 14%, implying that, on average, 6 percentage points (pp) of the difference between full CAPI and PAPI is the result of checks, whereas 7 pp are the result of other CAPI features. Raising the threshold to 1800 kcal/AEU/day still shows a significant difference between full CAPI and PAPI (at the 0.05% significance level), but the effect is completely due to the checks and not to the bundle of other CAPI features. When we consider 2200 kcal/AEU/day as the malnutrition threshold, we do not find any significant differences between CAPI and PAPI.

2.5.2. Total consumption, poverty and inequality

To arrive at a consumption measure, we place a monetary (Tanzanian shilling (TZS)) value on food consumption. For purchased items, this comes directly from the respondent's assessment of the value, whereas for gifts and home-produced items unit prices are used to convert the quantity estimations into monetary values. To this food consumption is added a non-food consumption component, which was asked directly in

[15] For example, we assumed that a missing source entry indicated zero consumption for that particular source, as discussed in section 2.4.4.

Table 2.7 Consumption, poverty and inequality

	Full CAPI (1)	Restricted CAPI (2)	PAPI (3)	Significance of *t*-test (1) = (2)	(2) = (3)	(1) = (3)
Total consumption per AEU mean and (standard deviation)	435,251 (222,881)	475,889 (279,349)	546,223 (337,538)	***	***	***
Poverty headcount (%)	83.0	74.6	68.3	***	*	***
Gini (95% CI)	0.24 (0.22–0.25)	0.29 (0.26–0.32)	0.30 (0.27–0.32)	**	=	***
Income elasticity of food consumption	0.89 (0.01)	0.95 (0.02)	0.94 (0.02)	*	=	***
N	1200	319	319			

Notes: Standard errors are shown in parentheses. Standard errors of the Gini coefficients were calculated using the jacknife method. Calorific intake values include meals taken outside of the household by household members. AEU positively adjusted for number of meals taken by guests in the household. Two PAPI outlier observations with total consumption over 3,000,000 are dropped. Asterisks indicate *p*-values for a one-sided *t*-test whether value is significantly different from PAPI. PAPI food consumption is calculated using the small unit assumption. Income elasticities are those of the interaction term between log total consumption and a dummy indicating which sample the observations is from in a regression that includes only the two samples being tested.
CAPI, computer-assisted personal interviewing; Full CAPI, CAPI with checks; Restricted CAPI, CAPI with checks disabled; AEU, adult equivalent unit; CI, confidence interval; N/A, not applicable; PAPI, pen-and-paper interviewing.
***$p<0.01$, **$p<0.05$, *$p<0.1$, =$p\geq0.1$.

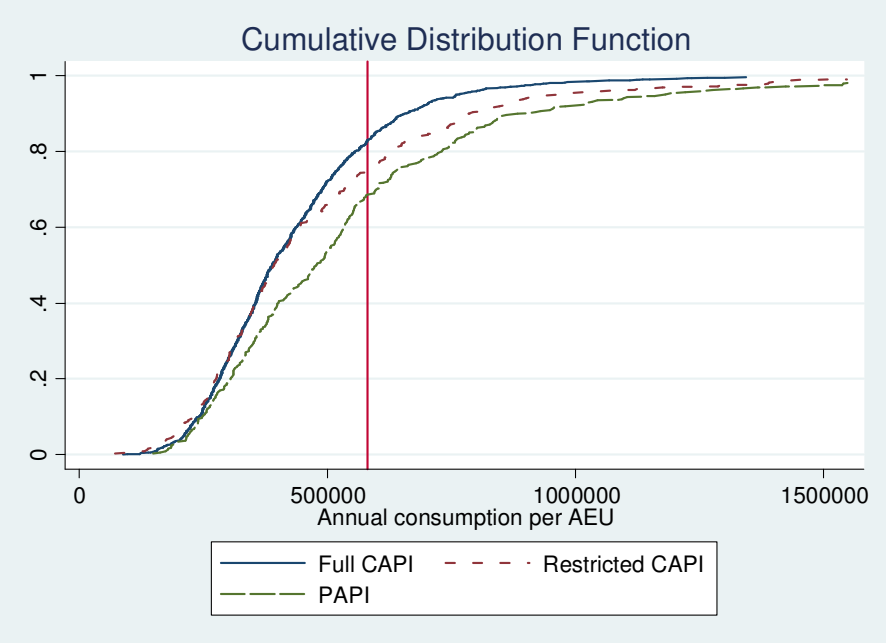
monetary value both on paper and in CAPI. As Pemba is a very small island and the survey was concentrated in the northern half only, we do not correct for prices. The regression analysis, however, will always verify robustness to inclusion of cluster fixed effects.

Table 2.7 shows that average consumption increases by 9% as one moves from full CAPI to restricted CAPI and by another 15% when moving to PAPI, creating a jump of 25% from full CAPI to PAPI. These mean differences also translate into very different conclusions regarding the number of people who live below the basic need poverty line, with the poverty headcount going down from 83% in full CAPI to 68% in PAPI.[16] Note

[16] Our poverty line is set at a value of TZS 580,832 of annual consumption per AEU. To construct this poverty line, we started from the basic need poverty lines of Wete and Micheweni (where this experiment was conducted) as defined by the Zanzibar Household Budget Survey (HBS) 2005. We then adjusted this

that the 2005 Zanzibar Budget Household Survey (Office of Chief Government Statistician Zanzibar, 2006) (conducted on paper) reported a poverty headcount of 72.54% on average for the region we consider (Wete and Micheweni districts). Because the differences both between full and restricted CAPI as well as between restricted CAPI and PAPI are significant, we conclude that both the checks and the bundle of other CAPI features are important. Interestingly, the cumulative distribution functions shown in Figure 2.3 show that the effect of the checks depends on where one draws the poverty line. For poor households, there is no effect of the checks, whereas the effects start appearing from around TZS 400,000 onwards. This evidence is consistent with the fact that rich households have more complicated consumption patterns, for which the power of the computer is important to summarise the information in an intelligible way.

Figure 2.3 CDF of annual consumption per adult equivalent unit (AEU)



Notes: 15 values over 1,600,000 are not displayed on the graph for a better presentation. The vertical line presents the poverty line. CAPI, computer-assisted personal interviewing; PAPI, pen-and-paper interviewing.

poverty line for inflation and differences in survey methods by multiplying it by a factor reflecting the difference in median consumption between the 2005 HBS dataset and our own 2009 dataset.

Next, we investigate what this means for inequality. Consistent with expectations, we see that full CAPI lowers the Gini coefficient to 0.24 from 0.30 in PAPI (pb0.01), a difference almost entirely attributable to the checks and not to the bundle of other CAPI features. At least some of the variation picked up in restricted CAPI and PAPI is actually measurement error rather than real inequality. Finally, we estimate Engel curves (log food consumption regressed on a constant and the log of total consumption) within each of the three different datasets and show, in Table 2.7, what the implied differences are with respect to the calculation of the income elasticity of food consumption. We see, as with the Gini coefficient, pronounced differences when moving from full CAPI to restricted CAPI, but not from restricted CAPI to PAPI. We conclude that it is mainly the checks that explain the difference between paper and electronic questionnaires with respect to Gini coefficients and income elasticities.

2.5.3. Testing for classical measurement error in consumption

The previous results demonstrate a tendency for the mean and spread of total consumption to go up on paper, compared with full CAPI, translating into lower measured poverty and higher measured inequality in PAPI. We also noted that this can lead to significantly different coefficients on the total consumption variable when estimating Engel curves. From section 2.4 we know that simple error counts in consumption data depend on wealth, household size and the occupation of the head.

In this section, we further analyse the nature of the measurement error (after cleaning, see footnote 15). We closely follow Bound et al. (2001) by considering a true linear model $y^* = X^* \beta + \varepsilon$. Rather than observing the true scalar y^* and the true vector X^*, we observe $X = X^* + \rho$ and $y = y^* + \varphi$. The measurement error is called classical if ρ and φ are uncorrelated to X^*, y^* or ε, that is, when the error is uncorrelated with the true value of the measured variable, the true values of other variables in the model and

the stochastic disturbance term. Lacking validation data, that is, information on the true values X^* and y^*, it is difficult to test these assumptions underlying the classical measurement error model directly. However, we can test whether our data are consistent with standard classical measurement error results: classical measurement error causes no bias when only the dependent variable has error and attenuates the estimated coefficient on a single error-ridden explanatory variable. The remainder of this section uses that insight to set up two tests for classical measurement error.

In the first test, we exploit the insight that an explanatory variable, measured with classical error, will lead to attenuation bias. In order to test the hypothesis of zero attenuation bias, we estimate the following estimation equation for three different outcomes O_i that are expected to be affected by consumption:

$$O_i = \alpha + \beta C_i + \gamma V_i + \delta LnCons_i + \phi LnCons_i.C_i + \rho LnCons_i.V_i + \varepsilon_i \qquad (2.3)$$

where C_i and V_i are defined as before and $LnCons_i$ is the log of total consumption per AEU of household i. Finding $|\delta| < |\delta + \phi|$ and/or $|\delta| < |\delta + \phi + \rho|$ would be consistent with attenuation bias with PAPI, assuming that the CAPI data are correct.

Table 2.8 shows the results of three regressions. In the first, i is a child between 7 and 14 years old and O_i is the number of years of formal schooling completed by the child, controlling for age fixed effects to compare only children of the same age. In the second specification, i is a school-going child in the sample and O_i is the amount spent by the household on his or her education (in which case we drop the education expenditure components from the consumption aggregate to avoid spurious correlation). The third is a regression where O_i is a dummy for whether the child (0–14 years old) slept under a treated mosquito net the night prior to the survey, again controlling for age fixed effects. In all cases we would expect family wealth and its correlates, measured

Table 2.8 Implications of measurement error in the explanatory variable

	No. of years of schooling (children aged 7–14 years)	Schooling expenditures on school-going children[a]	Child slept under a treated bednet night before survey (children aged 0–14 years)
	(1)	(2)	(3)
	Age FE	Age FE	LPM age FE
CAPI (β)	–0.451	39,915	–0.035
	(2.316)	(28,616)	(0.629)
Checks (γ)	–3.808**	–82,896***	–1.963***
	(1.842)	(22,376)	(0.509)
Log total consumption per AEU (δ)	0.204	8000***	0.134***
	(0.135)	(1676)	(0.036)
Interaction of log total consumption per AEU with:			
CAPI (ϕ)	0.054	–2958	0.004
	(0.180)	(2219)	(0.049)
Checks (ρ)	0.292**	6477***	0.154***
	(0.144)	(1,745)	(0.040)
N	2683	2137	5148
p-value of *F*-test ($\delta + \phi + \rho = 0$)	0.00	0.00	0.00

Notes: Standard errors are shown in parentheses. All regressions control for age fixed effects (FE). All estimates are robust to cluster and interviewer fixed effects, controlling for respondent characteristics (age, head of household, sex, literacy) and household characteristics (number of household members) and interview characteristics (day of field visit).

[a]To avoid spurious correlation in the results of column 2, we exclude the education component from the consumption aggregate used in that regression.

AEU, adult equivalent unit; CAPI, computer-assisted personal interviewing;

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

through consumption, to influence the outcome in question, but children's outcomes to have no influence on consumption.

The results, displayed in Table 2.8, show that, on paper, the number of years of schooling a child has attained is independent of household consumption. Once checks are included, however, a positive association emerges, consistent with attenuation bias in the PAPI and restricted CAPI results. Total consumption does explain schooling expenditures for children at school and whether or not a child slept under a bednet, but the size of the effect is estimated at around half of what full CAPI estimates, again consistent with attenuation bias. The coefficient on CAPI is not significant, indicating

that it is the consumption reports and validation checks that are responsible for these differences and not the bundle of other CAPI features.

One critique of these regressions could be that the left-hand-side variables are themselves measured more accurately in CAPI than in PAPI. The computer verified during the interview, for example, that the grade attained was sensible given the age of the child, flagged zero education expenditures for school-going children and expenditures over TZS 3 million as unlikely and cross-checked the transport costs for going to school with other entries in the transport section. Finally, there was a cross-validation check between bednet use and ownership of a bednet in the assets section. However, estimating $O_i = \alpha + \beta C_i + \gamma V_i + \varepsilon_i$ yielded insignificant coefficients on C_i and V_i, giving some confidence that the results are not affected by this possibility.

We also estimated all coefficients, restricting ourselves to within-sample estimates and artificially (and randomly) reducing the full CAPI sample to have the same observations as restricted CAPI and PAPI. All results remained intact after this exercise. Results are also robust to inclusion of cluster fixed effects, controlling for respondent characteristics (age, head of household, sex, literacy), household characteristics (household size), interviewer fixed effects and interview characteristics (conducted on the first day of the team's visit). In 13 PAPI child records at least one of these variables was missing, in which case we included a dummy indicating this.

The second test starts from the standard result that there is no bias if just the dependent variable suffers from classical measurement error. To investigate whether the evidence of this experiment is consistent with the latter property of classical measurement error, we run a regression explaining log total consumption per AEU with various factors typically included in the analysis of household poverty, and interaction effects between each of them with C_i and V_i. Under conditions further specified below,

significant interaction effects would prove, by contraposition, that measurement error is not classical.

In specifying our set of explanatory variables, we follow the guidelines provided by Haughton and Khandker (2009). The explanatory variables (household characteristics) included in the first regression in Table 2.9 are: household size, dependency ratio (counting adults as people aged 15–65 years), a dummy for whether the household head is female, average number of years of education of adult household members (aged 15 and above), number of days the household head was ill in the last 4 weeks, dummy for whether the household owns its house, dummy for whether the house has a modern roof (made of iron, concrete/cement, tiles or asbestos), the proportion of employed adults in the household and the acres of land owned. The regressions also control for cluster fixed effects.

We verified that all these variables are themselves not suffering from attenuation bias by regressing each one on C_i and V_i and ensuring that both coefficients are insignificant. We considered, but omitted, other variables, such as a dummy equal to 1 if the household head is *not* employed in the agricultural sector, the number of productive assets and the number of livestock owned by the household, because they were dependent on C_i and V_i and so could suffer from attenuation bias.

Column 1 in Table 2.9 displays the results of this regression and finds no significant interaction effects of CAPI and checks with determinants of poverty. This finding is consistent with the dependent variable having classical measurement error. Dropping the insignificant regressors from the analysis does not alter the results.

Table 2.9 Implications of measurement error in the dependent variable

	Log total consumption/AEU		Log total consumption/AEU	
	Final set of regressors (1)		Initial set of regressors (2)	
CAPI	–0.203	(0.196)	–0.279	(0.209)
Checks	0.055	(0.154)	0.124	(0.163)
HH size	–0.077***	(0.011)	–0.075***	(0.011)
Dependency ratio (adults 15–65 years old)	0.020	(0.023)	0.021	(0.023)
Dummy = 1 if HH head is female	–0.233***	(0.066)	–0.212***	(0.067)
Average education (years) adults (15+)	0.018**	(0.008)	0.015*	(0.008)
Number of days HH head ill in last 4 weeks	–0.007**	(0.003)	–0.007**	(0.003)
Dummy = 1 if HH owns its house	0.105	(0.079)	0.092	(0.079)
Dummy = 1 if house has robust roof	0.085*	(0.049)	0.063	(0.050)
Proportion of employed adults in the HH	0.211*	(0.128)	0.227*	(0.130)
Acres of land owned	0.028*	(0.015)	0.029*	(0.015)
Dummy = 1 if head *not* employed in agriculture			0.062	(0.052)
Number of productive assets owned by the HH			0.056	(0.037)
Number of livestock owned by the HH			–0.000	(0.002)
CAPI interacted with:				
HH size	0.010	(0.015)	0.002	(0.015)
Dependency ratio (adults 15–65 years old)	0.015	(0.034)	0.013	(0.033)
Dummy = 1 if HH head is female	0.068	(0.088)	0.098	(0.089)
Average education (years) adults (15+)	0.012	(0.011)	0.012	(0.012)
Number of days HH head ill in last 4 weeks	0.007	(0.004)	0.007	(0.004)
Dummy = 1 if HH owns its house	–0.055	(0.112)	–0.048	(0.111)
Dummy = 1 if house has robust roof	–0.028	(0.069)	–0.046	(0.069)
Proportion of employed adults in the HH	–0.109	(0.184)	–0.095	(0.187)
Acres of land owned	0.020	(0.023)	–0.003	(0.023)
Dummy = 1 if head *not* employed in agriculture			0.020	(0.072)
Number of productive assets owned by the HH			0.047	(0.050)
Number of livestock owned by the HH			0.017***	(0.005)
Checks interacted with:				
HH size	0.008	(0.011)	0.016	(0.011)
Dependency ratio (adults 15–65 years old)	–0.023	(0.027)	–0.024	(0.027)
Dummy = 1 if HH head is female	–0.048	(0.066)	–0.061	(0.067)
Average education (years) adults (15+)	–0.007	(0.009)	–0.009	(0.010)
Number of days head ill in last 4 weeks	–0.002	(0.003)	–0.002	(0.003)
Dummy = 1 if HH owns its house	–0.053	(0.089)	–0.052	(0.087)
Dummy = 1 if house has robust roof	0.036	(0.054)	0.064	(0.054)
Proportion of employed adults in the HH	–0.073	(0.147)	–0.067	(0.150)
Acres of land owned	–0.023	(0.019)	–0.003	(0.019)
Dummy = 1 if head *not* employed in agriculture			0.000	(0.057)
Number of productive assets owned by the HH			–0.078**	(0.039)
Number of livestock owned by the HH			–0.012***	(0.004)
_cons	13.242***	(0.143)	13.164***	(0.151)
N	1,837		1,837	

Notes: Standard errors are shown in parentheses. Both equations control for cluster fixed effects. Robust roof materials are considered to be iron, concrete/cement, tiles or asbestos. Robust to different measures of household health (e.g. average health household members) and household educational level (e.g. average level of education of all household members). The choice of measures used in this table was based on minimising the difference in means of the measures between CAPI and pen-and-paper interviewing to avoid attenuation bias affecting the results. Three observations with outliers in land area owned were dropped from the sample.

CAPI, computer-assisted personal interviewing; HH, household. ***$p<0.01$, **$p<0.05$, *$p<0.1$.

For the sake of completeness, column 2 adds the problematic regressors and finds their interactions are indeed significant. This does not lead us to reject the classical measurement hypothesis, as these results could simply be driven by attenuation bias on the PAPI estimates (insignificant level effects), which is alleviated through CAPI (significant interaction effects).

Finally, we looked at the joint significance of all interaction terms with C_i, with V_i and both groups of interaction terms together and rejected all three times the null that they are not significantly different from zero in column 1 of Table 2.9. We also ran more parsimonious regressions – one for each of the regressors – and found that, with the exception of the regression that includes the proportion of household members who are employed, all interaction effects were insignificant in all regressions. We do not consider our one exception sufficient evidence to reject the null of classical measurement error: considering the number of coefficients we are scrutinising, we would expect some false positives to be present.

2.6. Further dimensions of comparison

CAPI has a larger fixed cost component: up-front outlays for the development of the software and the purchase of the hardware. But many of the variable costs (per interview) that are incurred for PAPI, such as printing and data entry, are eliminated in CAPI. This makes CAPI, budget-wise, a more viable option for larger surveys. As a smooth functioning rental market for hand-held devices does not exist, surveys with fewer interviewers, spread over a longer time, will be cheaper, as fewer machines need to be bought. Organisations that regularly conduct surveys could share machines across projects to overcome this problem.

The software used in this experiment (and the baseline survey on which it was based) is based on Microsoft Access and had been under development for 2 years.

However, it still required approximately 50 consultancy days each of a senior and junior program developer to adapt it for this specific survey. The survey had approximately 80% similarity with other surveys already conducted with the same program. This comes to a fixed cost of US$40,000 in consultancy fees for making the program. A data entry application could be developed for around US$4000 using consultants at similar rates. The UMPCs used in this experiment, including peripherals such as extra batteries, a replacement battery after 2 years of use, GPS units, bags, charging equipment and transport, were approximately US$1800 a piece and have an estimated life time of 600 interviewing days or approximately US$3 per day. Interviewers conducted three interviews per day in this survey, so the variable UMPC cost per questionnaire was approximately US$1 per questionnaire. The variable costs per paper questionnaire were approximately US$4 for data entry clerks and desktop computers, US$4 per questionnaire for data entry management and supervision (including adjudication of errors) and US$2 for printing a single questionnaire.

Thus, in the context of this survey solving $40{,}000 + X = 4000 + 10X$ for X gives a break-even point of 4000 questionnaires. Below this, paper is cheaper and above this CAPI is cheaper. For example, a survey of 2500 households would be US$13,500 more expensive on CAPI, whereas a survey of 10,000 households would be US$54,000 cheaper on CAPI. Adding wasted observations to this changes this number. For example, if a paper survey needs to collect 10% more observations, then the break-even point drops to 3600 questionnaires. The break-even point can drop even further if one considers reduced interview length (see below) and reduced data-cleaning efforts.

Driven by the popularity of Apple's iPad, many hardware manufacturers are now developing their own UMPCs, and it seems likely that the price of a machine will go

down to around US$300.[17] Also, in terms of software, this project used a hard-coded questionnaire that can only be adapted by experienced software engineers. This system was used because existing products (e.g. the program used by Fafchamps et al. (2012)) did not allow us to build in the complexity we needed. Once the market develops off-the-shelf products that do not require software engineers to be involved, there should be no reason to believe that making an electronic questionnaire takes longer than making a data entry program. Fafchamps et al. (2012) cite 75 hours of researcher time to build an enterprise survey and 20 hours to program the follow-up questionnaires; without hiring the services of a software specialist.

Table 2.1 showed that interview time is reduced in CAPI by 10% (and by 14% in the restricted CAPI without validation checks). Examples of CAPI features that may be responsible for this reduction are the automatic routing system and the use of drop-down menus from which to select responses (instead of codes listed in a box somewhere else on the page). Finally, we see virtually no differences in the respondent's perception (e.g. degree of intimidation, perception of confidentiality, etc.) between CAPI and PAPI (see Appendix 2.A.3).

2.7. Conclusion

Many researchers and survey implementers are keen to switch from paper to electronic surveys, but there is currently little quantitative, empirical information available to inform that choice. This chapter uses data from a survey experiment to identify differences between PAPI and CAPI and finds that errors leading to missing variables in PAPI are virtually eliminated in CAPI. A simple, compensatory increase in sample size on paper cannot adequately deal with this problem, because observations are not randomly dropped. The effect of CAPI is particularly evident when measuring

[17] At the time of submission of this thesis, non-touch-screen UMPC units had already gone down to US$400 whereas touch-screen UMPCs were being sold for approximately US$650 per unit. Battery equipment costs currently range from US$100 to US$350.

consumption. We find that paper questionnaires can lead to estimates of higher mean consumption, lower poverty and higher inequality.

We performed a number of regression analyses using the consumption aggregates as dependent and explanatory variables. Paper and restricted CAPI suffer from attenuation bias when consumption aggregates are used as regressors. We do not, however, find evidence of bias when consumption aggregates are used as the left-hand-side variable of a regression model. Hence, our evidence is consistent with classical measurement error.

Although there is scope for mimicking some of the CAPI features on paper, this seems unfeasible for the checks and reports made available to the interviewer during the interview. Results show that these two features play a key role, especially for reducing errors in consumption data. We further explain why this specific CAPI product leads to higher fixed up-front costs, but lower variable per-questionnaire costs. Finally, we show that interview times are significantly lower in CAPI, whereas there is no change in respondents' experience.

Some analytical caveats remain, however. First, the experiment does not provide an independent validation of the data, as in, for example, Bound and Krueger (1991) and Bound et al. (1994). One could argue that in the analysis in section 2.4, where we use error counts at the left-hand side of the regression models, any dependency between the checks and the error count variables could bias the results.

A second critique is that an (apparently) error-free questionnaire is not the same as one that reflects reality. One may worry that interviewers can now simply enter any data that the computer is willing to accept. An unscrupulous interviewer could simply change a value to anything that suppresses the error message rather than make the effort to obtain the correct value. This survey was subject to intense quality control: supervisors carried out daily direct observations and brief re-interviews of respondents to check the validity of the data. An independent quality controller went to a random

subset of 18 EAs to conduct re-interviews. The interviewers were aware that these random quality control visits would take place and no false data were detected in either PAPI or CAPI. As members of staff of a survey research company (EDI), most interviewers intend to be on board for several surveys and several years. Moreover, this CAPI application kept time stamp records of all response entries and all consistency alerts, which allowed the supervisors to closely monitor the sequence and the rate at which the survey questions were answered. However, not every CAPI survey project will be able to afford such intensive monitoring efforts. Therefore, more generally, concerns regarding the possibility that consistency checks may help enumerators in faking answers or exerting insufficient caution remain.

Similarly, there are other possible threats to the assumption that a 'consistent' questionnaire (i.e. one which passes the myriad of built-in consistency checks) equals a questionnaire that reflects reality. For instance, validation alerts may force respondents to change truthful answers for the sake of enforcing consistency in the survey, whilst generating false complacency among researchers about the quality of the data they collected. The enumerators who were employed for the specific experiment used in this chapter were intensively trained in terms of their active interviewing skills (as opposed to passively following consistency alerts and blindly recording responses provided by the respondent). However, as before, it is fair to say that not all CAPI projects are expected to provide such intensive training to their interviewers.

Third, CAPI is not a panacea. We have found that the success of CAPI depends substantially on the effort spent programming, piloting and testing the application, as well as on careful consideration of the underlying data management and transfer systems. This specific experiment was designed by survey implementers experienced in the field realities of complex household surveys, spending considerable resources refining the application before taking it to the field. Without such preparation PAPI may

well outperform CAPI. More generally, one needs to keep in mind that this experiment had no variation with respect to CAPI applications.

The literature from the developed world, 10–15 years ago, talks about the inevitability of the switch from CAPI to PAPI; despite the challenges, the benefits are so attractive that a switch seems irresistible. It is our contention that a similar desire is growing among survey practitioners working in lower income countries, and with good reason, as this chapter illustrates.

2.A. Chapter 2 Appendix

2.A.1. Screen shots of the CAPI application used in the experiment

Figure 2.4 Screen shot of section 'household head information'

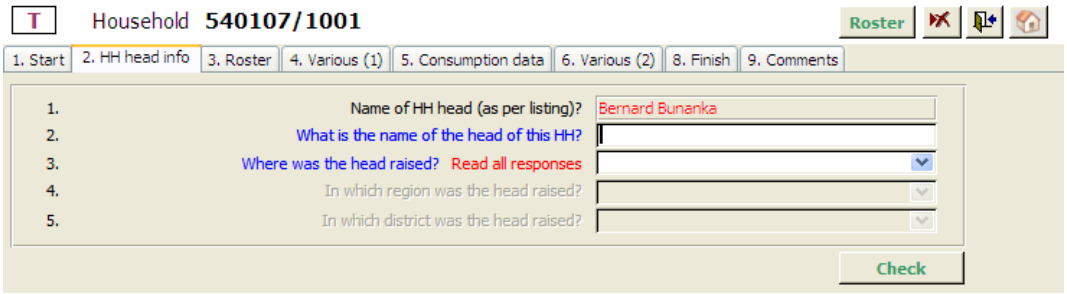
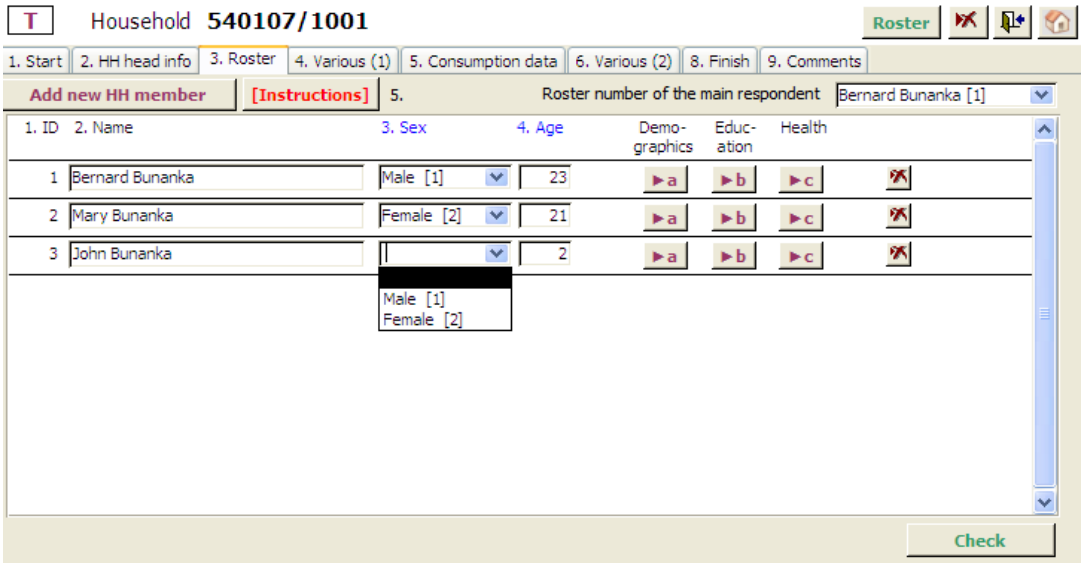


Figure 2.5 Screen shot of section 'roster'



2.A.2. Validation checks in full CAPI

Table 2.10 Subset of a total of 366 validation checks used in full CAPI

Section	Type	Warning message
Agriculture	Error	Duplicate crops have been selected
Agriculture	Missing	Missing question 5: Do you have a title deed/offer letter for any land that you own?
Amenities	Missing	Missing question 5: Do you have any documentation of ownership of the dwelling?
Amenities	Warning	It is uncommon to have electricity in a thatch roofed house. Please verify whether this information is correct.
Assets	Warning	Are you sure the household owns a TV antenna or video/DVD but does not own a TV?
Assets	Warning	It is unlikely that someone with main activity self-employed farming does not own any farming equipment at all (wheelbarrow/cart, nor harrow/plough, nor other farming equipment). Please verify
Assets	Warning	At least one household member slept under a mosquito net last night although the household claims not to have any mosquito nets. This is possible, but please verify
Consumption	Error	Respondent to the consumption questions should not be less than 12 years of age
Finish	Error	Entry in disabled question 2: How proficient was the respondent in Swahili?
Food consumption	Error	The household has consumed this item, but no source for obtaining this consumption item (Purchased/Produced/Other sources) has been selected
Demographics	Error	The relationship of [NAME] is Wife/Husband, although the marital status is not married or no non-formal union
Demographics	Warning	This person is very old to be in boarding school. Are you sure this information is correct? Please verify
Demographics	Warning	Person is said to be in boarding school but its main occupation is said to be different than ‘student’. Please verify
Demographics	Warning	The man is a polygamist, hence it is unlikely that he is present all year long, unless all his wives live in the same household. Please verify
Education	Error	The last grade of examination cannot be higher than 1 grade higher than the highest level of completed education
Education	Error	The age when he/she started school cannot exceed his/her current age
Education	Missing	Missing question 16: Who runs/manages school [NAME] is attending?
Education	Warning	Main activity was recorded as full-time student but person is not currently in school? Please verify
Health	Error	It is not possible that a male person has pregnancy related problems.
Health	Error	The days ill in 4 weeks cannot exceed 28
Livestock	Warning	It is uncommon to own such a high number of goats. Are you sure the given information is correct?
Roster	Error	Household head claims to have monogamous marriage but there is more than one wife

Table 2.11 Fifteen most common missing values outside the consumption section in pen-and-paper interviewing (PAPI)

Section	Validation check message	% of surveys where error occurred at least once			Average frequency of error in surveys where error occurred		
		Full CAPI	Restricted CAPI	PAPI	Full CAPI	Restricted CAPI	PAPI
Start and finish	Error in interview duration calculation	0.9	0.3	23.8			
Agriculture	Missing: Over the past year, how much did you spend on transport in order to sell the products from [Crop Item]?	0.0	0.0	9.4	0.0	0.0	1.4
Health	Missing: In the past 5 years, has [HH Member] given birth to children (including children dead)?	0.1	3.1	8.1	1.0	1.0	1.1
Education	Missing: How much was spent by the members of your household on [HH Member]'s transport to/from school?	0.0	0.0	7.2	0.0	0.0	1.3
Agriculture	Missing: In which of those 2 places (i.e. selling places) do you fetch the highest price per unit of [Crop Item].	0.0	0.3	6.9	0.0	1.0	1.4
GPS	Error with GPS location	0.6	1.3	6.6			
Education	Missing: Has [HH Member] ever attended school?	0.0	0.0	6.3	0.0	0.0	1.3
Health	Missing: Is [HH Member] permanently physically or mentally disabled in any way which limits or prevents normal daily activities or work?	0.0	0.0	5.6	0.0	0.0	2.4
Education	Missing: What is the one way fare to go to school using [Transportation Mode]?	0.0	0.3	5.6	0.0	1.0	1.3
Agriculture	Missing: Do you own any agricultural land/farm (including grazing and fallow land)?	0.0	0.0	5.3			
Agriculture	Missing: Rank the most important crops for generating cash income	0.0	0.0	4.7	0.0	0.0	1.6
Demographics	Missing: What is [HH Member]'s marital status?	0.0	0.6	4.4	0.0	1.0	1.6
Education	Missing: Can [HH Member] read and write?	0.0	0.0	4.4	0.0	0.0	1.0
Demographics	Missing: Is [HH Member] an actual member of this household (satisfying some membership criteria, such as being present at the household for at least 9 out of 12 past months)?	0.0	0.0	4.1	0.0	0.0	1.2
Health	If the number of days person was too ill to perform his/her normal daily activities in the past 4 weeks is greater than 0, the answer to 'In the past 12 months was [HH Member] too ill to perform his/her normal daily activities?' cannot be 'no'	0.0	0.0	3.8	0.0	0.0	1.5

CAPI, computer-assisted personal interviewing; PAPI, pen-and-paper interviewing; GPS, global positioning system; HH, household.

Table 2.12 Ten most common missing values in consumption section in pen-and-paper interviewing (PAPI)

Section	Validation check message	% of surveys where error occurred at least once			Average frequency of error in surveys where error occurred		
		Full CAPI	Restricted CAPI	PAPI	Full CAPI	Restricted CAPI	PAPI
Consumption	The total amount of consumption of this food item differs from the sum ('How much came from purchases' + 'How much came from own production' + 'How much came from gifts and other sources')	0.1	2.8	17.2	1.0	1.0	1.5
Consumption	Missing: How much [Food Item] came from 'gifts and other sources' in the past [Recall Period]?	0.0	5.0	14.0	0.0	1.0	1.3
Consumption	Missing: How much [Food Item] came from 'own production' out of what was spent in the last [Recall Period]?	0.0	3.4	10.3	0.0	1.0	1.2
Consumption	Missing: UNIT of 'How much [Food Item] did your household consume in the past [Recall Period]'?	0.0	0.0	10.0	0.0	0.0	1.2
Consumption	Missing: How much [Food Item] came from 'purchases' out of what was spent in the last [Recall Period]?	0.0	0.6	7.2	0.0	1.0	1.9
Consumption	Missing: How much expenditure information for [HH Member] is not captured in what you have mentioned to me?	0.1	0.9	6.3	1.0	1.3	2.0
Consumption	Missing: In the past 7 days did household consume any [Food Item]?	0.0	9.1	5.6	0.0	1.3	3.8
Consumption	Missing: How much [Food Item] did your household consume in the past [Recall Period]?	0.0	0.3	5.0	0.0	1.0	1.0
Consumption	Missing: In the past [Recall Period] did household consume/purchase any [Non Food Item]?	0.0	5.6	4.1	0.0	1.2	1.1
Consumption	The household has consumed [Food Item], but no source for obtaining this item (Purchased/Production/gifts) has been selected	0.0	0.0	1.9	0.0	0.0	1.0

CAPI, computer-assisted personal interviewing; PAPI, pen-and-paper interviewing; HH, household.

2.A.3. Respondent's perception of the interview

Table 2.13 Respondent's perception by survey method (%)

Question	Response options	Full CAPI	Restr. CAPI	PAPI
What did you think of the duration of the interview?[a,b]	Short (1)	26.3	35.1	37.5
	Long (2)	73.7	64.9	62.5
	Sd	0.4	0.5	0.5
Did you enjoy participating in the interview?	Yes (1)	93.2	90.9	89.8
	No (2)	6.8	9.1	10.2
	Sd	0.3	0.3	0.3
How smooth did you find the interview in terms of flow of questions?	Bad (1)	1.2	2.0	1.0
	Normal (2)	29.3	30.2	27.3
	Rather good (3)	61.7	60.1	63.5
	Very good (4)	7.7	7.8	8.3
	Sd	0.6	0.6	0.6
Do you believe your answers will get used for policy making?	Yes (1)	88.8	87.8	87.4
	No (2)	11.2	12.2	12.6
	Sd	0.3	0.3	0.3
We are not talking about this specific interview, but do you think that results of these types of surveys are generally reliable?	Not reliable at all (1)	2.5	2.9	2.2
	Rather reliable (2)	90.7	91.2	90.5
	Very reliable (3)	6.8	5.8	7.0
	Sd	0.3	0.3	0.3
Do you believe that the information you provided in the interview is 100% confidential?	Yes (1)	95.1	98.1	96.1
	No (2)	4.9	2.0	3.8
	Sd	0.2	0.1	0.2
If we went through the survey again, do you think any answers would change?	Yes (1)	25.0	20.2	20.1
	No (2)	75.0	79.8	79.9
	Sd	0.4	0.4	0.4
Did you feel comfortable talking to the interviewer?	Not at all (1)	4.3	2.0	3.5
	A little (2)	34.0	27.6	32.5
	Very much (3)	61.7	70.5	64.0
	Sd	0.6	0.5	0.6
How nervous did you feel during the interview?	Not at all (1)	86.7	90.9	86.7
	A little (2)	11.1	8.1	9.8
	A lot (3)	2.2	1.0	3.5
	Sd	0.4	0.3	0.5
How difficult did you find the questions?[c]	Easy	38.3	29.6	35.7
	Difficult (1)	61.7	70.5	64.3
	Sd	0.5	0.5	0.5
If we were not recording the answers (just talking to you), do you think you would have answered anything differently?	Yes (1)	29.6	31.8	27.7
	No (2)	70.4	68.2	72.3
	Sd	0.5	0.5	0.5
Would you participate to this survey again next year?	Yes (1)	95.4	97.0	98.1
	No (2)	3.1	1.7	1.3
	Maybe (3)	1.5	1.3	0.6
	Sd	0.3	0.3	0.2

Notes: [a]Full CAPI covered more modules than restricted CAPI and PAPI and hence had a longer interview time, which may affect responses (see section 2.2). [b]Answers grouped as: short = 'short' + 'just fine'; long = 'long' + 'too long' (with no change to results). [c]Answers grouped as: easy = 'very easy' + 'rather easy'; difficult = 'rather difficult' + 'very difficult' (with no change to results). sd, standard deviation.

2.A.4. More examples of pictures used in the CAPI consumption section

Figure 2.6 Bunches of bananas (with a bike and chair as a reference)

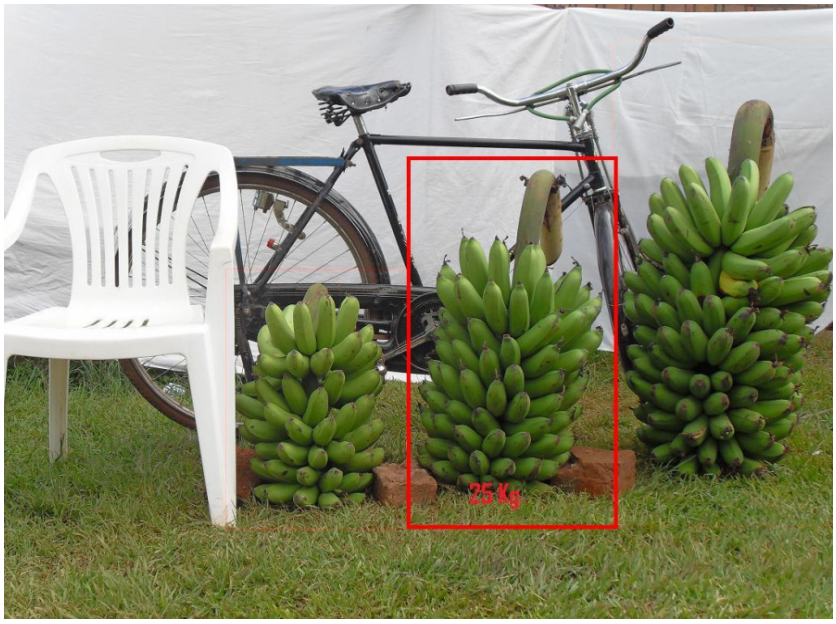
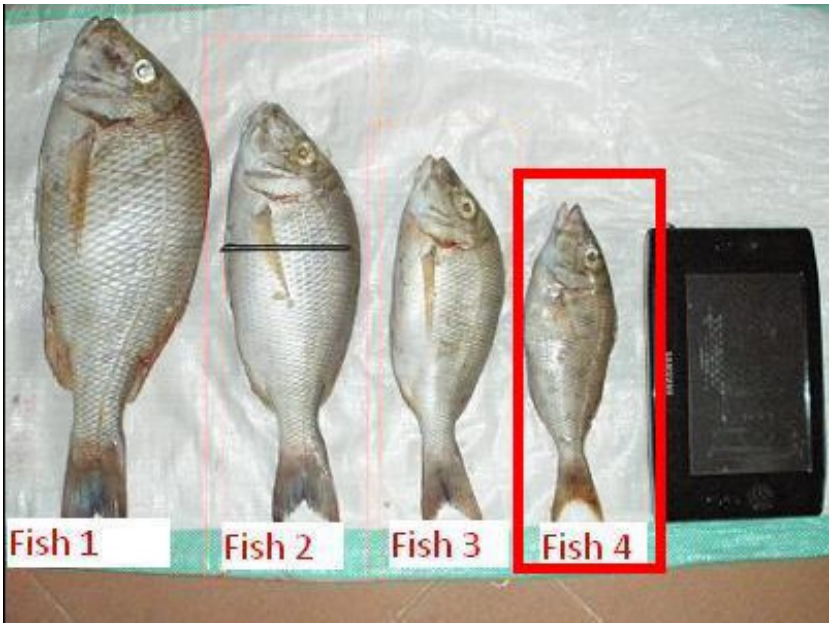


Figure 2.7 Fish (with a UMPC unit as a reference)



Chapter 3

Exclusion bias in empirical social interaction models: causes, consequences and solutions[18]

3.1. Introduction

In the social interaction literature, the standard model for estimating endogenous peer effects is the following linear-in-means specification (for a review, see Durlauf and Ioannides, 2010):[19]

$$y_{li} = \beta_0 + \beta_1 \bar{y}_{k_{li}} + \varepsilon_i \qquad (3.1)$$

where y_{li} denotes the outcome of individual i in group l and $\bar{y}_{k_{li}}$ is the average outcome of individual i's k peers, which exclude individual i herself. The identification of the peer effect β_1 typically encounters various empirical challenges extensively discussed in the literature, such as the reflection problem (Manski, 1993), endogenous peer group sorting and correlated effects (Brock and Durlauf, 2001, 2007; Moffitt, 2001; Graham and Hahn, 2005; Soetevent, 2006; Graham, 2008). Various methodologies have been suggested to deal with these identification obstacles (e.g. Laschever, 2005; Lee, 2007; Lin, 2007; Graham, 2008; Bramouille et al., 2009; De Giorgi et al., 2009).

One such strategy is to use randomisation methods to assign peers to individuals. Until recently, it was assumed that such a randomisation strategy ensures that no *ex ante* systematic relationship exists between individuals and their peers. As a result, any observed correlation between outcomes *ex post*, conditional on common shocks, was attributed to peer effects. For instance, this strategy has been used to study peer effects on educational outcomes, where y_{li} in equation (3.1) indicates the test score of student i

[18] I gratefully acknowledge the feedback I received from the audience of a Gorman Research Workshop at the Economics Department of the University of Oxford, UK, where I presented this work. I would also like to thank Matthew Collin and Simon Quinn for their constructive feedback on this chapter.

[19] Although my basic model considers a simple bivariate regression, this chapter also briefly discusses the implications for the results of the inclusion of control variables to the model.

in classroom l and $\bar{y}_{k_{li}}$ denotes the average test score of that student's k peers (e.g. Sacerdote, 2001).

However, Guryan–Kroft–Notowidigdo (2009), henceforth GKN (2009), have argued that, even with random peer assignment, a negative relationship exists between people's characteristics and those of their peers. The intuition for this result builds on the fact that, in equation (3.1), individuals cannot be their own peers, so that the pool of potential peers from which an individual's peers are drawn systematically excludes the individual herself. Consequently, the pool of potential peers differs for each individual, and its expected value will be negatively correlated with the characteristics of the individual. Using Monte Carlo simulations, GKN (2009) show how this correlation yields a downward bias in the ordinary least squares (OLS) estimate of the endogenous peer effect, β_1, in typical tests of random peer assignments (where the true $\beta_1 = 0$ in a regression on baseline characteristics). Henceforth, I refer to this type of bias as the *exclusion* bias, reflecting the fact that the bias is caused by an exclusion of individuals from their own pool of potential peers. To my knowledge, the only attempt to formalise the intuitive findings of GKN (2009) has been made by Wang (2010). His study shows, both analytically and using simulations, that standard tests of random assignment of peers to students are biased, for the specific case where a student's peers are defined as all his/her classmates.

This chapter contributes to the social interaction literature in a number of ways. First, it provides a more general formalisation of the proposition that OLS estimates of the endogenous peer effect are biased downwards in standard peer effects models. I generalise the linear-in-means model in GKF (2009) and Wang (2010) in two ways: (i) I allow the number of peers k to differ from the total size of the group l in equation (3.1) (i.e. peer groups do not necessarily have to consist of *all* a student's classmates); (ii) I

allow the true endogenous peer effect to be greater than zero $(\beta_1 \geq 0)$, rather than focusing on a model that tests for random peer assignment $(\beta_1 = 0)$.

Next, I derive a clear analytical expression for the magnitude of the bias, and discuss the underlying parameters that determine it. When deriving this expression in a model that allows for cluster sampling, I make a careful distinction between models that add cluster fixed effects (e.g. classroom dummies) and models that do not. My results show that, in some common empirical settings, the exclusion bias becomes significantly more severe in models adding cluster fixed effects. Moreover, I show that in the latter type of models, the bias does not disappear as the sample size tends to infinity.

Furthermore, I suggest a new solution to the bias. I list various limitations to the method suggested by GKF (2009) and I show that two-stage least squares (2SLS) procedures are generally more useful in eliminating the bias. With the latter result, this chapter further contributes to the literature by explaining a counter-intuitive yet common finding in peer effects studies. Many studies on social interactions obtain 2SLS estimates of endogenous peer effects that are significantly larger than OLS estimates (e.g. Goux and Maurin, 2007; De Giorgi et al., 2010; Brown and Laschever, 2012; de Melo, 2011; Helmers and Patnam, 2012; Krishnan and Patnam, 2012; Naguib, 2012; Collin, 2013).[20] This is counter-intuitive, as one would expect the OLS estimate to be biased *upwards* as a result of other sources of bias such as the reflection bias (Manski, 1993), endogenous peer group formation and/or unobserved correlated effects (Brock and Durlauf, 2001; Moffitt, 2001). This finding has mainly been ignored in the literature to date, except by authors such as Goux and Maurin (2007), de Melo (2011), Helmers and Patnam (2012) and Collin (2013), who attribute it to classical measurement error and/or to the local average treatment interpretation of 2SLS. This chapter provides an

[20] These studies are examples in which both OLS and 2SLS estimates are reported. Many other studies present only 2SLS results, and comparisons with OLS cannot be made (e.g. Lalive and Cattaneo, 2009; Fletcher, 2012).

alternative explanation for this finding, by showing that the negative exclusion bias that is present in OLS estimation is eliminated in 2SLS estimation strategies.

Although the exclusion bias is also present in models in which peer selection is non-random, the focus of this chapter is on random peer assignment, which allows me to evaluate the magnitude of the bias without conflating it with the bias caused by endogenous peer group formation. Moreover, this chapter shows that the exclusion bias is only a concern in models in which (i) each individual is excluded from her own peer group, but is included in the peer group of other sampled individuals, and (ii) peer effects specifications fail to account for the difference in expected peer group outcomes that such self-exclusion generates. In the concluding section of this chapter, I discuss examples of empirical specifications that are expected to be free of the exclusion bias (e.g. Lyle, 2007, 2009; Duflo et al., 2011; Fafchamps and Quinn, 2012).

In section 3.2, I begin by constructing a basic model to illustrate that the OLS estimate of the endogenous social interaction effect is biased downwards in specifications where individuals are excluded from their own peer groups. I proceed by deriving an analytical expression for the magnitude of the bias, and obtain results on the parameters that determine it. In section 3.3, I show that in some common empirical specifications the exclusion bias is aggravated in models adding cluster fixed effects. In section 3.4, I evaluate the method suggested by GKF (2009) to eliminate the bias, and argue that 2SLS procedures are generally more effective in doing so. In section 3.5, I present simulation results, which confirm the theoretical findings of the chapter. Section 3.6 discusses extensions of the basic model and considers how these would alter the main results. Finally, in section 3.7, I discuss the implications of my results for the empirical peer effects literature.

3.2. Basic model

Suppose we have a sampled network Ω of N individuals. For instance, we have a school Ω with N students. Let subset $\Pi_i \subset \Omega$ be the pool of N_P potential peers for individual i, from which individual i is randomly assigned k peers. In my basic model, I ignore potential clustering and I assume that peers are drawn at the level of the sampled network Ω (e.g. school), that is, $\Pi_i = \Omega$ for all $i \in \Omega$. In the next section, I will consider a model in which peers are drawn from a subset $\Pi_i = l$ of size $L < N$ (e.g. a classroom within the school). Denote the person-specific group of randomly selected k peers by P_i. Throughout this chapter, I assume that individual i is excluded from her own reference group, that is, $i \notin P_i$. Note that in this basic model, we assume that peer groups are person-specific and not necessarily overlapping; that is, the random selection of peers for one person does not depend on the random selection of peers for another person. In the next section, we will consider a more general model that allows for grouping of friends. Nevertheless, the current basic model is interesting. For instance, it forms the basis of specifications in which students are asked to identify their individual friends from the school roster (e.g. Fletcher, 2012).

Let y_i be an outcome variable of interest of individual $i \in \Omega$ (e.g. a student's test score). Consider a simple bivariate linear-in-means social interaction model in which an individual i's outcome y_i is regressed on the average outcome of her k (randomly assigned) peers $\bar{y}_{k_i}$:[21]

$$y_i = \beta_0 + \beta_1 \bar{y}_{k_i} + \varepsilon_i \qquad (3.2)$$

where parameter β_1 captures the peer effect. The exposition in this chapter assumes either positive or zero true peer effects, that is, $\beta_1 \geq 0$, which is natural in this setting.

[21] In section 3.6, I discuss how the addition of control variables affects the results of the model.

The error term ε_i reflects unobservable characteristics associated with *i*. To focus on the effects of exclusion bias, I assume zero correlated effects. The main result of this section is to formally show that, despite the random peer assignment, the explanatory variable $\bar{y}_{k_i}$ will be correlated with the error term ε_i, leading to bias in OLS estimates of the social interaction effect β_1.

To see this, we expand $\bar{y}_{k_i}$ in equation (3.2). In our basic model, the person-specific peer group P_i consists of a random set of *k* individuals from the pool of $(N-1)$ individuals in network Ω (excluding individual *i* herself). Let $\bar{y}_{-i}$ be the average outcome of this pool of $(N-1)$ individuals. Given random peer selection, the expected value of the average outcome of the *k* selected peers will be equal to $\bar{y}_{-i}$, that is, $E(\bar{y}_{k_i}) = \bar{y}_{-i}$. The actual draw, $\bar{y}_{k_i}$, will deviate from $\bar{y}_{-i}$ by a random component u_i:

$$\bar{y}_{k_i} = \bar{y}_{-i} + u_i \tag{3.3}$$

where $E(u_i) = 0$.

Inserting equation (3.3) into equation (3.2), we obtain:

$$y_i = \beta_0 + \beta_1(\bar{y}_{-i} + u_i) + \varepsilon_i \tag{3.4}$$

where we assume that $E(u) = E(\varepsilon) = 0$, $\text{cov}(u, \varepsilon) = 0$, $\text{var}(u) = \sigma_u^2$ and $\text{var}(\varepsilon) = \sigma_\varepsilon^2$.

It is important to note here that, in a model where peers form part of the study sample, σ_u^2 will be correlated with σ_ε^2. Specifically, in Appendix 3.A.1 I show that:

$$\sigma_u^2 = \frac{(N-1-k)}{(N-1)k}\sigma_\varepsilon^2 < \sigma_\varepsilon^2 \tag{3.5}$$

Equation (3.5) shows that, in this basic model in which peers are randomly drawn from within the sampled network Ω of size *N*, σ_u^2 is decreasing in the peer group size *k*

and increasing in the sample size *N*. These findings will be important when interpreting the parameters affecting the magnitude of the exclusion bias below.

Returning to equation (3.4), note that $\bar{y}_{-i}$ can be written as:

$$\bar{y}_{-i} = \frac{\left(\sum_{i=1}^{N} y_i\right) - y_i}{N-1} \tag{3.6}$$

As a result, equation (3.4) can be rewritten as follows:

$$y_i = \beta_0 + \beta_1 \left(\frac{\left(\sum_{i=1}^{N} y_i\right) - y_i}{N-1} + u_i \right) + \varepsilon_i \tag{3.7}$$

Note that it is the presence of the dependent variable y_i on the right-hand side of equation (3.7) that leads the OLS estimate of β_1 to be biased downwards. To derive this correlation more formally, inserting equation (3.4) into equation (3.6) results in the reduced form of $\bar{y}_{-i}$:

$$\bar{y}_{-i} = \frac{\sum_{j=1}^{N} y_j - \beta_0}{N-1+\beta_1} - \frac{\beta_1 u_i}{N-1+\beta_1} - \frac{\varepsilon_i}{N-1+\beta_1} \tag{3.8}$$

Using equation (3.8) and the assumption that $\text{cov}(u, \varepsilon) = 0$, it is now straightforward to see that $\text{cov}(\bar{y}_{k_i}, \varepsilon_i) \neq 0$ and, therefore, that OLS estimation of equation (3.2) leads to a biased estimate of the social interaction effect β_1:

$$\text{cov}(\bar{y}_{k_i}, \varepsilon_i) = \text{cov}(\bar{y}_{-i} + u_i, \varepsilon_i) = \text{cov}(\bar{y}_{-i}, \varepsilon_i) = \frac{-\sigma_\varepsilon^2}{N-1+\beta_1} < 0 \tag{3.9}$$

Using equation (3.9), together with the expression for $\text{var}(\bar{y}_{k_i})$ derived in Appendix 3.A.2, we can now determine the magnitude of the bias in OLS:

$$E\left(\hat{\beta}_1^{OLS}\right) = \beta_1 + \frac{\text{cov}\left(\bar{y}_{k_i}, \varepsilon_i\right)}{\text{var}\left(\bar{y}_{k_i}\right)} = \beta_1 + \frac{\frac{-\sigma_\varepsilon^2}{(N-1+\beta_1)}}{\frac{(N-1)^2\sigma_u^2 + \sigma_\varepsilon^2}{(N-1+\beta_1)^2}} = \beta_1 - \frac{\sigma_\varepsilon^2(N-1+\beta_1)}{(N-1)^2\sigma_u^2 + \sigma_\varepsilon^2}$$

Inserting the expression for σ_u^2 presented in equation (3.5), I obtain the first result of this chapter, summarised by proposition 3.1.

Proposition 3.1. Suppose each individual *i* in a sampled network Ω of *N* individuals is randomly assigned *k* peers $\neq i$ from within the sampled network Ω. In a linear-in-means social interaction model, specified as in equation (3.2), the OLS estimate of the peer effect, $\hat{\beta}_1^{OLS}$, will be biased downwards according to the following expression:

$$E\left(\hat{\beta}_1^{OLS}\right) = \beta_1 - \frac{(N-1+\beta_1)k}{(N-1)(N-k-1)+k} < \beta_1$$

Proposition 3.1 tells us how, in a basic model in which peers are randomly selected from the total sample of *N* individuals, the magnitude of the exclusion bias is driven by three important parameters. Specifically, denoting the bias by

$$|bias| = \left|\frac{(N-1+\beta_1)k}{(N-1)(N-1-k)+k}\right|$$

we have:

1 $\frac{\Delta|bias|}{\Delta N} < 0$, *ceteris paribus*, the larger the size of the sampled network, from within which peers are drawn, the smaller the bias. This result, which is consistent with the simulation results in GKN (2009), is intuitive. The mean of the pool will be less sensitive to an exclusion of one of the outcome variables as the pool of potential peers becomes larger. Formally, as shown in equation (3.5), the larger the *N*, the larger the variance of *u*, σ_u^2. From equation (3.3) we know

that, as σ_u^2 becomes larger, more variation in $\bar{y}_{k_i}$ will be explained by the random component u_i rather than being governed by the mean of the pool of potential peers, $\bar{y}_{-i}$. This implies that, *ceteris paribus*, studies that consider a larger pool of potential peers will have less severe exclusion bias. In the basic model, in which the pool from which peers are selected coincides with the sampled network Ω, this result implies that the bias disappears as the sample size tends to infinity. In the next section, we will see that this is not necessarily the case in models in which peers are selected from a sub-set of the sampled network.

2 $\frac{\Delta|bias|}{\Delta k} > 0$, *ceteris paribus*, the larger the peer group size *k*, the larger the bias.

Note that the magnitude of the bias is not linear in *k*. If it were, then $\frac{\Delta|bias|}{\Delta k} > 0$ could simply reflect the fact that, although the absolute bias increases with the number of peers, the bias *per additional peer* remains constant. Proposition 3.1 indicates, however, that the bias per peer also increases with the total number of peers considered.[22] Intuitively, the larger the number of people who are considered in the peer group, the more closely the average peer group outcome $\bar{y}_{k_i}$ will follow the average outcome of the pool from which those peers are drawn, $\bar{y}_{-i}$. This means that $\bar{y}_{k_i}$ will be more sensitive to changes to the mean, $\bar{y}_{-i}$, such as the exclusion of an individual's outcome. Formally, as shown above in equation (3.3), as *k* becomes larger, the variance σ_u^2 will become smaller. From equation (3.4) we know that, as σ_u^2 becomes smaller, less of the variation

[22] From proposition 3.1 it is straightforward to see that the bias per peer will be equal to $\left|\frac{(N-1+\beta_1)}{(N-1)(N-1-k)+k}\right|$.

in $\bar{y}_{k_i}$ will be explained by the random component, u_i, rather than being governed by the mean of the pool of potential peers, $\bar{y}_{-i}$. This indicates that, for a given pool size, studies that consider larger peer groups will have a larger exclusion bias.

3 $\frac{\Delta|bias|}{\Delta\beta_1} > 0$, that is, the greater the true social interaction effect β_1, the larger the bias. Also, this result is intuitive, as the greater the true peer effect β_1, the more similar the outcomes y_i, and therefore the smaller their variance, σ_ε^2. From equation (3.3) we know that, if σ_ε^2 is small, σ_u^2 will be relatively small as well. Again, from equation (3.4), we know that, as σ_u^2 becomes smaller, less of the variation in $\bar{y}_{k_i}$ will be explained by the random component u_i and the more sensitive $\bar{y}_{k_i}$ will be to the exclusion of an individual's outcome.

A detailed discussion of the implications of these results for the interpretation of estimation results in the social interaction literature, in particular the effect of the pool size and peer group size, will follow in section 3.7.

3.3. Model allowing for cluster sampling

3.3.1. Motivation

In the basic model described in the previous section, we assumed that peers are randomly drawn from the pool of (N – 1) observations within that network. In practice, however, sampled observations are often clustered into groups within a network, and peers are often restricted to those clusters. Moreover, the researcher often relies on the addition of cluster fixed effects to control for unobserved correlated effects. For example, whereas a basic OLS estimator would consider the variation across all N individuals in a school, a model adding classroom dummies would allow us to consider

the variation across students *within* different classes of size $L < N$. This is done, for instance, to control for common characteristics of students within groups, such as a teacher's contribution to the student's test score.

In this section, I show how the exclusion bias differs between models with and without cluster fixed effects, while introducing the possibility that peers are selected from within a pool $l \prec \Omega$ (e.g. a classroom within the school). I show that, in some common empirical specifications, the downward exclusion bias will be more severe in OLS models that add cluster fixed effects and the bias will not decrease as the sample size tends to infinity. Specifically, the aggravation of the bias will occur whenever peer group formation is correlated with cluster formation. The intuition for this condition will be conveyed in the next subsection, in which I set out two different empirical scenarios. The analytical discussion in section 3.3.3 will be structured around these two different cases.

3.3.2. Two empirical scenarios

Case 1: Peers assigned at the network level Ω $(N_P = N)$

All N students in a school Ω are randomly assigned k other students in the school as reference group, and classes indexed by l of size $L < N$ are formed independently from this peer allocation process.[23] That is, peer group formation operates independently from cluster formation. Thus, peers are not restricted to the class level, as the pool of potential peers consists of all students in the school (excluding oneself). In the next section, I will show that, in this case, the exclusion bias will be the same in models with and without classroom dummies.

To my knowledge, there are no studies that randomly assign peers to students at the school level. However, several studies consider non-randomly selected reference groups that contain peers from outside the classroom. For instance, studies in which peer

[23] This case still assumes person-specific peer groups.

groups are defined by individuals identified as friends from the school roster (Fletcher and Ross, 2012; Halliday and Kwak, 2012) or by all other students in the same year group and school (e.g. Hoxby, 2000; Hanushek et al., 2003; Angrist and Lang, 2004; Fletcher, 2012). To the extent that peers are selected independently from the classroom in these studies, case 1 applies.

Case 2: Peers assigned at the cluster level l $(N_P = L)$

Now consider an alternative case in which all N students in a school are, first, randomly assigned to classes indexed by l of size $L < N$, and k peers are selected from *within* the class l. I will show that, in this case, in which peer group formation is correlated with cluster formation, the exclusion bias will be more severe in a model adding classroom fixed effects than in the basic OLS model. I will also show that the bias in OLS models excluding cluster effects will be larger than their equivalents in case 1 described above.

Note that case 2 allows for both (i) person-specific peer groups within the class, such that peer groups do not necessarily overlap and (ii) grouping of friends at the cluster level l. Whereas in (i) one would set $k < L-1$, in (ii) one would set $k = L-1$.[24] An example of (i) is de Melo (2011), in which each individual was asked to select friends from the class roster.[25] An example of (ii) is the study by Duflo et al. (2011), who randomly assigned students in non-tracking schools to one of two sections and who then defined peers by all students in the section excluding the student himself/herself.[26] Similar examples include the studies by Sacerdote (2001), Glaeser et al. (2002) and

[24] If grouping occurs at a level that is lower than the classroom, we would re-specify the model and denote l to be that lower level group (instead of the classroom). The same model would apply.
[25] Note that such person-specific peer groups may still overlap, but, for the purpose of simplicity, we assume here that this overlap occurs randomly.
[26] Although their randomisation strategy is in line with case 2, the study by Duflo et al. (2011) is not affected by the exclusion bias because of the empirical model they use (see section 3.7).

Zimmerman (2003), who randomly assigned students to rooms in dormitories and defined peers by roommates.[27]

3.3.3. Model

In this section, we extend the basic model by allowing for clustering of sampled observations into $\frac{N}{L}$ groups of equal size *L*, within the network Ω . For instance, this model allows for clustering of *N* students into $\frac{N}{L}$ classrooms within a school. In section 3.6 I will discuss the results of a specification in which I allow the size of groups to vary within the network. For simplicity, we assume that all individuals within the same group l use the same pool $\Pi_l \subset \Omega$, from which to (randomly) draw peers. Suppose that the size of the pool Π_l is the same for all groups l, that is, N_P. Following the two scenarios described in the previous section, this pool is either the entire network Ω (e.g. school), as in case 1, or a subset $l \subset \Omega$ (e.g. classroom), as in case 2.

The equivalent equations of (3.2) and (3.4) become:

$$y_{li} = \beta_0 + \beta_1 \bar{y}_{k_{li}} + \varepsilon_{li} \tag{3.10}$$

and

$$y_{li} = \beta_0 + \beta_1 \left(\bar{y}_{\Pi_l, -i} + u_{li} \right) + \varepsilon_{li} \tag{3.11}$$

where y_{li} denotes outcome *y* of individual *i* in group l; variable $\bar{y}_{\Pi_l, -i}$ denotes the average outcome of the pool of potential peers of all individuals in group l, excluding

[27] Although one would not add classroom fixed effects in (ii), one could add fixed effects at a level that is higher than the classroom, but lower than the school (such as grade fixed effects). It can easily be shown that such models form special cases of the model we consider in this section.

individual *i* herself. As before, assume that $E(u) = E(\varepsilon) = 0$, $\text{cov}(u, \varepsilon) = 0$, $\text{var}(u) = \sigma_u^2$ and $\text{var}(\varepsilon) = \sigma_\varepsilon^2$.[28]

Whereas, in the analytical discussion that follows, the models without cluster fixed effects will be based on equation (3.10), models including cluster fixed effects will be based on a transformation of this equation. Specifically, the observations of interest to the fixed effect estimator are the individual specific deviations of outcomes from their respective cluster average. Define $\ddot{x} = x_{li} - \bar{x}_l$ for $x = y, \bar{y}_k, \varepsilon$. Then, differencing equation (3.10) gives:

$$\ddot{y} = \beta_1 \ddot{y}_k + \ddot{\varepsilon} \qquad (3.12)$$

3.3.4. Exclusion bias in models adding cluster fixed effects

Using a similar procedure as before, in Appendix 3.A.3 I derive the cluster fixed effect equivalent of proposition 3.1, for a cluster sampling model. Proposition 3.2 summarises the results.

Proposition 3.2. Suppose each individual *i* in cluster *l* in a clustered network Ω of size *N* is randomly assigned *k* peers $\neq i$ from within a subset $\Pi_l \subset \Omega$ of size $N_P \leq N$. In a linear-in-means social interaction model adding cluster fixed effects, specified as in equation (3.12), the OLS estimate of the peer effect, $\hat{\beta}_1^{FE}$, will be downward biased according to the following expression:

$$E\left(\hat{\beta}_1^{FE}\right) = \beta_1 - \frac{(N_P - 1 + \beta_1)k}{(N_P - 1)(N_P - 1 - k) + k} < \beta_1$$

Note that proposition 3.2 is similar to proposition 3.1, except that now the magnitude of the bias depends on the size of the potential pool $N_P \leq N$, rather than on

[28] Note that in reality, in the context of cluster sampling, the assumption $E(\varepsilon) = 0$ will usually not be satisfied. Although the results are expected to hold in models allowing for correlated effects, in this chapter I assume zero correlated effects, to focus on the partial effect of the exclusion bias.

the sampled network size *N*. Importantly, proposition 3.2 indicates that, for models adding cluster fixed effects, the exclusion bias is not necessarily only a small sample property. With peer pool sizes fixed at $N_P < N$, the magnitude of the bias will not decrease as *N* tends to infinity.

In section 3.3.2 I described two different empirical scenarios: one in which peers are selected at the network level Ω ($N_P = N$) and one in which peers are drawn from within the cluster *l* of size $L < N$ ($N_P = L$). In what follows, I discuss how proposition 3.2 applies to each of these two special cases.

Case 1: Peers assigned at the network level Ω $(N_P = N)$

This case refers to the scenario in which all *N* students in a school are randomly assigned to other students in the school to become friends, and classes are formed independently from this peer allocation process. In this case, the pool of potential peers equals all students in the school (excluding oneself), with size N_P equal to *N* for all observations in the sample. Applying the result in proposition 3.2, we find that, in this case, the exclusion bias in the model adding cluster fixed effects will be identical to the basic case considered in section 3.2:

$$E\left(\hat{\beta}_1^{FE}\right) = \beta_1 - \frac{(N-1+\beta_1)k}{(N-1)(N-1-k)+k} < \beta_1 \qquad (3.13)$$

As before in the basic model, since the pool of potential peers coincides with the sampled network Ω, the bias will disappear as the sample size *N* tends to infinity.

Case 2: Peers assigned at the cluster level *l* $(N_P = L)$

This case reflects the scenario in which all *N* students in a school are, first, randomly allocated to $\left(\frac{N}{L}\right)$ classes of size *L*. Within each class, students are then randomly assigned to other students in the class to become friends. In this case, peers are

restricted to the class. This means that N_P equals L for all observations in the sample. Again using proposition 3.2, we find that, in this case, the exclusion bias in the model adding cluster fixed effects becomes:

$$E\left(\hat{\beta}_1^{FE}\right) = \beta_1 - \frac{(L-1+\beta_1)k}{(L-1)(L-1-k)+k} < \beta_1 \qquad (3.14)$$

In this case, in which the pool of potential peers is a sub-set l of the sampled network Ω, the bias depends on the size L of the sub-set l. With fixed L, the bias will not decrease as N tends to infinity.

3.3.5. Exclusion bias in models omitting cluster fixed effects

Case 1: Peers assigned at the network level Ω $(N_P = N)$

When the pool from which an individual's peers are drawn equals the total network Ω of N individuals (excluding the individual herself), and when one ignores cluster fixed effects, then it is clear that the results reduce to the basic model we considered in section 3.2. We obtain proposition 3.3, which is very similar to proposition 3.1, except that we now consider a clustered network.

Proposition 3.3. Suppose each individual i in cluster l in a clustered network Ω of size N is randomly assigned k peers $\neq i$ from within the network Ω. In a linear-in-means social interaction model omitting cluster fixed effects, specified as in equation (3.10), the OLS estimate of the peer effect, $\hat{\beta}_1^{OLS}$, will be downward biased according to the following expression:

$$E\left(\hat{\beta}_1^{OLS}\right) = \beta_1 - \frac{(N-1+\beta_1)k}{(N-1)(N-k-1)+k} < \beta_1$$

Case 2: Peers assigned at the cluster level l $(N_P = L)$

The results are very different, however, when peers are assigned at the level $l \prec \Omega^N$. Within the framework of hierarchical models, in which data are clustered at a sublevel smaller than the total sample size N, it has been shown that the basic OLS estimator

$\hat{\beta}_1^{OLS}$ (i.e. OLS in a model excluding cluster fixed effects) is a weighted combination of the cluster fixed effects estimator $\hat{\beta}_1^{FE}$ (or *within* estimator) and the *between* estimator $\hat{\beta}_1^{BE}$ (Raudenbush and Bryk, 2002):

$$\hat{\beta}_1^{OLS} = \eta^2 \hat{\beta}_1^{BE} + (1-\eta^2)\hat{\beta}_1^{FE} \tag{3.15}$$

where η^2 is the ratio of the between sum of squares of the independent variable of interest, $\bar{y}_{k_i}$, to its total sum of squares. The within estimator $\hat{\beta}_1^{FE}$ is that presented in equation (3.14). The between estimator $\hat{\beta}_1^{BE}$ is the OLS estimator from a regression of $\bar{y}_l$ on an intercept and $\bar{\bar{y}}_{k_l}$, where $\bar{y}_l$ denotes the average outcome y_{li} over the individuals in the group and $\bar{\bar{y}}_{k_l}$ denotes the average peer group outcome of the group:

$$\bar{y}_l = \beta_0 + \beta_1 \bar{\bar{y}}_{k_l} + \bar{\varepsilon}_l \tag{3.16}$$

In Appendix 3.A.4 I show that, in the limit, the between estimator $\hat{\beta}_1^{BE}$ will approach:

$$p\lim\left(\hat{\beta}_1^{BE}\right) = \beta_1 + \frac{(L-1)k}{(L-1)k + (L-1-k)} > \beta_1 \tag{3.17}$$

which, in contrast to $\hat{\beta}_1^{BE}$, is *larger* than the true value β_1.

Combining equations (3.14), (3.15) and (3.17) and the results on η^2 derived in Appendix 3.A.5, we arrive at proposition 3.4.

Proposition 3.4. Suppose each individual *i* in a clustered network Ω is randomly assigned *k* peers $\neq$ *i* from within their cluster *l* of size *L*. In a linear-in-means social interaction model ignoring cluster fixed effects, as specified in equation (3.10), the OLS estimate of the peer effect, $\hat{\beta}_1^{OLS}$, will be downward biased and inconsistent according to the following expression:

$$p\lim\left(\hat{\beta}_1^{OLS}\right) = p\lim\left(\eta^2\right)p\lim\left(\hat{\beta}_1^{BE}\right) + \left[1 - p\lim\left(\eta^2\right)\right]p\lim\left(\hat{\beta}_1^{FE}\right)$$

where:

$$p\lim\left(\hat{\beta}_1^{BE}\right) = \beta_1 + \frac{(L-1)k}{(L-1)k + (L-1-k)} > \beta_1$$

$$p\lim\left(\hat{\beta}_1^{FE}\right) = \beta_1 - \frac{(L-1+\beta_1)k}{(L-1)(L-1-k)+k} < \beta_1$$

$$0 < p\lim\left(\eta^2\right) = \frac{\dfrac{(Lk-2k+L-1)}{k(L-1)}}{\dfrac{(Lk-2k+L-1)}{k(L-1)} + \left[\dfrac{(L-1)^2(L-1-k)+(L-1)k}{k(L-1+\beta_1)^2}\right]} < 1$$

Given that $0 < p\lim\left(\eta^2\right) < 1$, we know that:

$$p\lim\left(\hat{\beta}_1^{FE}\right) < p\lim\left(\hat{\beta}_1^{OLS}\right) < \beta_1 < p\lim\left(\hat{\beta}_1^{BE}\right)$$

Note that the result summarised in proposition 3.4 is only valid in the limit, that is, as the sample size *N* tends to infinity. In Appendix 3.A.5 I explain why the downward bias in $\hat{\beta}_1^{OLS}$ is expected to be more severe the smaller the sample of observations that is available. An analytical derivation of the magnitude of the exclusion bias, when using small samples, in OLS models omitting cluster fixed effects, is beyond the scope of this thesis (see Appendix 3.A.5). However, in section 3.5 I will use simulations to show that, in small samples, we also have $\hat{\beta}_1^{FE} < \hat{\beta}_1^{OLS} < \beta_1$.

3.3.6. Comparison results of models with and without cluster fixed effects

Comparing the results provided by propositions 3.1–3.4, I conclude this section with a fifth proposition, in which $\left|bias^{FE}\right|$ denotes the absolute value of the magnitude of the exclusion bias in a model adding cluster fixed effects, and $\left|bias^{OLS}\right|$ denotes the absolute value of the magnitude of the exclusion bias in a model ignoring fixed effects.

Proposition 3.5.

When peer group formation occurs at the total network level Ω (e.g. case 1):

$$\left|bias^{FE}\right| = \left|bias^{OLS}\right|$$

When peer group formation occurs at the cluster level $l \prec \Omega$ (e.g. case 2):

$$\left|bias^{FE}\right| > \left|bias^{OLS}\right|$$

The intuition behind proposition 3.5 is the following. Recall that the exclusion bias arises if (i) an individual is excluded from her own potential peer group and (ii) if that individual is a potential peer of other individuals in the set within which one analyses the variation in outcomes. The second part of this condition is important: the exclusion bias arises because of the presence of observations for which individual *i* is included in her pool of potential peers. The smaller the share of such 'problematic' observations in the set of individuals across which one analyses the variation in outcomes, the less severe the exclusion bias will be. Providing peers are selected only from *within* an individual *i*'s cluster (case 2), then the potential peers of the individuals *outside i*'s cluster will never include *i.* Given that, in models adding cluster fixed effects, the researcher is interested only in the variation of outcomes *within* clusters, the presence of such 'unproblematic' observations outside the cluster will not reduce the exclusion bias. In contrast, in models without cluster fixed effects, the researcher is interested in the

variation of outcomes across *all* individuals in the sample, which means that the exclusion bias will become relatively less severe.

An important remark is in order when comparing proposition 3.2 with proposition 3.4. Whereas in models adding cluster fixed effects the exclusion bias does *not* decrease with sample size N (see proposition 3.2), proposition 3.4 shows that, in models without cluster fixed effects, the exclusion bias *does* decrease as N tends to infinity.[29] Intuitively, with fixed cluster sizes and with peers selected at the cluster, an increase in the sample size N implies a larger number of 'unproblematic observations' in the sample. I have discussed above why this leads to a reduction in the magnitude of the exclusion bias in models without cluster fixed effects, whereas models adding cluster fixed effects remain unaffected by this.

The results summarised in proposition 3.5 are important. They caution researchers against naive comparisons of OLS peer effects estimation results between models with and those without cluster fixed effects. A more detailed discussion of the implications of these results for the social interaction literature follows in section 3.7.

3.4. Solutions to the bias

3.4.1. Method suggested by GKN (2009)

As discussed above, the exclusion bias is driven by the fact that an individual cannot be her own peer and therefore is excluded from the pool of individuals from which her peers are drawn. This empirical construction results in a correlation between the mean of the pool of potential peers, $\bar{y}_{-i}$, and the characteristics of the individual. I have shown above how this causes the OLS coefficient estimate of the endogenous peer effect to be biased downwards in empirical social interaction models.

[29] Formally, this result can be derived by considering the result in Appendix 3.A.5 that, when N tends to infinity, more weight is given to the between estimator in the computation of $\hat{\beta}_1^{OLS}$, and, therefore, $\hat{\beta}_1^{OLS}$ becomes less negative.

In a simple model of randomised peer group formation, GKN (2009) deals with the exclusion bias by controlling for the average outcome of the pool of potential peers, $\bar{y}_{-i}$, in the above regression (equation (3.2)). To highlight how, under specific conditions, this effectively deals with the exclusion bias, consider a model similar to equation (3.2), but now adding the term $\bar{y}_{-i}$. Substituting in for equation (3.3) and rearranging, I obtain:

$$y_i = \beta_0 + \beta_1 \bar{y}_{k_i} + \beta_2 \bar{y}_{-i} + \varepsilon_i \tag{3.18}$$

$$= \beta_0 + \beta_1 (\bar{y}_{-i} + u_i) + \beta_2 \bar{y}_{-i} + \varepsilon_i$$

$$= \beta_0 + (\beta_1 + \beta_2)\bar{y}_{-i} + \beta_1 u_i + \varepsilon_i \tag{3.19}$$

The inclusion of $\bar{y}_{-i}$ in the model acts as a proxy, soaking up the effect of the non-random component of $\bar{y}_{k_i}$. As a result, the coefficient estimate $\hat{\beta}_1$ measures the partial effect of the random component u_i. Since $E(u_i \varepsilon_i) = 0$ under the assumption of random peer selection, $p\lim(\hat{\beta}_1) = \beta_1$, and OLS will yield consistent estimates of the social interaction effect β_1.

However, there are a few limitations to the method suggested by GKN (2009). If the pool of potential peers is identical for each sampled individual, excluding *i* (such as in case 1 described above), then all variation in $\bar{y}_{-i}$ will be driven by y_i. Thus, β_1 will not be identified. If the pool of potential peers is identical for individuals sharing the same cluster, a similar problem occurs whenever cluster fixed effects are added to the model and β_1 will not be identified. To formalise the latter result, reconsider the model in equation (3.12) but now we add the difference between $\bar{y}_{-i}$ and its average over the cluster group, $\bar{\bar{y}}_{-i,l}$:

$$(y_{li} - \bar{y}_l) = \beta_1(\bar{y}_{k_i} - \bar{\bar{y}}_{k_l}) + \beta_2(\bar{y}_{-i} - \bar{\bar{y}}_{-i,l}) + (\varepsilon_{li} - \bar{\varepsilon}_l)$$

$$= \beta_1(\bar{y}_{k_i} - \bar{\bar{y}}_{k_l}) + \beta_2\left(\left(\frac{\sum_{j=1}^{L} y_{lj} - y_{li}}{L-1}\right) - \frac{1}{L}\sum_{i=1}^{L}\left(\frac{\sum_{j=1}^{L} y_{lj} - y_{li}}{L-1}\right)\right) + (\varepsilon_{li} - \bar{\varepsilon}_l)$$

$$= \beta_1(\bar{y}_{k_i} - \bar{\bar{y}}_{k_l}) + \beta_2\left(\left(\frac{\sum_{j=1}^{L} y_{lj} - y_{li}}{L-1}\right) - \frac{\sum_{j=1}^{L} y_{lj} - \bar{y}_l}{L-1}\right) + (\varepsilon_{li} - \bar{\varepsilon}_l)$$

$$= \beta_1(\bar{y}_{k_i} - \bar{\bar{y}}_{k_l}) + \beta_2\left(\left(\frac{\sum_{j=1}^{L} y_{lj} - y_{li}}{L-1}\right) - \frac{\sum_{j=1}^{L} y_{lj}}{L}\right) + (\varepsilon_{li} - \bar{\varepsilon}_l)$$

$$= \beta_1(\bar{y}_{k_i} - \bar{\bar{y}}_{k_l}) + \beta_2\left(\frac{y_{li} - \bar{y}_l}{1-L}\right) + (\varepsilon_{li} - \bar{\varepsilon}_l)$$

These derivations show that the term $(\bar{y}_{-i} - \bar{\bar{y}}_{-i,l})$ (the within-model equivalent of the term suggested by GKN (2009)) equals $\left(\frac{y_{li} - \bar{y}_l}{1-L}\right)$, when peers are selected from within the cluster. Thus, all the variation in $(y_{li} - \bar{y}_l)$ will be captured by $\left(\frac{y_{li} - \bar{y}_l}{1-L}\right)$ and β_1 will not be identified.

Lastly, even if these issues are not a concern, the GKN (2009) method requires perfect knowledge of the pool from which peers are selected. However, as I will discuss in section 3.6, in practice, this is not always the case.

3.4.2. 2SLS estimation strategies

The social interaction literature frequently relies on 2SLS strategies to address other empirical challenges in the identification of peer effects, such as the reflection problem and correlated effects (for a discussion of these issues, see Manski 1993; Brock and Durlauf, 2001; Moffitt, 2001). For instance, various studies have instrumented the average peer group outcome by the proportion of peers that are randomly treated by an

intervention that affects their outcomes (e.g. Duflo and Saez, 2002; Lalive and Cattaneo, 2009). Other studies have used a set of average exogenous characteristics of the peers of a person's peers as identifying instruments, while controlling for the individual's own exogenous characteristics (Bramouille et al., 2009; De Giorgi et al., 2010; de Melo, 2011; Helmers and Patnam, 2012). Although these studies use 2SLS to address traditional challenges to identification, I will show that it also overcomes the exclusion bias.

To see this, let $\bar{z}_{k_i}$ be the peer group average of a characteristic *z*, and assume that it is considered as an instrument for the average peer group outcome $\bar{y}_{k_i}$. If $\bar{z}_{k_i}$ is informative in the first stage of a 2SLS estimation strategy, then z_i will usually also be informative about y_i, and therefore it is often included in the second stage of 2SLS.[30] The first and second stages of such SLS estimation strategies can be modelled as follows:

First stage: $\bar{y}_{k_i} = \pi_0 + \pi_1 \bar{z}_{k_i} + \pi_2 z_i + v_i$

Second stage: $y_i = \beta_0 + \beta_1 \hat{\bar{y}}_{k_i} + \beta_2 z_i + \varepsilon_i$ (3.20)

where $E(z_i \varepsilon_i) = 0$, $E(\varepsilon_i) = 0$ and $\hat{\bar{y}}_{k_i} = \hat{\pi}_0 + \hat{\pi}_1 \bar{z}_{k_i} + \hat{\pi}_2 z_i$ is the fitted value of the first-stage regression.

Expanding the second-stage 2SLS equation (equation (3.20)), using the expression for the fitted values, it is straightforward to see that $\text{cov}(\hat{\bar{y}}_{k_i}, \varepsilon_i \mid z_i) = 0$ and therefore $\hat{\beta}_1^{2SLS}$ does not suffer from the exclusion bias:

[30] Note that it is not necessary to control for *z* whenever *z* is purely random. This is the case, for instance, when an intervention is randomised only for peers and not for oneself. If all individuals in the sample are subject to the randomisation, however, it is essential that *z* is controlled for.

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 \hat{\bar{y}}_{k_i} + \beta_2 z_i + \varepsilon_i \\
&= \beta_0 + \beta_1 \left(\hat{\pi}_0 + \hat{\pi}_1 \bar{z}_{k_i} + \hat{\pi}_2 z_i \right) + \beta_2 z_i + \varepsilon_i \\
&= \beta_0 + \beta_1 \left[\hat{\pi}_0 + \hat{\pi}_1 \left(\frac{\sum_{j=1}^{N_P} z_j - z_i}{N_P - 1} + \tilde{u}_i \right) + \hat{\pi}_2 z_i \right] + \beta_2 z_i + \varepsilon_i \\
&= \beta_0 + \beta_1 \left(\hat{\pi}_0 + \frac{\hat{\pi}_1 \sum_{j=1}^{N_P} z_j}{N_P - 1} + \hat{\pi}_1 \tilde{u}_i \right) + \left(\beta_1 \hat{\pi}_2 + \beta_2 \frac{-\hat{\pi}_1}{N_P - 1} \right) z_i + \varepsilon_i
\end{aligned}
\qquad (3.21)
$$

where $\tilde{u}$ is the equivalent of *u* but now for *z* instead of *y* (see (3.3)). Note that, if y_i and z_i are correlated (i.e. if $\beta_2 \neq 0$), we would expect $\bar{z}_{k_i}$ to also be mechanically correlated with y_i, through $\bar{z}_{k_i} = \frac{\sum_{j=1}^{N_P} z_j - z_i}{N_P - 1} + \tilde{u}_i$. However, the expression derived in equation (3.21) shows how controlling for z_i in the regression prevents this mechanical relationship from existing. The presence of z_i in the model captures the effect that the exclusion of z_i has on the computation of the average peer group characteristic $\bar{z}_{k_i}$, allowing $\hat{\beta}_1^{2SLS}$ to be an unbiased estimate of the peer effect.

3.5. Simulation results

To illustrate the presence and the consequences of the exclusion bias, this section reports on the results of various Monte Carlo simulations. The outcomes of the simulations confirm each of the theoretical predictions derived earlier in this chapter. Specifically, they confirm the predicted magnitude of the bias for various values of the network size *N*, the cluster size *L* and the peer group size *k*. They confirm the difference in results between models with and without cluster fixed effects whenever peers are

assigned at the cluster level. Finally, they confirm the ability of 2SLS strategies to eliminate the bias.

Although the results are also confirmed for simulations with $\beta_1 > 0$, for simplicity I focus here on the case of zero true peer effects, that is, $\beta_1 = 0$. In each simulation, I generate a dataset of *N* observations and randomly assign a normally distributed characteristic *y* to each of them. The underlying data generation process (DGP) for the first set of Monte Carlo simulations is the following:

$$y_{li} = 10 + \varepsilon_{li},\ \varepsilon_{li} \sim N(0,1) \tag{3.22}$$

For the simulations based on the cluster sampling models, I then randomly assign the *N* individuals to $\frac{N}{L}$ clusters of *L* persons each. For each individual in the sampled network, I randomly select *k* reference persons out of a specified pool $\Pi_l \subset \Omega$ of potential peers. Each simulation differs in the value assigned to the network size *N*, the cluster size *L* and the peer group size *k*. The simulations also differ in the definition of the pool Π_l. This is either the total network Ω $(N_P = N)$ or the cluster $l \prec \Omega$ $(N_P = L)$.

Once the dataset is compiled, I run regressions, both with and without cluster fixed effects, of the outcome y_{li} on $\bar{y}_{k_i}$:

$$y_{li} = \beta_0 + \beta_1 \bar{y}_{k_i} + \varepsilon_{li} \tag{3.23}$$

where y_{li} is the outcome of individual *i* in cluster *l* and $\bar{y}_{k_i}$ is the average outcome of that individual's peer group. As noted above, the DGP is based on a true $\beta_1 = 0$, thus an unbiased estimator would be expected to yield $\hat{\beta}_1 = 0$. In each simulation, I store the coefficient estimate $\hat{\beta}_1$ and repeat this procedure 10,000 times. Tables 3.1–3.4 show the simulation results for the estimate of the endogenous peer effect, $\hat{\beta}_1$, where (S)

Table 3.1 Simulation results $\hat{\beta}_1$: basic model

		$N_P = N$		
		$N = 50$	$N = 100$	$N = 200$
		OLS	OLS	OLS
		(1)	(2)	(3)
$k = 1$	T	–0.021	–0.010	–0.005
	S	–0.019	–0.009	–0.006
		(0.001)	(0.001)	(0.001)
	% rejection 0.05	4.7	4.7	4.9
	% rejection 0.10	9.6	10.1	9.7
$k = 5$	T	–0.113	–0.053	–0.026
	S	–0.113	–0.049	–0.028
		(0.003)	(0.002)	(0.002)
	% rejection 0.05	6.1	5.5	5.5
	% rejection 0.10	11.8	10.5	10.7
$k = 10$	T	–0.255	–0.112	–0.053
	S	–0.257	–0.105	–0.055
		(0.005)	(0.003)	(0.002)
	% rejection 0.05	8.2	6.1	5.7
	% rejection 0.10	0.140	0.116	0.111
$k = 15$	T	–0.437	–0.178	–0.082
	S	–0.442	–0.171	–0.083
		(0.007)	(0.004)	(0.003)
	% rejection 0.05	9.7	6.8	5.7
	% rejection 0.10	16.6	12.4	11.2

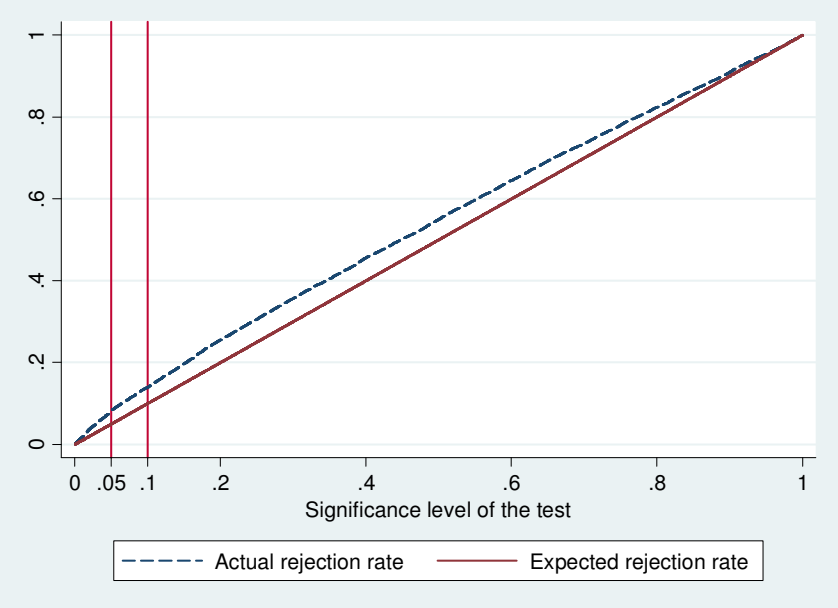
Notes: No cluster sampling – peer selection at the level Ω. Data generation process, $y_{li} = 10 + \varepsilon_{li}$, $\varepsilon_{li} \sim N(0,1)$. Standard deviations of $\hat{\beta}_1$ are shown in parentheses. OLS, ordinary least squares; S, estimated $\hat{\beta}_1$ over 10,000 Monte Carlo repetitions; T, coefficient estimate predicted by proposition 3.1. % rejection 0.05, percentage of Monte Carlo repetitions for which we reject $H_0 : \beta_1 = 0$ at significance level 0.05; % rejection 0.05, percentage of Monte Carlo repetitions for which we reject $H_0 : \beta_1 = 0$ at significance level 0.10.

represents the simulation result and (T) indicates what the theory outlined above would predict given the respective values of *k*, *N* and *L*. The tables also indicate, for each simulation, the proportion of repeated simulation rounds (10,000 in total) for which we reject the hypothesis that $\beta_1 = 0$, at the 5% and 10% significance levels.

Table 3.1 presents the results of simulations predicated on the basic model presented in section 3.2, with an underlying DGP that ignores clustering at a level lower than Ω. That is, all *N* sampled students belong to the same school but are not grouped into classes. The *k* peers are randomly assigned at the school level, $N_p = N$, where we vary

the values of *k* and *N*. The analytical expression of the exclusion bias that is relevant here is provided in proposition 3.1. Note (by comparing (T) with (S)) that each of the simulations yields results for the magnitude of the bias that are very similar to what proposition 3.1 would predict. As expected in this model, for a given peer group size, the exclusion bias becomes less severe as the sample size grows. Moreover, with the sample size fixed, the magnitude of the bias increases as the peer group grows in size.

Figure 3.1 Rejection rates $H_0 : \beta_1 = 0$ (basic model OLS: $N = 50$; $k = 10$)



Note, however, that the rejection rates reported in Table 3.1 are relatively low. To interpret these numbers, consider Figure 3.1, which plots the rates at which we reject the test $H_0 : \beta_1 = 0$ in simulation $N = 50$ and $k = 10$, for various significance levels. Given that the significance level states the tolerated probability of rejecting H_0 when it is in fact true, the 45° line shows us the rejection rates that we would expect if the estimator was unbiased. Actual rejection rates that are higher than the 45° line are an indication that the estimator is biased. Figure 3.1 indicates that, for $N = 50$ and $k = 10$, the estimate is slightly biased: whereas, at the 5% significance level, we would expect

to reject H_0 5% of the time, in our simulations we reject it approximately 8% of the time (out of 10,000 repetitions). However, although the magnitude of the bias appears to be quite significant in most of the specifications in Table 3.1 (e.g. for $N = 50$ and $k = 15$, we obtain a coefficient estimate of the endogenous peer effect of –0.442, which is equivalent to a bias of 3 percentage points per peer), the rejection rates are relatively low. This is not surprising, given the small sample sizes and the associated high standard errors of the estimates.

Table 3.2 Simulation results $\hat{\beta}_1$: case 1 ($N_P = N = 200$)

		$L = 20$		$L = 50$	
		FE	OLS	FE	OLS
		(1)	(2)	(3)	(4)
$k = 1$	T	–0.005	–0.005	–0.005	–0.005
	S	–0.004	–0.004	–0.004	–0.004
		(0.001)	(0.001)	(0.001)	(0.001)
	% rejection 0.05	5.6	5.1	5.1	5.1
	% rejection 0.10	10.8	9.9	10.1	9.9
$k = 5$	T	–0.026	–0.026	–0.026	–0.026
	S	–0.023	–0.024	–0.023	–0.024
		(0.006)	(0.006)	(0.002)	(0.002)
	% rejection 0.05	5.4	4.9	5.0	4.9
	% rejection 0.10	10.8	10.0	10.5	10.0
$k = 10$	T	–0.053	–0.053	–0.053	–0.053
	S	–0.049	–0.050	–0.050	–0.050
		(0.002)	(0.002)	(0.002)	(0.002)
	% rejection 0.05	6.0	5.6	5.7	5.6
	% rejection 0.10	11.2	10.4	10.7	10.4
$k = 15$	T	–0.082	–0.082	–0.082	–0.082
	S	–0.078	–0.079	–0.078	–0.079
		(0.003)	(0.003)	(0.003)	(0.003)
	% rejection 0.05	6.3	5.8	6.0	5.8
	% rejection 0.10	11.7	11.2	11.5	11.2

Notes: cluster sampling level $l \prec \Omega$ – peer selection level Ω. Data generation process, $y_{li} = 10 + \varepsilon_{li}$, $\varepsilon_{li} \sim N(0,1)$. Standard deviations of $\hat{\beta}_1$ are shown in parentheses. FE, OLS regressions adding cluster fixed effects; OLS, OLS regressions ignoring cluster fixed effects; S, estimated $\hat{\beta}_1$ over 10,000 Monte Carlo repetitions; T, coefficient estimate predicted by proposition 3.2 and proposition 3.3. % rejection 0.05, percentage of Monte Carlo repetitions for which we reject $H_0 : \beta_1 = 0$ at significance level 0.05; % rejection 0.05, percentage of Monte Carlo repetitions for which we reject $H_0 : \beta_1 = 0$ at significance level 0.10.

Tables 3.2 and 3.3 report on the results of simulations in which N is fixed at 200, but in which now the N individuals are randomly allocated to groups of size L. We present the results for $L = 20$ and $L = 50$. Table 3.2 is based on a DGP reflecting case 1, in which k peers are randomly assigned at the school level, independently from the classroom formation process. Again, the simulation results for both models with cluster fixed effects (FE) and without cluster fixed effects (OLS) are consistent with the values predicted by the analytical expressions in proposition 3.2 and proposition 3.3, respectively. Specifically, the magnitude of the bias depends only on the network size N and not on the cluster size L; hence, the similarities between the results presented in Table 3.2 and the results presented in column (3) in Table 3.1.

Table 3.3 shows the results of simulations with an underlying DGP mimicking case 2, in which the k peers are randomly assigned from within the cluster l. From section 3.3.5 we know that, in the context of cluster sampling, the OLS estimate of the peer effect in a model omitting cluster fixed effects will be a weighted average of the between estimator (BE) and the cluster fixed effect estimator (FE). The analytical expressions for the between estimator and the fixed effects estimator were presented in equation (3.14) and equation (3.17), respectively. Recall from the discussion of proposition 3.4 that for small samples, such as those we use in our simulations, it is computationally difficult to analytically derive predicted values for the coefficient estimate in OLS models omitting cluster fixed effects, that is, $\hat{\beta}_1^{OLS}$. Hence, those predicted values are indicated as 'not applicable' (N/A) in Table 3.3. However, the simulation results confirm the prediction made in section 3.3.6 that $\hat{\beta}_1^{OLS}$ always lies somewhere in between the between estimator and the fixed effect estimator results.

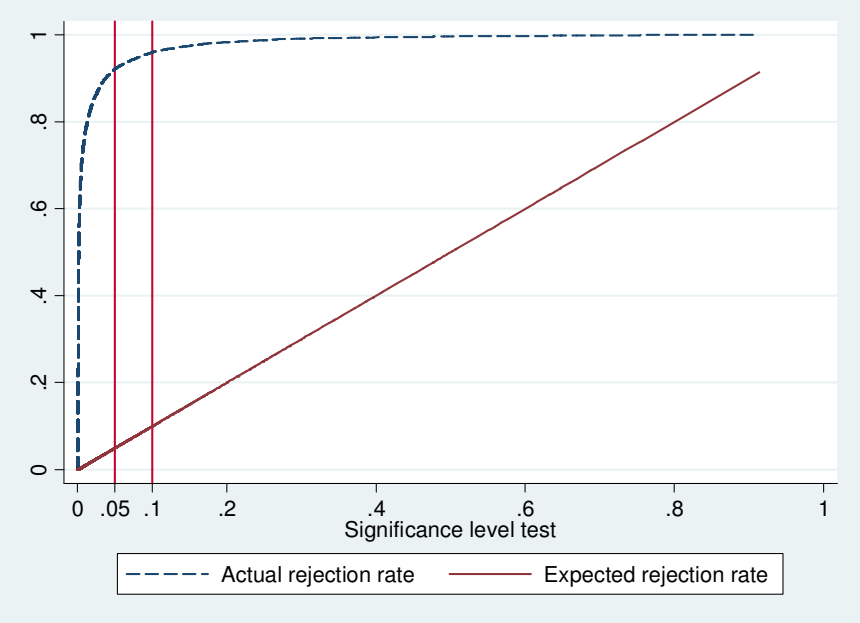
Table 3.3 Simulation results $\hat{\beta}_1$: case 2 ($N = 200$; $N_P = L$)

		$L = 20$			$L = 50$		
		BE	FE	OLS	BE	FE	OLS
		(1)	(2)	(3)	(4)	(5)	(6)
$k = 1$	T	0.514	–0.055	N/A	0.505	–0.021	N/A
	S	0.515	–0.056	–0.006	0.505	–0.021	–0.005
		(0.002)	(0.001)	(0.001)	(0.005)	(0.001)	(0.001)
	% rejection 0.05		13.0	6.2		6.0	5.1
	% rejection 0.10		20.6	11.7		11.3	10.2
$k = 5$	T	0.872	–0.351	N/A	0.848	–0.113	N/A
	S	0.870	–0.351	–0.036	0.852	–0.110	–0.026
		(0.001)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)
	% rejection 0.05		49.0	8.4		10.0	5.9
	% rejection 0.10		61.2	14.8		17.0	11.2
$k = 10$	T	0.955	–1.050	N/A	0.926	–0.255	N/A
	S	0.955	–1.049	–0.099	0.925	–0.251	–0.065
		(0.001)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
	% rejection 0.05		92.3	12.3		17.8	7.2
	% rejection 0.10		0.961	0.193		0.271	0.134
$k = 15$	T	0.986	–3.132	N/A	0.986	–3.132	N/A
	S	0.986	–3.145	–0.212	0.955	–0.432	–0.114
		(0.001)	(0.004)	(0.005)	(0.002)	(0.003)	(0.003)
	% rejection 0.05		100	17.8		26.5	8.8
	% rejection 0.10		100	25.8		37.9	15.5

Notes: cluster sampling level $l \prec \Omega$ – peer selection level *l*. Data generation process, $y_{li} = 10 + \varepsilon_{li}$, $\varepsilon_{li} \sim N(0,1)$. Standard deviations of $\hat{\beta}_1$ are shown in parentheses. BE, between-group estimation model; FE, OLS regressions adding cluster fixed effects (within group estimation model); N/A: not applicable; OLS, OLS regressions ignoring cluster fixed effects; S, estimated $\hat{\beta}_1$ over 10,000 Monte Carlo repetitions; T, coefficient estimate predicted by equation (3.14), equation (3.17) and proposition 3.4. % rejection 0.05, percentage of Monte Carlo repetitions for which we reject $H_0 : \beta_1 = 0$ at significance level 0.05; % rejection 0.05, percentage of Monte Carlo repetitions for which we reject $H_0 : \beta_1 = 0$ at significance level 0.10.

Importantly, note the high rates at which we reject the null hypothesis that $\beta_1 = 0$ at the 0.05 and 0.10 significance levels, especially in the models adding cluster fixed effects (FE). Figure 3.2 shows the rejection rates for the cluster fixed effects estimation results for $N = 200$, $L = 20$ and $k = 10$.

Figure 3.2 Rejection rates $H_0 : \beta_1 = 0$ (case 2 – FE: $N = 200$; $L = 20$; $k = 10$)



Finally, in order to test the hypothesis that 2SLS is effective in eliminating the exclusion bias, I generate an alternative DGP that is guided by the discussion in section 3.4.2:

$$y_{li} = 10 + 5x_{li} + \varepsilon_{li},\ \varepsilon_{li} \sim N(0,1) \qquad (3.24)$$

where x_{li} is a random uniformly distributed variable. Based on this DGP, I carry out simulations (again with 10,000 repetitions) in which each time I run 2SLS regressions on:

$$y_{li} = \beta_1 + \beta_2 \bar{y}_{k_i} + x_{li} + \varepsilon_{li}$$

instrumenting $\bar{y}_{k_i}$ by the average x characteristic of the peer group, $\bar{x}_{k_i}$.

Table 3.4 displays the 2SLS equivalent estimation results for each of the cases considered in Tables 3.1–3.3. Since the results are similar for all values of k, the table only reports on the results for $k = 10$. The results are consistent with our analytical results in section 5.2, that is, the exclusion bias is eliminated in 2SLS.

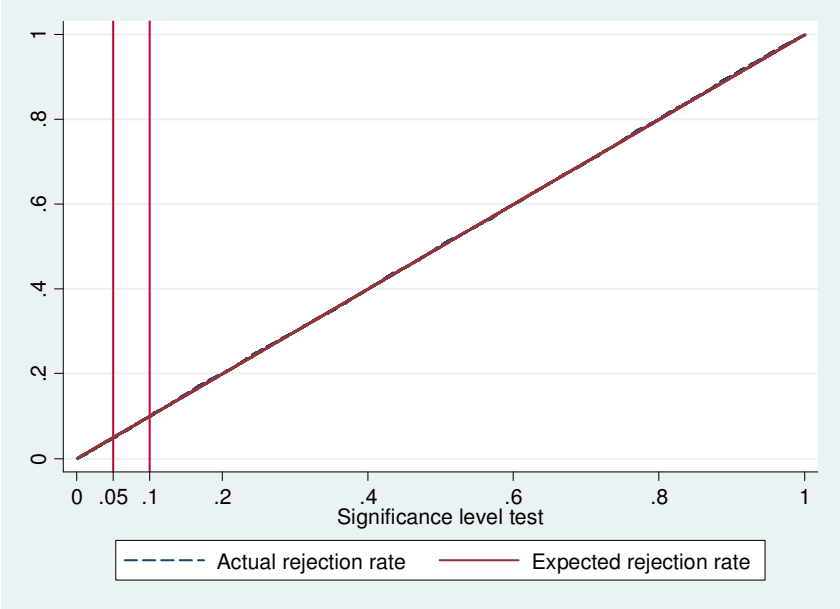
Table 3.4 Simulation results on $\hat{\beta}_1$: 2SLS

		Base model	Case 1: $N = 200$; $N_P = N$				Case 2: $N = 200$; $N_P = L$			
		$N = 200$	$L = 20$		$L = 50$		$L = 20$		$L = 50$	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$k = 10$	T	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	S	0.002	0.001	0.001	0.001	0.002	0.002	–0.011	0.003	–0.001
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
	% rejection 0.05	4.9	5.6	4.9	5.1	4.9	5.9	4.7	5.5	4.8
	% rejection 0.10	9.5	10.9	10.0	10.4	10.0	11.0	9.5	10.4	10.1
	Cluster FE	No	Yes	No	Yes	No	Yes	No	Yes	No

Notes: Data generation process, $y_{li} = 10 + 5x_{li} + \varepsilon_{li}$, $\varepsilon_{li} \sim N(0,1)$. Standard deviations of $\hat{\beta}_1$ are shown in parentheses. FE, fixed effects; T, coefficient estimate predicted by theory outlined in section 3.4.2; S, estimated $\hat{\beta}_1$ over 10,000 Monte Carlo repetitions; % rejection 0.05, percentage of Monte Carlo repetitions for which we reject $H_0 : \beta_1 = 0$ at significance level 0.05; % rejection 0.05, percentage of Monte Carlo repetitions for which we reject $H_0 : \beta_1 = 0$ at significance level 0.10.

Figure 3.3 shows the test rejection rates in case 2 ($N = 200$; $L = 20$; $k = 10$), now based on the 2SLS estimate of the endogenous peer effect β_1. The actual rejection rates now closely follow the expected rejection rates (i.e. 45 degree line), suggesting that the 2SLS estimate of the endogenous peer effect is, indeed, unbiased.

Figure 3.3 Rejection rates $H_0 : \beta_1 = 0$ (case 2 – 2SLS: $N = 200$; $L = 20$; $k = 10$)



3.6. Extensions of the basic model

In practice, determining the magnitude of the exclusion bias will not always be as straightforward as in the basic models considered above. Until now, we have assumed that all peers are randomly drawn from within a well-specified pool of individuals, of fixed size N_P. In this section, I briefly discuss some possible extensions of the standard model and anticipate how these would alter the main results.

First, in the basic cluster sampling models considered in section 3.3, I assumed all clusters (e.g. classrooms) were of equal size *L*. In reality, however, groups often vary in size. Allowing for varying group sizes L_l, in Appendix 3.A.5, I show how the analytical expression for the magnitude of the bias in a cluster fixed effects model (with peers selected at the cluster level) becomes:

$$|Bias| = \frac{\left[\sum_{l=1}^{\frac{N}{L}} \frac{L_l - 1}{(L_l - 1 + \beta_1)L_l} \times L_l\right]}{\left[\sum_{l=1}^{\frac{N}{L}} \frac{(L_l - 1)^2 \frac{(L_l - 1 - k)}{(L_l - 1)k} + 1}{(L_l - 1 + \beta_1)^2} \times L_l\right]} \tag{3.25}$$

Intuitively, expression (3.25) is the result of assigning different weights to the sampled observations, in which the weight is determined by the share of the individuals' group l in the total sample size, $\frac{L_l}{N}$. To test the validity of expression (3.25), I run a simulation with 200 individuals ($N = 200$) allocated to five different groups that each differ in size: $l_1 = 20$, $l_2 = 30$, $l_3 = 40$, $l_4 = 50$, $l_5 = 60$. I set each individual to randomly select five peers from within their clusters ($k = 5$). The result of a simulation with a DGP structure identical to that used in Table 3.3 (with 10,000 repetitions) indicates that the exclusion bias as estimated by this simulation, –0.147, is very similar to the predicted exclusion bias of –0.146 that is obtained by inserting the values for k and the various values for L_l into equation (3.25).

Second, the pool of potential peers might be correlated with, but not necessarily perfectly overlap, the cluster. For instance, although students tend to select many of their friends from within their classrooms, they usually also have peers other than classmates. Such empirical specifications would be characterised by properties of both case 1 and case 2 considered above: peer group formation is likely to be correlated with cluster formation (case 1), but some peers are selected from outside the cluster (case 2). Formally, consider a specification in which a proportion, θ, of peers is selected from within the classroom and a proportion, $(1 - \theta)$, is selected from outside the classroom.

Instead of $E(\bar{y}_{k_{li}}) = \bar{y}_{l,-i}$ that we considered in the previous sections, we would now model $E(\bar{y}_{k_{li}})$ as follows:

$$E(\bar{y}_{k_{li}}) = \theta \bar{y}_{l,-i} + (1-\theta)\bar{y}_{\Omega,-l} \tag{3.26}$$

where $0 \leq \theta \leq 1$ and $\bar{y}_{\Omega,-l}$ is the average outcome of the entire network Ω, excluding individual *i*'s group *l*. Although a further treatment of this extension is beyond the scope of this thesis, it is expected that, in a classroom fixed effects model, the exclusion bias would be smaller when at least one peer is selected from outside the cluster ($\theta < 1$) than when all peers are drawn from within the classroom ($\theta = 1$). Intuitively, in a model adding cluster fixed effects (which looks only at the variation in outcomes within the cluster), the exclusion bias is driven by a dependency of the expected peer group outcome $E(\bar{y}_{k_{li}})$ on $\bar{y}_{l,-i}$, and this dependency will be smaller whenever $\theta < 1$.

It is important to note here a related property of models that add cluster fixed effects: whereas an increase in peer group size unambiguously increases the magnitude of the exclusion bias as long as all peers are drawn from within the cluster (i.e. $\frac{\Delta|bias^{FE}|}{\Delta k^{inside}} > 0$), the bias becomes insensitive to additional peers who are drawn from outside the cluster (i.e. $\frac{\Delta|bias^{FE}|}{\Delta k^{outside}} = 0$).

A third complication is that, in practice, the pool of potential peers is not always well defined. Empirical studies differ in the value they – implicitly – attach to θ in equation (3.26). In some studies on peer effects in educational achievement, survey respondents are asked to identify peers from the entire school roster (e.g. Halliday and Kwak, 2012; Fletcher, 2012), thus $\theta < 1$.[31] In other studies, people are restricted to a list

[31] If it is further assumed that peers are as likely to be selected from within the classroom as from outside the classroom, we would have $\theta = \frac{L}{N}$.

of all students in the classroom in the identification of their peers (e.g. de Melo, 2011), thus $\theta = 1$. Fletcher and Ross (2012), in an attempt to control for correlated effects, construct 'clusters of observationally equivalent individuals who face the same friendship opportunity set and make the same type of friendship choices' within the school. To the extent that students select peers from within these clusters, it would be the size of these groups that would be relevant for the derivation of the magnitude of the exclusion bias. In reality, however, the boundaries of the pools from which peers are selected are usually difficult to draw.

Fourth, even if the exact pool of potential peers is perfectly known, in reality peers are usually not randomly drawn from within those pools. Therefore, the expected value of the outcomes of the selected peers might differ from the average of those pools (excluding individual *i*). In such cases, although we would expect the bias to be similarly present, the derivation of its magnitude would become considerably more complex.

Lastly, in this chapter we have only considered simple bivariate models. However, the addition of statistically significant control variables to the model is expected to reduce the magnitude of the exclusion bias. This is because, to the extent that these covariates explain some variation in the outcome variable *y*, they will partially control for the difference between *i* and the other individuals in the sample. Hence, they will partially capture the difference in expected value of the pools from which peers are drawn. Simulations indicate (not reported) that the reduction in the magnitude of the exclusion bias in models that add control variables with high coefficient parameters (e.g. 0.8) can be very significant. The reduction is negligible, however, when the coefficient parameters are relatively small (e.g. 0.20).

In general, when it is difficult to rely on theory for reasons outlined in this section, simulations of the type we ran in the previous section might be more useful in estimating the magnitude of the exclusion bias.

3.7. Conclusion

This chapter has formalised a cited (Guryan et al., 2009) but so far unproven source of downward estimation bias of OLS in standard empirical peer effects models. The main results derived in this chapter have various implications for the empirical social interaction literature.

The finding that the exclusion bias becomes more severe as the number of peers in the reference group grows implies that comparisons of estimates between models that vary in peer group size can be misleading. For instance, Glaeser et al. (2002) compare the estimates of peer effects on test scores of students at Dartmouth College, Hanover, NH, for three different units of peer group aggregation: The dormitory room (containing on average 2.3 students), the dormitory floor (containing on average eight students) and the dormitory (containing on average 27 students). According to this study, the estimated peer effect decreases with the level of aggregation. Although this finding may reflect diminishing peer effects for peers who are more distant, it may also be driven by the aforementioned property of the exclusion bias. Similarly, Halliday and Kwak (2012), whose specific aim is to compare peer effects estimates for different definitions of peer groups in the education literature, find that the estimated peer effect is significantly smaller when school grade cohorts are used instead of smaller circles of friends. Note that this result also applies to any heterogeneity analysis of peer effects across cohorts that differ in peer group size. For instance, if female students tend to have more friends than male students, then estimated peer effects that are smaller for females than for males may be confounded by the exclusion bias.

Furthermore, although I certainly do not argue that one should prefer models ignoring cluster fixed effects, I do caution against naive comparisons of these models with those that do control for such effects, in terms of their respective estimation results. As shown above, the exclusion bias is aggravated in the latter type of models, whenever peer group formation is correlated with cluster formation. For example, studies adding classroom effects (e.g. de Melo, 2011), dormitory effects (e.g. Sacerdote, 2001) or school effects (e.g. Fletcher, 2012) will be more severely affected by the exclusion bias if peers are selected – partially or completely – from within the respective clusters. The literature tends to interpret a drop in the estimate of the endogenous peer effect in models adding cluster fixed effects as evidence of unobserved covariates at the cluster level. Although such correlates will often matter, the results in this chapter suggest that such interpretations of the results may be confounded by the presence of the exclusion bias.

The finding that the downward exclusion bias, present in OLS estimation, is eliminated by 2SLS estimation strategies is equally important. It provides an alternative explanation for the common but counter-intuitive tendency of peer effects studies to obtain 2SLS estimates of endogenous peer effects that are larger than their OLS counterparts (e.g. Goux and Maurin, 2007; De Giorgi et al., 2010; de Melo, 2011; Brown and Laschever, 2012; Helmers and Patnam, 2012; Krishnan and Patnam, 2012; Naguib, 2012; Collin, 2013).

Finally, note that not all peer effects studies are affected by the exclusion bias. For example, there would be no problem if the peer group consists of *all* the individuals in student *i*'s classroom, including student *i* herself. Including the individual in her own peer group average outcome computations is rarely done in practice, however, because of the associated reflection problem (Manski, 1993). Nor would there be a problem if the pool of potential peers consists entirely of people who do not form part of the study

sample. For example, a study on peer effects between male students from an all-boys school and female students of a neighbouring all-girls school would not be affected by this bias if the units of interest are restricted to either male or female students. In most studies, however, peers are included in the study sample. The greater the overlap between the set of peers and the set of observations captured in the study sample, the greater the exclusion bias. Lastly, studies regressing an individual's current outcome levels on peers' baseline outcomes also prevent the exclusion bias from manifesting itself, as long as they control for the individual's own baseline outcome (e.g. Zimmerman, 2003; Lyle, 2007, 2009; Duflo et al., 2011; Fafchamps and Quinn, 2012).

If not properly addressed, however, the exclusion bias may significantly affect the results of social interaction studies. Its consequences may be especially problematic for empirical analyses that consider cluster fixed effects, for which the magnitude of the bias does not decrease as the sample size tends to infinity. This chapter has provided more insights into the cause, consequences and solutions of this bias, which has largely been ignored to date. Exploring other means of addressing this concern would be desirable in the future.

3.A. Chapter 3 Appendix

3.A.1. Relationship between σ_u^2 and σ_ε^2

We can rewrite equation (3.3) as follows:

$$u_i = \bar{y}_{k_i} - \bar{y}_{-i} = \frac{\sum_{j=1}^{k} y_{j_i}}{k} - \left(\frac{\sum_{j=1}^{N} y_{j_i} - y_i}{N-1} \right)$$

$$= \frac{\sum_{j=1}^{k} y_{j_i}}{k} - \left(\frac{\sum_{j=1}^{k} y_{j_i} + \sum_{j=k+1}^{N} y_{j_i} - y_i}{N-1} \right)$$

$$= \frac{(N-1)\sum_{j=1}^{k} y_{j_i}}{(N-1)k} - \frac{k\sum_{j=1}^{k} y_{j_i}}{(N-1)k} - \frac{\sum_{j=k+1}^{N} y_{j_i}}{(N-1)} + \frac{y_i}{(N-1)}$$

$$= \frac{(N-1-k)\sum_{j=1}^{k} y_{j_i}}{(N-1)k} - \frac{\left(\sum_{j=k+1}^{N} y_{j_i} - y_i \right)}{(N-1)}$$

Using this and $\text{var}(y_1) = ... = \text{var}(y_n) = \sigma_\varepsilon^2$, we can now derive $\text{var}[u_i]$:

$$\text{var}(u_i) = \sigma_u^2 = \frac{(N-1-k)^2 k\sigma_\varepsilon^2}{(N-1)^2 k^2} + \frac{(N-1-k)\sigma_\varepsilon^2}{(N-1)^2}$$

$$= \frac{(N-1-k)^2 k + k^2(N-1-k)}{(N-1)^2 k^2}\sigma_\varepsilon^2$$

$$= \frac{(N-1-k)k(N-1-k+k)}{(N-1)^2 k^2}\sigma_\varepsilon^2 \quad \Rightarrow \text{var}(u_i) = \sigma_u^2 = \frac{(N-1-k)}{(N-1)k}\sigma_\varepsilon^2 < \sigma_\varepsilon^2$$

$$= \frac{(N-1-k)}{(N-1)k}\sigma_\varepsilon^2$$

(3.27)

3.A.2. Deriving an expression for $\text{var}(\bar{y}_{k,-i})$

Using the reduced form of $\bar{y}_{-i}$ provided in equation (3.8), we obtain:

$$\text{var}(\bar{y}_{k,-i}) = \text{var}(\bar{y}_{-i} + u_i) = \text{var}(\bar{y}_{-i}) + 2\,\text{cov}(\bar{y}_{-i}, u_i) + \text{var}(u_i)$$

$$= \frac{\beta_1^2 \sigma_u^2}{(N-1+\beta_1)^2} + \frac{\sigma_\varepsilon^2}{(N-1+\beta_1)^2} + 2\frac{-\beta_1 \sigma_u^2}{(N-1+\beta_1)} + \sigma_u^2$$

$$= \frac{(N-1)^2 \sigma_u^2 + \sigma_\varepsilon^2}{(N-1+\beta_1)^2} \tag{3.28}$$

3.A.3. Exclusion bias in models adding cluster fixed effects (equal cluster sizes)

Starting from equation (3.11), we substitute for $\bar{y}_{\Pi,-i}$ by the cluster sampling equivalent of equation (3.6):

$$y_{li} = \beta_0 + \beta_1 \left(\frac{\left(\sum_{j=1}^{N_P} y_{lj} \right) - y_{li}}{N_P - 1} + u_{li} \right) + \varepsilon_{li} \tag{3.29}$$

Averaging equation (3.29) over all L observations in group l, we obtain the group average outcome $\bar{y}_l$:

$$\bar{y}_l = \beta_0 + \beta_1 \left[\frac{\sum_{i=1}^{L} \frac{\left(\sum_{j=1}^{N_P} y_{lj} \right) - y_{li}}{N_P - 1}}{L} + \bar{u}_l \right] + \bar{\varepsilon}_l = \beta_0 + \beta_1 \left[\frac{\left(\sum_{j=1}^{N_P} y_{lj} \right) - \bar{y}_l}{N_P - 1} + \bar{u}_l \right] + \bar{\varepsilon}_l$$

$$\Rightarrow \bar{y}_l = \beta_0 + \beta_1 \left[\frac{\left(\sum_{j=1}^{N_P} y_{lj} \right) - \bar{y}_l}{N_P - 1} + \bar{u}_l \right] + \bar{\varepsilon}_l \tag{3.30}$$

Subtracting equation (3.30) from equation (3.29), we derive the following expression for the cluster fixed effects model:

$$y_{li} - \bar{y}_l = \beta_1 \left[\frac{\left(\sum_{j=1}^{N_P} y_{lj} \right) - y_{li}}{N_P - 1} + u_{li} - \left[\frac{\left(\sum_{j=1}^{N_P} y_{lj} \right) - \bar{y}_l}{N_P - 1} \right] - \bar{u}_l \right] + \varepsilon_{li} - \bar{\varepsilon}_l$$

$$\Leftrightarrow y_{li} - \bar{y}_l = \beta_1 \left(\frac{\bar{y}_l - y_{li}}{N_P - 1} + u_{li} - \bar{u}_l \right) + \varepsilon_{li} - \bar{\varepsilon}_l \tag{3.31}$$

Denoting $\ddot{x} = x_{li} - \bar{x}_l$, for $x = y, u, \varepsilon$, we have:

$$\ddot{y} = \beta_1 \left(\frac{-\ddot{y}}{N_P - 1} + \ddot{u} \right) + \ddot{\varepsilon} \tag{3.32}$$

Using the properties of the covariance and variance operators, I obtain the following expression for the cluster fixed effects estimate of β_1:

$$E\left(\hat{\beta}_1^{FE}\right) = \beta_1 + \frac{\operatorname{cov}\left(\frac{-\ddot{y}}{N_P - 1} + \ddot{u}, \ddot{\varepsilon} \right)}{\operatorname{var}\left(\frac{-\ddot{y}}{N_P - 1} + \ddot{u} \right)}$$

$$= \beta_1 + \frac{\operatorname{cov}\left(\frac{-\ddot{y}}{N_P - 1}, \ddot{\varepsilon} \right) + \operatorname{cov}(\ddot{u}, \ddot{\varepsilon})}{\operatorname{var}\left(\frac{-\ddot{y}}{N_P - 1} \right) + 2\operatorname{cov}\left(\frac{-\ddot{y}}{N_P - 1}, \ddot{u} \right) + \operatorname{var}(\ddot{u})} \tag{3.33}$$

In order to expand equation (3.33), we consider:

$$\operatorname{cov}(\ddot{u}, \ddot{\varepsilon}) = E(\ddot{u}\ddot{\varepsilon}) = E\left[(u_{li} - \bar{u}_l)(\varepsilon_{li} - \bar{\varepsilon}_l) \right]$$

$$= E(u_{li}\varepsilon_{li}) - E(\bar{u}_l \varepsilon_{li}) + E(\bar{u}_l \bar{\varepsilon}_l) - E(u_{li}\bar{\varepsilon}_l) = 0 \tag{3.34}$$

and

$$\operatorname{var}(\ddot{u}) = \operatorname{var}(u_{li} - \bar{u}_l) = \operatorname{var}(u_{li}) - 2E(u_{li}\bar{u}_l) + \operatorname{var}(\bar{u}_l)$$

Furthermore, since $\bar{u}_l = \frac{\sum_{i=1}^{L} u_{li}}{L}$ and u is assumed independent across individuals:

$$\begin{cases} E(u_{li}\bar{u}_l) = \dfrac{E(u_{li}^2)}{L} = \dfrac{\sigma_u^2}{L} \\ \text{var}(\bar{u}_l) = \text{var}\left(\dfrac{\sum_{i=1}^{L} u_{li}}{L}\right) = \dfrac{\sum_{i=1}^{L} \text{var}(u_{li})}{L^2} = \dfrac{L\sigma_u^2}{L^2} = \dfrac{\sigma_u^2}{L} \end{cases}$$

$$\Rightarrow \text{var}(\ddot{u}) = \sigma_u^2 - 2\frac{\sigma_u^2}{L} + \frac{\sigma_u^2}{L} = \frac{(L-1)\sigma_u^2}{L} \tag{3.35}$$

Similarly:

$$\text{var}(\ddot{\varepsilon}) = \frac{(L-1)\sigma_\varepsilon^2}{L} \tag{3.36}$$

Using equation (3.32), we derive the reduced form of $\left(-\dfrac{\ddot{y}}{N_P - 1}\right)$:

$$\ddot{y} = \beta_1\left(\frac{-\ddot{y}}{N_P - 1} + \ddot{u}\right) + \ddot{\varepsilon} \Leftrightarrow \left[\frac{N_P - 1 + \beta_1}{N_P - 1}\right]\ddot{y} = \beta_1\ddot{u} + \ddot{\varepsilon}$$

$$\Leftrightarrow -\frac{\ddot{y}}{N_P - 1} = \frac{-\beta_1\ddot{u}}{N_P - 1 + \beta_1} - \frac{\ddot{\varepsilon}}{N_P - 1 + \beta_1} \tag{3.37}$$

Using $E(\ddot{\varepsilon}) = E(\varepsilon_{li} - \bar{\varepsilon}_l) = 0$, equations (3.34), (3.36) and (3.37), we derive:

$$\text{cov}\left(\frac{-\ddot{y}}{N_P - 1}, \ddot{\varepsilon}\right) = E\left[\left[\frac{-\ddot{y}}{N_P - 1} - E\left(\frac{-\ddot{y}}{N_P - 1}\right)\right]\ddot{\varepsilon}\right]$$

$$= E\left[-\frac{\beta_1\ddot{u}\ddot{\varepsilon}}{N_P - 1 + \beta_1}\right] + E\left[\frac{-\ddot{\varepsilon}\ddot{\varepsilon}}{N_P - 1 + \beta_1}\right]$$

$$= \frac{-\beta_1 E(\ddot{u}\ddot{\varepsilon}) - \text{var}(\ddot{\varepsilon})}{N_P - 1 + \beta_1}$$

$$= \frac{-\text{var}(\ddot{\varepsilon})}{N_P - 1 + \beta_1}$$

$$= -\frac{1}{N_P - 1 + \beta_1} \cdot \frac{(L-1)\sigma_\varepsilon^2}{L} \tag{3.38}$$

Similarly,

$$2\,\mathrm{cov}\left(\frac{-\ddot{y}}{N_P-1},\ddot{u}\right)=-2\frac{\beta_1\,\mathrm{var}(\ddot{u})}{N_P-1+\beta_1}-2\frac{E(\ddot{u}\ddot{\varepsilon})}{N_P-1+\beta_1}$$

$$=-2\frac{\beta_1\,\mathrm{var}(\ddot{u})}{N_P-1+\beta_1}=-\frac{2\beta_1}{N_P-1+\beta_1}.\frac{(L-1)\sigma_u^2}{L} \tag{3.39}$$

Again using equation (3.37), we have:

$$\mathrm{var}\left(\frac{-\ddot{y}}{N_P-1}\right)=\mathrm{var}\left[\frac{-\beta_1\ddot{u}}{N_P-1+\beta_1}-\frac{\ddot{\varepsilon}}{N_P-1+\beta_1}\right]$$

$$=\mathrm{var}\left(\frac{-\beta_1\ddot{u}}{N_P-1+\beta_1}\right)+\frac{2\beta_1\,\mathrm{cov}(\ddot{u},\ddot{\varepsilon})}{(N_P-1+\beta_1)^2}+\mathrm{var}\left(\frac{-\ddot{\varepsilon}}{N_P-1+\beta_1}\right)$$

$$=\frac{\beta_1^2\dfrac{(L-1)\sigma_u^2}{L}}{(N_P-1+\beta_1)^2}+\frac{\dfrac{(L-1)\sigma_u^2}{L}}{(N_P-1+\beta_1)^2} \tag{3.40}$$

Using equations (3.34)–(3.40), we obtain:

$$\mathrm{cov}\left(\frac{-\ddot{y}}{N_P-1}+\ddot{u},\ddot{\varepsilon}\right)=\left(-\frac{1}{N_P-1+\beta_1}.\frac{(L-1)\sigma_\varepsilon^2}{L}\right) \tag{3.41}$$

and

$$\mathrm{var}\left(\frac{-\ddot{y}}{N_P-1}+\ddot{u}\right)=\left[\frac{(N_P-1)^2\sigma_u^2+\sigma_\varepsilon^2}{(N_P-1+\beta_1)^2}\right]\frac{(L-1)}{L} \tag{3.42}$$

Hence, we can expand equation (3.33) as follows:

$$E\left(\hat{\beta}_1^{FE}\right)=\beta_1+\frac{\mathrm{cov}\left(\dfrac{-\ddot{y}}{N_P-1}+\ddot{u},\ddot{\varepsilon}\right)}{\mathrm{var}\left(\dfrac{-\ddot{y}}{N_P-1}+\ddot{u}\right)}=\beta_1+\frac{\left(-\dfrac{1}{N_P-1+\beta_1}\dfrac{(L-1)\sigma_\varepsilon^2}{L}\right)}{\left[\dfrac{(N_P-1)^2\sigma_u^2+\sigma_\varepsilon^2}{(N_P-1+\beta_1)^2}\right]\dfrac{(L-1)}{L}}$$

$$\Rightarrow E\left(\hat{\beta}_1^{FE}\right)=\beta_1-\frac{\sigma_\varepsilon^2(N_P-1+\beta_1)}{(N_P-1)^2\sigma_u^2+\sigma_\varepsilon^2}$$

Inserting the expression for σ_u^2 given by equation (3.27), we finally have:

$$E\left(\hat{\beta}_1^{FE}\right) = \beta_1 - \frac{\left(N_P - 1 + \beta_1\right)k}{\left(N_P - 1\right)\left(N_P - 1 - k\right) + k} \tag{3.43}$$

3.A.4. Exclusion bias in the between estimation model

The between-group model equivalent of the reduced form equation in (3.8) is

$$\begin{aligned}\bar{y}_{l,-i} &= \frac{\sum_{j=1}^{L} y_j - \beta_0}{L - 1 + \beta_1} - \frac{\beta_1 \bar{u}_l}{L - 1 + \beta_1} - \frac{\bar{\varepsilon}_l}{L - 1 + \beta_1} \\ &= \frac{L\bar{y}_l - \beta_0}{L - 1 + \beta_1} - \frac{\beta_1 \bar{u}_l}{L - 1 + \beta_1} - \frac{\bar{\varepsilon}_l}{L - 1 + \beta_1}\end{aligned} \tag{3.44}$$

where $\bar{y}_{l,-i}$ is the average outcome over the individuals in the group, excluding individual *i*, and $\bar{y}_l$, $\bar{u}_l$ and $\bar{\varepsilon}_l$ denote the group averages of *y*, *u* and ε, respectively.

Using equation (3.44), we find $\mathrm{cov}\left(\bar{\bar{y}}_{k_l}, \bar{\varepsilon}_l\right)$ and $\mathrm{var}\left(\bar{\bar{y}}_{k_l}\right)$, which we will need to find $p\lim\left(\hat{\beta}_1^{BE}\right)$ below:

$$\begin{aligned}\mathrm{cov}\left(\bar{\bar{y}}_{k_l}, \bar{\varepsilon}_l\right) &= \mathrm{cov}\left(\bar{y}_{l,-i} + \bar{u}_l, \bar{\varepsilon}_l\right) = \mathrm{cov}\left(\bar{y}_{l,-i}, \bar{\varepsilon}_l\right) \\ &= L\frac{E\left(\bar{\varepsilon}_l^2\right)}{L - 1 + \beta_1} - \frac{E\left(\bar{\varepsilon}_l^2\right)}{L - 1 + \beta_1} = \frac{L - 1}{L - 1 + \beta_1}\mathrm{var}\left(\bar{\varepsilon}_l\right) \\ &= \frac{\left(L - 1\right)\sigma_\varepsilon^2}{\left(L - 1 + \beta_1\right)L} = \frac{L - 1}{L - 1 + \beta_1}\mathrm{var}\left(\frac{\sum_{j=1}^{L}\varepsilon_j}{L}\right)\end{aligned} \tag{3.45}$$

$$\text{var}\left(\bar{\bar{y}}_{k_l}\right) = \text{var}\left(\frac{\sum_{i=1}^{L}\bar{y}_{k_i}}{L}\right) = \frac{1}{L^2}\text{var}\left(\sum_{i=1}^{L}\left(\frac{\sum_{j=1}^{L} y_{lj} - y_{li}}{L-1}\right) + \sum_{i=1}^{L} u_{li}\right)$$

$$= \frac{1}{L^2}\text{var}\left(\frac{L\sum_{i=1}^{L} y_{li}}{L-1} - \frac{\sum_{i=1}^{L} y_{li}}{L-1} + \sum_{i=1}^{L} u_{li}\right)$$

$$= \frac{1}{L^2}\text{var}\left(\sum_{i=1}^{L} y_{li} + \sum_{i=1}^{L} u_{li}\right) = \frac{\sigma_\varepsilon^2 + \sigma_u^2}{L}$$

$$= \frac{\left[1 + \frac{(L-1-k)}{(L-1)k}\right]\sigma_\varepsilon^2}{L} \tag{3.46}$$

where in the last step we used equation (3.27).

Using equations (3.45) and (3.46), we finally obtain:

$$p\lim\left(\hat{\beta}_1^{BE}\right) = \beta_1 + \frac{\text{cov}\left(\bar{\bar{y}}_{k_l}, \bar{\varepsilon}_l\right)}{\text{var}\left(\bar{\bar{y}}_{k_l}\right)} = \frac{\frac{(L-1)\sigma_\varepsilon^2}{(L-1+\beta_1)L}}{\frac{\left[1 + \frac{(L-1-k)}{(L-1)k}\right]\sigma_\varepsilon^2}{L}}$$

$$\Rightarrow p\lim\left(\hat{\beta}_1^{BE}\right) = \beta_1 + \frac{(L-1)^2 k}{(L-1+\beta_1)[(L-1)k + (L-1-k)]} \tag{3.47}$$

3.A.5. Derivation of $p\lim(\eta^2)$

Weight parameter η^2 in equation (3.15) is the ratio of the between-group sum of squares of the independent variable of interest, $\bar{y}_{k_i}$, to its total sum of squares:

$$\eta^2 = \frac{SS_{\bar{y}_k}^{BG}}{SS_{\bar{y}_k}^{Total}} = \frac{SS_{\bar{y}_k}^{BG}}{SS_{\bar{y}_k}^{BG} + SS_{\bar{y}_k}^{Within}} \tag{3.48}$$

Specifically, $SS_{\bar{y}_k}^{BG}$ is the sum of all the squared differences between each of the cluster group means and the overall sample mean, multiplied by the number of observations in the group *L*. In other words:

$$SS_{\bar{y}_k}^{BG} = SS_{\bar{y}_k}^{BE} \times L \tag{3.49}$$

where $SS_{\bar{y}_k}^{BE}$ is the sum of squares of $\bar{\bar{y}}_{k_i}$ in the between estimation regression (3.35) in section 3.3.5. Furthermore, using the definition of the variance operator, we know that:

$$\text{var}\left(\bar{\bar{y}}_{k_l}\right) = \frac{SS_{\bar{y}_k}^{BE}}{\left(\frac{N}{L} - 1\right)} \Rightarrow SS_{\bar{y}_k}^{BE} = \text{var}\left(\bar{\bar{y}}_{k_l}\right) \times \left(\frac{N}{L} - 1\right) \tag{3.50}$$

Using equations (3.48)–(3.50), we obtain:

$$SS_{\bar{y}_k}^{BG} = \text{var}\left(\bar{\bar{y}}_{k_l}\right) \times \left(\frac{N}{L} - 1\right) \times L$$

Substituting in for the expression of $\text{var}\left(\bar{\bar{y}}_{k_l}\right)$ given by equation (3.46), we finally have:

$$SS_{\bar{y}_k}^{BG} = \frac{\left(Lk - 2k + L - 1\right)}{k\left(L-1\right)}\left(\frac{N}{L} - 1\right)\sigma_\varepsilon^2 \tag{3.51}$$

Next, $SS_{\bar{y}_k}^{Within}$ is the sum of the squared differences between each individual's average peer group outcome, $\bar{y}_{k_i}$, and its average for the individual's group $\bar{\bar{y}}_{k_l}$. Similarly to equation (3.50), we have:

$$\text{var}\left(\bar{y}_{k_i} - \bar{\bar{y}}_{k_l}\right) = \frac{SS_{\bar{y}_k}^{Within}}{N-1} \Rightarrow SS_{\bar{y}_k}^{Within} = \text{var}\left(\bar{y}_{k_i} - \bar{\bar{y}}_{k_l}\right) \times \left(N-1\right)$$

From the above, we know that $\text{var}\left(\bar{y}_{k_i} - \bar{\bar{y}}_{k_l}\right) = \text{var}\left(\frac{-\ddot{y}}{N_P - 1} + \ddot{u}\right)$. Therefore, we can substitute in for the expression of $\text{var}\left(\bar{y}_{k_i} - \bar{\bar{y}}_{k_l}\right)$ by using equations (3.42) and (3.27).

We have:

$$SS_{\bar{y}_k}^{Within} = \left[\frac{(L-1)^2(L-1-k)+(L-1)k}{k(L-1+\beta_1)^2}\right]\frac{1}{L}(N-1) \tag{3.52}$$

Combining equations (3.48), (3.51) and (3.52), we obtain:

$$\eta^2 = \frac{SS_{\bar{y}_k}^{BG}}{SS_{\bar{y}_k}^{BG} + SS_{\bar{y}_k}^{Within}} \tag{3.53}$$

where:

$$\begin{cases} SS_{\bar{y}_k}^{BG} = \dfrac{(Lk-2k+L-1)}{k(L-1)}\left(\dfrac{N}{L}-1\right)\sigma_\varepsilon^2 \\ SS_{\bar{y}_k}^{Within} = \left[\dfrac{(L-1)^2(L-1-k)+(L-1)k}{k(L-1+\beta_1)^2}\right]\dfrac{(N-1)}{L}\sigma_\varepsilon^2 \end{cases}$$

Finally, taking probability limits, we obtain the following expression for $p\lim(\eta^2)$:

$$p\lim(\eta^2) = \frac{p\lim(SS_{\bar{y}_k}^{BG})}{p\lim(SS_{\bar{y}_k}^{BG}) + p\lim(SS_{\bar{y}_k}^{Within})}$$

$$= \frac{\dfrac{(Lk-2k+L-1)}{k(L-1)}}{\dfrac{(Lk-2k+L-1)}{k(L-1)} + \left[\dfrac{(L-1)^2(L-1-k)+(L-1)k}{k(L-1+\beta_1)^2}\right]} \tag{3.54}$$

Note that this result only holds in the limit, that is, when the sample size *N* tends to infinity. For small samples, however, $E(\eta^2) < p\lim(\eta^2)$. To see this, consider:

$$E(\eta^2) = E\left[\frac{(SS_{\bar{y}_k}^{BG})}{(SS_{\bar{y}_k}^{BG}) + (SS_{\bar{y}_k}^{Within})}\right]$$

$$= E(SS_{\bar{y}_k}^{BG})E\left(\frac{1}{(SS_{\bar{y}_k}^{BG}) + (SS_{\bar{y}_k}^{Within})}\right) + \operatorname{cov}\left(SS_{\bar{y}_k}^{BG}, \frac{1}{(SS_{\bar{y}_k}^{BG}) + (SS_{\bar{y}_k}^{Within})}\right)$$

Although the derivation of $\operatorname{cov}\left(SS_{\bar{y}_k}^{BG}, \frac{1}{\left(SS_{\bar{y}_k}^{BG}\right)+\left(SS_{\bar{y}_k}^{Within}\right)} \right)$ is beyond the scope of this thesis, we know that $\operatorname{cov}\left(SS_{\bar{y}_k}^{BG}, \frac{1}{\left(SS_{\bar{y}_k}^{BG}\right)+\left(SS_{\bar{y}_k}^{Within}\right)} \right) < 0$ (this result is also confirmed by simulations). Using this result, it can easily be shown that:

$$E\left(\eta^2\right) < p\lim\left(\eta^2\right) \tag{3.55}$$

The smaller the sample, the smaller $E\left(\eta^2\right)$, the more weight is given to the cluster fixed effect estimator, and, therefore, the more severe the exclusion bias.

3.A.6. Exclusion bias in models adding cluster fixed effects, with varying cluster sizes

As before, the cluster fixed effects estimate for β_1 in $y_{li} - \bar{y}_l = \beta_1\left(\bar{y}_{k_{li}} - \bar{\bar{y}}_{k_l}\right) + \varepsilon_{li} - \bar{\varepsilon}_l$ is approached by:

$$p\lim\left(\hat{\beta}_1^{FE}\right) = \beta_1 + \frac{\operatorname{cov}\left(\bar{y}_{k_{li}} - \bar{\bar{y}}_{k_l}, \varepsilon_{li} - \bar{\varepsilon}_l\right)}{\operatorname{var}\left(\bar{y}_{k_{li}} - \bar{\bar{y}}_{k_l}\right)}$$

However, now the within covariance and within variance will be weighted averages of the respective within covariances and within variances of groups *l* in the sample. Using equations (3.41) and (3.42), we obtain:

$$\begin{aligned} \left|Bias\right| &= -\frac{\sum_{l=1}^{\frac{N}{L}} \operatorname{cov}\left(\bar{y}_{k_{li}} - \bar{\bar{y}}_{k_l}, \varepsilon_{li} - \varepsilon_l\right) \times \frac{L_l}{N}}{\sum_{l=1}^{\frac{N}{L}} \operatorname{var}\left(\bar{y}_{k_{li}} - \bar{\bar{y}}_{k_l}\right) \times \frac{L_l}{N}} \\ &= \frac{\left[\sum_{l=1}^{\frac{N}{L}} \frac{L_l - 1}{\left(L_l - 1 + \beta_1\right)L_l} \times L_l\right]}{\left[\sum_{l=1}^{\frac{N}{L}} \frac{\left(L_l - 1\right)^2 \frac{\left(L_l - 1 - k\right)}{\left(L_l - 1\right)k} + 1}{\left(L_l - 1 + \beta_1\right)^2} \times L_l\right]} \end{aligned} \tag{3.56}$$

PART II

Empirical analysis

Chapter 4

Peer effects in development programme awareness of vulnerable groups in rural Tanzania[32]

4.1. Introduction

There is a growing literature on the effectiveness of community-based development (CBD) projects (for a recent review, see Mansuri and Rao, 2013). A pre-condition for grassroots participation, key for CBD success, is widespread programme knowledge among the eligible population (Khemani, 2007). Transparency allows community members to participate and to hold those in charge of resource spending to account. Asymmetric information, with poor and illiterate households having less access to information than richer and more powerful families, may result in elite capture in decentralised public service delivery (Bardhan and Mookherjee, 2006; Conning and Kevane, 2002). The current literature on local participatory institutions mainly focuses on village meetings and media campaigns as a means to strengthen community awareness (Goetz and Jenkins, 2001; Banerjee et al., 2010; Bjorkman and Svensson, 2007; Khemani, 2007; Olken, 2007; Shankar et al., 2010). The role played by social interactions in this process has received little attention to date. This chapter investigates the extent to which well-informed peers may influence the development programme awareness of vulnerable groups in rural Tanzania, the elderly and the disabled in particular.

Recent years have seen a growth in social interaction literature, especially in the area of education (Borjas, 1995; Gaviria and Raphael, 2001; Sacerdote, 2001; Hanushek et

[32] I thank the World Bank for sharing with me the baseline data collected for the impact evaluation of TASAF II. Specifically, I gratefully appreciate the support I received from Sarah Baird, Craig McIntosh and Berk Ozler during the data collection process. I also thank the Department of Economics at the University of Oxford, UK, for funding an additional trip to Tanzania to collect qualitative data for this project. I very much benefited from the support and assistance I received from Lies Caeyers during my last field trip to Tanzania.

al., 2003; Cipollone and Rosolia, 2007), but also in other areas such as technology adoption (Bandiera and Rasul, 2006; Conley and Udry, 2010; Van den Broeck and Dercon, 2011) and criminal activity (Glaeser et al., 1996; Bayer et al., 2009). In this literature, individuals' outcomes are influenced not only by their own characteristics but also by interactions with their peers. Specifically, since the pioneering work of Manski (1993), the probability of observing an individual having a certain outcome is modelled as a function of her own characteristics ('individual effects'), of the exogenous characteristics of her peers ('contextual peer effects'), the prevalence of this outcome among her peers ('endogenous peer effects') and some unobserved characteristics of the individual's environment ('correlated effects').

There is some evidence from the developed world to indicate that endogenous peer effects may also be present in the diffusion of information on welfare programmes (Duflo and Saez, 2003). The evidence from developing countries is limited. Information sharing among peers may well be crucial for a vulnerable household's awareness of, and hence participation in, CBD programmes. Most community-based initiatives rely heavily on village meetings and media campaigns (e.g. radio, television and newspapers) to sensitise communities (Mansuri and Rao, 2013). However, as a result of their physical and/or socioeconomic conditions, various disadvantaged demographic groups, such as the elderly and disabled, have only restricted access to such traditional sources of information about development programmes. To the extent that these groups have only limited direct access to news sources, social interactions with well-informed peers may be crucial for their inclusion in bottom-up development approaches.

In this chapter, I examine whether endogenous peer effects were present in the public diffusion of information about the Tanzania Social Action Fund II (TASAF II), one of the flagship community-driven development (CDD) projects of the government

of Tanzania.[33] Evaluating the targeting performance of the 'vulnerable group' (VG) component of TASAF II, Baird et al. (2011) find that 49% of the eligible non-beneficiary households had never heard about the programme.[34] Their study identified the household's direct access to information (such as education, village meeting attendance rate and the ownership of a radio/phone) and its direct political connectedness (such as being blood related to the village elite) as dominant determinants of being informed. However, Baird et al. (2011) do not consider the potential role of social interactions with informed peers in the information diffusion process. In view of the earlier discussion, the current chapter concentrates on two specific VGs targeted by TASAF II, namely the elderly (defined, in my base model, as those older than 55 years) and the (mentally or physically) disabled. The aim of this chapter is to measure the extent to which these VGs in rural Tanzania benefited from having well-informed individuals living in their neighbourhood, in terms of information exchange about TASAF II.

The household data that I will use to answer this empirical question come from a substantial baseline listing exercise organised by the World Bank in April–November 2008 to stratify and sample households to be followed up for an impact evaluation of the VG component of TASAF II. This unique dataset captures details of 30,339 households, including the 19,916 households of *all* elderly and disabled people, in 100 villages in rural Tanzania. The dataset includes, among others, global positioning system (GPS) coordinates of each of these households and information about these households' knowledge of TASAF.[35] The unit of study in this chapter is a household in

[33] Whereas CBD is an umbrella term for projects that engage communities in their design and management, CDD refers to CBD projects in which communities have control over key project decisions, including management of funds.
[34] The VG component of this programme was targeted at the elderly, the disabled, people infected with HIV/AIDS, orphans, widow(ers) and unemployed youth.
[35] The data were collected using advanced computer-assisted personal interviewing (CAPI) techniques, the comparative advantages of which were the focus of Chapter 2.

which there is at least one elderly and/or disabled person (henceforth 'elderly/disabled household').

My identification strategy is based on Manski's standard linear-in-means model (Manski, 1993). A set of empirical problems related to identifying peer effects have been extensively discussed in the literature (e.g. Manski, 1993; Brock and Durlauf, 2001; Moffitt, 2001). The first problem relates to defining and measuring the relevant peer group. I will motivate the choice of the set of 10 nearest neighbours (including both vulnerable and non-vulnerable households) as the relevant peer group in my main specifications. I will test the robustness of my results to alternative neighbours set definitions. The second identification problem, the presence of 'correlated effects', refers to the possibility that individuals may have similar outcomes to their peers because of some common characteristics (in the case of peer group sorting) or because of similar conditions they simultaneously face. The third problem is the 'reflection problem' highlighted by Manski (1993), which is essentially an identification problem in a system of simultaneous equations. Finally, given that I consider a linear-in-means model in which households are excluded from their own peer groups, the exclusion bias (discussed in Chapter 3) is likely to present itself.

To deal with these empirical issues, I employ a two-stage least squares (2SLS) estimation strategy recently developed by Bramouille et al. (2009) and De Giorgi et al. (2010). This strategy uses exogenous characteristics of peers' peers ('excluded peers') as instruments for the peer group's average outcome. In addition, I control for 'hamlet' fixed effects (something that is rarely feasible in practice because of data limitations) and a set of observed neighbours set and individual household characteristics.[36]

My findings are consistent with a significant and large endogenous peer effect: *each* additional informed neighbour in the set of 10 nearest sampled neighbours raises an

[36] The hamlet (*kitongoji*) is the smallest administrative unit in rural Tanzania.

elderly/disabled household's probability of being informed by 8 percentage points (pp). This compares to an increase by 14 pp of the probability of being informed if a household attended a village meeting (i.e. TASAF's most important sensitisation medium). Importantly, the unique size of the dataset also allows me to identify the type of elderly/disabled households that are most/least receptive to information in the neighbourhood, and those households (vulnerable and non-vulnerable) that are most effective in transmitting information to others. From a policy perspective, the finding of such a significant endogenous peer effect is of primary importance, in view of its social multiplier effect. Further, these results provide development intervention planners, targeting elderly and disabled households, with a better understanding of which individuals in the village network to target in their sensitisation campaigns.

In addition to identifying the presence of endogenous peer effects in development programme awareness, in Appendix 4.A.3, I use simulations to illustrate the importance and properties of the exclusion bias, formally discussed in Chapter 3; the ordinary least squares (OLS) are adjusted accordingly, allowing to correctly interpret the observed differences (i) between OLS and 2SLS estimation results and (ii) between the estimates of models with and without cluster fixed effects. Determining the proportion of the overall OLS bias that can be attributed to the exclusion bias informs on the true extent to which unobserved correlated effects are affecting the results.

This chapter is organised as follows. Section 4.2 summarises the specific TASAF II programme objectives, resource allocation procedures and awareness-building strategies. Section 4.3 subsequently describes the data and motivates the peer group measure used in this study. Before moving on to the empirical analysis, section 4.4 carefully discusses the empirical issues involved and the identification strategy employed. The empirical results of the chapter are presented in section 4.5. The main

results of the study are tested against robustness to alternative specifications in section 4.6. Section 4.7 concludes.

4.2. Background of Tanzania Social Action Fund II

4.2.1. Programme objectives and resource allocation process

Tanzania is currently implementing TASAF II, a large-scale CDD fund financed by the World Bank. Launched in May 2005 as an important component within the national framework of the Tanzanian National Strategy for Growth and Reduction of Poverty (popularly known as MKUKUTA), TASAF II follows its predecessor TASAF I (2000–2005) in the aim of empowering local communities to participate in development initiatives. Through an elaborate screening process, eligible communities throughout Tanzania could collectively apply for TASAF funding to implement projects corresponding to their most urgent needs.

The benefiting community could be an entire ward, an entire village or a subgroup of the village (e.g. the food insecure, widows, elderly). Among the communities explicitly targeted by TASAF II were the VGs, comprising orphans, disabled, elderly, widows/widowers and those infected by HIV/AIDS.[37] These VGs could collectively apply for funding to start up income-generating group activities (Tanzania Social Action Fund, 2011a). Examples of VG income-generating activities funded by TASAF include poultry farming, dairy cattle keeping and tailoring.

In accordance with the principles of CDD, TASAF's institutional structure is highly decentralised. Officials at the national level allocate funds to districts, and within those districts resource allocation follows a bottom-up approach (Tanzania Social Action Fund, 2011a). The village council is responsible for sensitising all members of its

[37] TASAF II also targeted (i) service-poor communities, which could apply for projects improving basic social and market services such as education, water and sanitation, roads, banking and markets, and (ii) communities wishing to apply for funding to start up public work programmes targeted to the food insecure, that is, able-bodied members of food-insecure households.

village about TASAF and for encouraging all eligible communities to participate in the programme. Any interested community can indicate its interest in a particular project to the village council. Officials at the village and district level assess applications for eligibility and merit, after which village assemblies call all village members to a village meeting where the selected projects are discussed. To ensure that the selected projects are indeed those desired by the majority of the community, attendees vote on pairwise combinations of these projects, but only if at least 50% of the targeted community members are present at the meeting.

At the outset, it was expected that one-third of all 11,000 Tanzanian villages would have at least one project funded by TASAF II by the end of 2010, targeting 5950 projects in total.[38] In June 2008, at the time of baseline data collection for the impact evaluation (see below), a total of 4582 subprojects had been selected for funding by TASAF II, of which 1091 (23.8%) were for VGs. The total number of VG beneficiaries was 432,268, of which the elderly group captured 28.5%, the widow/widowers group 21.1%, the HIV/AIDS group 14.8%, the disabled group 10.6% and the unemployed youths 12.7% (Tanzania Social Action Fund, 2008).[39] In this chapter, I focus on the elderly and the disabled VGs. Although some VGs existed before TASAF II, most of the TASAF groups were formed only after people were informed about the programme.

4.2.2. Official programme awareness strategies and the scope for peer effects

TASAF II's flagship feature is its CDD development approach. Listed among its 12 guiding principles are the following three items (Tanzania Social Action Fund, 2011a): 'Demand-driven and follow a bottom-up planning and decision-making approach'; 'Transparent and demonstrate full public accountability'; and 'Strengthen community

[38] In June 2011, the number of projects funded amounted to 10,526, far exceeding the original target of 5950 projects, thanks to additional resources obtained (Tanzania Social Action Fund, 2011b).
[39] Note that, in June 2011, the distribution of subprojects across beneficiary groups had moved to 41% service poor, 50% VGs and 9% food insecure (Tanzania Social Action Fund, 2011b).

empowerment'. An essential prerequisite for adherence to these principles is widespread programme awareness. Successful CDD approaches are expected to be inclusive (i.e. to empower all minority groups to identify and express their priorities). To achieve this objective, it is important that programme awareness reaches all members of the community.

The importance of programme awareness is also recognised by TASAF itself: 'Enhancement of skills and knowledge at all levels is a crucial element for participatory development' (Tanzania Social Action Fund, 2008). One of the main activities implemented by TASAF during the first 2 years of implementation (2005–2007) was, therefore, building programme awareness through various sensitisation methods (Tanzania Social Action Fund, 2006, 2007). Technical launch workshops were organised for politicians and district technical staff to inform them about the delivery details of TASAF II. Communication and promotional materials, such as brochures, posters, t-shirts and caps, were distributed to districts, which in turn were responsible for further distribution to their respective communities. District officials were also in charge of training village officials in their respective communities (e.g. through workshops). In June 2008, there were 3030 trained experts at district and national level in charge of training communities in project implementation.

Every one of Tanzania's 11,000 villages were supposed to be visited by a district official and given details about the programme. The village council members in turn were in charge of raising awareness to the larger public through village meetings. The extent to which this exercise was successful varies from district to district, and from village to village. Annual TASAF reports highlight that information dissemination was challenging in large districts with sparse populations because of high operational costs. They also highlight that village leaders played a crucial role in the dissemination of

TASAF information to all households within the village, by organising village meetings.

Information about TASAF was further disseminated through various media. Apart from regular coverage on general radio programmes at zonal level, as of June 2006 TASAF broadcasted a half-hour weekly magazine programme via the Dar es Salaam-based Radio Uhuru and Radio Tanzania to various districts. The programmes discussed the various stages of the programme, from project application to implementation. TASAF was also covered in national and regional newspapers (English and Kiswahili) and on national television (TBC1). In addition to national TASAF newsletters that were distributed to all districts on a quarterly basis, some districts received bi-annual regional newsletters. These newsletters, including descriptions of TASAF success stories, were intended for distribution to all villages within the districts.

To summarise, the official TASAF information campaigns relied heavily on village meetings, radio and television programmes, newspapers and newsletters to reach eligible communities. However, elderly and disabled individuals' access to these primary sources of information is likely to be restricted in rural Tanzania, owing to their physical and socioeconomic constraints. For instance, physical disabilities may prevent individuals from regularly attending village meetings. Even if they do attend those meetings (village meeting attendance is compulsory in most of rural Tanzania), a lack of education may prevent individuals from effectively capturing information during those meetings.[40] Given that village meetings are always conducted in Swahili, those not familiar with this language are unlikely to benefit from them. Moreover, to the extent that elderly/disabled individuals have higher illiteracy rates, they are less likely to read newspapers and newsletters. Poor economic conditions may deny them access to radios

[40] In many villages, village meeting attendance is compulsory for all citizens older than 18 years. Some demographic groups are exempt from this norm, such as women taking care of children and very old people.

and televisions. Social interactions with well-informed peers may well compensate for this lack of direct access to information.

4.3. Data

4.3.1. Data source

The data in this chapter come from a listing exercise organised at baseline (April–November 2008) of an evaluation of the VG component of TASAF II, conducted by the World Bank. Twenty villages of five Tanzanian districts (100 villages in total) were randomly selected for the evaluation.[41] The data were collected using CWEST (capture with enhanced survey techniques), a CAPI survey application for which the comparative benefits were extensively discussed in Chapter 2. The five districts we visited were Moshi, Lushoto, Makete, Kwimba and Nzega (Figure 4.1). Each of the 100 sampled villages had at least one VG selected for TASAF funding.[42] The deposit of the funds was pending until after the baseline survey visit and therefore no TASAF VG activities had been undertaken in those areas at the time of our survey visit.

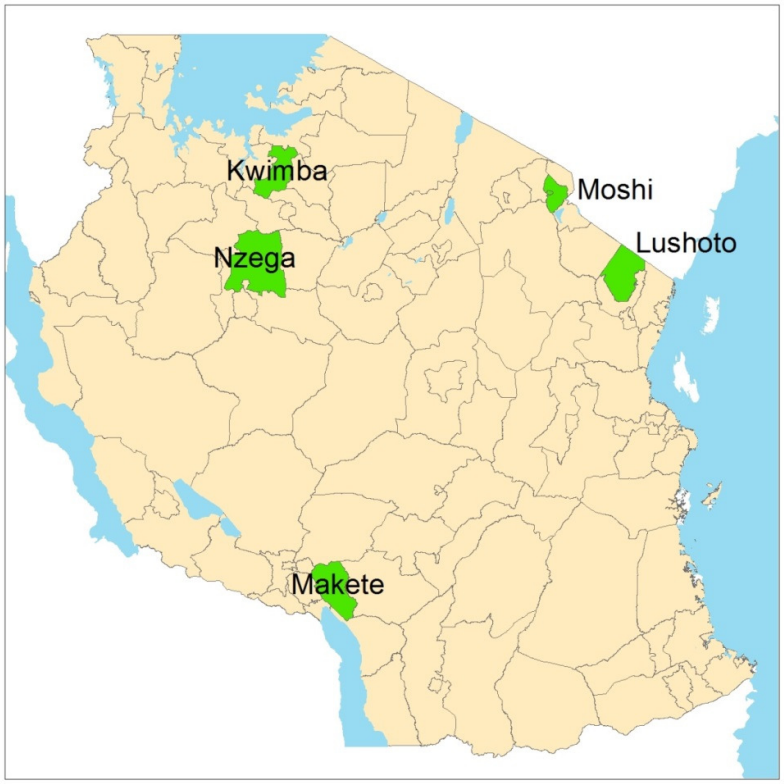
Visits of *all* households living in these 100 villages resulted in a dataset of 61,610 households. A short listing survey was conducted of each of these households, collecting basic demographic data and detailed information about the vulnerability status of the household members: the presence of any male or female elderly (defined as those older than 55 years), any widows/widowers, any orphans and any disabled individuals (including the type of disability). Henceforth, in this chapter, I denote

[41] In the position of project manager at Economic Development Initiatives (EDI), the Tanzania-based survey company that carried out the data collection for the TASAF II impact evaluation, I was in charge of the entire data collection process for this project.

[42] Less than 5% of the households in the sample contain TASAF VG beneficiary members. This relatively low number of TASAF beneficiaries does not mean that TASAF II was not considered an important development programme. By sampling design, all villages in the sample are those that ended up selecting VG activities as development initiatives to be funded by TASAF. The latter activities tend to benefit only a relative small number of households. However, the decision to select TASAF VGs as the type of activities to be funded by TASAF was the result of a long decision process, in which VG activities were balanced against projects improving basic social and market services, such as building schools and roads and public work programmes. The stakeholders of the last two project types are likely to be higher in number than those of VG activities. It was therefore in any community member's interest to be informed about TASAF.

households with at least one such member as 'VG households'. We collected GPS data for all dwellings (latitude, longitude and altitude), which allows me to map all households and to calculate distance measures. Automatic capturing of the GPS data ensured high quality of these data (see the results in Chapter 2).

Figure 4.1 Sampled districts (each containing 20 sampled villages)



Source: Baird et al. (2011).

In addition to the short-listing survey, a longer survey was carried out of a large subset of households in each village. These data include details on household head characteristics (e.g. education), basic consumption data, asset ownership and social/political connectedness of the household members. This longer survey also asked the respondents the question 'Have you personally ever heard of a development programme called TASAF?' I will use the answer to this question (yes or no) as the main proxy for a household's TASAF awareness. In section 4.6.2 I discuss the limitations of this measure and test the robustness of my results to an alternative proxy.

The subset of households receiving both the short- and the long-listing survey includes *all* village elites (village chairperson and village executive officer), *all* VG households (including the elderly/disabled households) and a randomly sampled subset of the non-VG households. The resulting sample contains 35,870 households. For 15% of these households, we have missing data because of the absence of household members during listing.[43] The resulting dataset of 30,339 households contains 23,865 VG households in total, of which 19,916 are elderly and/or disabled households. Appendix 4.A.1 summarises the population and sample characteristics.

4.3.2. Peer group measure

Nearest neighbours set

In the social interaction literature, peer groups have been defined in various ways and are often based on geographic and/or social proximity. Spatially restricted reference groups include farmers living in the same district (Case, 1992), households living in the same village (Munshi, 2004; Wydick et al., 2011), people living in the same neighbourhood (Glaeser et al., 1996; Aaronson, 1998; Bayer and Ross, 2009; Wydick et al.), students in the same school (Evans et al., 1992) and roommates living in the same dormitory (Sacerdote, 2001). Measures based on social proximity consider, for instance, similarities in ethnicity or caste (Borjas, 1992, 1995; Grimard, 1997; Fafchamps and Lund, 2003; Luke and Munshi, 2006; Munshi and Rosenzweig, 2006), family and friends (Bandiera and Rasul, 2006), and people of the same age group and/or education level and/or religion (Bandiera and Rasul, 2006).

However, most peer group definitions are ad hoc and rely on strong assumptions about the way people interact with each other (Manski, 2000). Without asking individuals directly about the set of people they interact with, as is done in Conley and

[43] This attrition rate is unlikely to be random, because the chances for mobile households to be absent at an unexpected visit are high relative to less mobile households. This means that attrition naturally leaves our sample with the least mobile vulnerable households in the village, which – for reasons discussed below – is actually beneficial for the estimation strategy used in this study.

Udry (2005), De Weerdt and Dercon (2006) and Van den Broeck and Dercon (2011), it is generally difficult to convincingly define peer groups. Reasonable exceptions to this are situations in which individuals are somehow constrained in their interactions. An example of this is the study by Bayer et al. (2009) on the influence of juvenile offenders on their peers in juvenile institutions. Given the legal constraint on the freedom of movements of the subjects in their study, it is reasonable to define peers as individuals who serve time in the same correctional facility, despite the lack of detailed social network data available. Given similar data limitations in this chapter, my peer group definition relies on the physical and socioeconomic constraints faced by elderly/disabled households in rural Tanzania. Specifically, I will assume that such households are restricted to their geographically nearest neighbours for receiving information on development initiatives such as TASAF.

In the developed world, there is a legitimate claim that geographic distance has become less restrictive for social relationships. In places such as rural Tanzania, however, geographic location remains a highly important factor, especially for people with physical disabilities and socioeconomic constraints. In our sample of 19,916 elderly/disabled households, only 2% own a motorbike, car or truck. Despite the proliferation of mobile phones throughout the continent in the previous decade, 76% of the elderly/disabled households in our sample do not own a mobile phone; 44% of them do not own a radio; and 43% of them are headed by a person who has not had any form of education, which makes them less likely to listen to radio programmes or to read newsletters and leaflets about TASAF. As a result, the majority of these groups heavily rely on their peers within the village for access to information.

If we further assume that elderly/disabled households need frequent interactions with well-informed people for information to be effectively transmitted to them, the assumption underlying the nearest neighbours set definition is not that stringent: it only

requires the assumption that elderly/disabled households *mainly* interact with their nearest neighbours about development programmes such as TASAF. Note that my study does allow for some form of interactions with individuals other than neighbours, such as irregular meetings at the village market or greetings on the street, as long as these do not result in effective information exchange about TASAF.

Qualitative evidence gathered for this chapter suggests that people in rural Tanzania mainly discuss welfare programmes with their blood relatives.[44] Hence, it is important to consider peer group definitions that include blood relatives. In fact, as land traditionally is passed on from father to son in these areas, people tend to live near to their relatives. Therefore, using the nearest neighbours set as reference group has the additional advantage of including most of the household's blood relatives as peers. On the other hand, the fact that peer groups systematically include blood relatives may cause problems of unobserved correlated effects, which is something that my identification strategy will have to take into account (see below). To allow for relatives living further away, my regressions will control for whether or not the household has blood relatives living in the village.

To address the concern that the respondent who answered the question 'Have you personally ever heard of a development programme called TASAF?' (my measure of TASAF awareness) was not necessarily the vulnerable person (and therefore not necessarily less mobile), in section 4.6.2 I will test whether my results are robust to restricting the sample to those households for which the survey respondent was the household head.[45] This test also addresses the possibility that elderly/disabled individuals may have heard about TASAF from more mobile members within their

[44] These contextual data were collected in November 2012 during an additional field trip organised in preparation of this thesis chapter.

[45] The data do not allow me to identify the elderly/disabled individuals within the household and therefore I cannot restrict the sample to such respondents. However, given that the elderly members tend to be the household heads in Tanzania, and given that the elderly also tend to have disabilities, I consider household head to be a reasonable proxy for the elderly/disabled individual in the household.

household. Since information tends to reach households via the household heads in Tanzania, restricting the sample to these individuals limits the likelihood of contamination by mobile members obtaining the information elsewhere.

Distances are calculated using three-dimensional Euclidean distance formulae between GPS locations of households, in which I also consider the elevation level of the household's dwelling. Given the specific survey sampling strategy applied during collection of data used in this study (see section 4.3.1), in which only a fraction of non-VG households were sampled whereas all VG households were sampled, I give each neighbour j in household i's neighbours set an importance weight, $\omega_{i,j}$, equal to the inverse probability of being sampled for the survey. When computing means of variables at the peer group level, I normalise the weights to sum to 1 within each group.

Base specification: 10 nearest neighbours

For reasons outlined next, in my base model I define the peer group by the set of 10 nearest (sampled) neighbours, restricted to those living within a maximum distance threshold of 1 kilometre.[46] This neighbours set may include both vulnerable and non-vulnerable households. I do not claim, however, that this measure is superior to other measures at all levels, such as the set of five, 15 or 20 nearest neighbours, or the set of neighbours living within a 300-metre distance band. In the sample, the maximum distance of the tenth nearest sampled household ranges from 92 to 1513 metres (see Appendix 4.A.1). By using the 10 nearest neighbours set, I risk excluding relevant households living nearby, or risk including households living too far. My measure, just like any other peer group measure, is likely to suffer from measurement error. It is important to test the robustness of my results to alternative nearest neighbours set definitions. I will do this in section 4.6.2.

[46] I choose 1 kilometre as a threshold in view of the evidence provided in the longer household questionnaire that many vulnerable individuals find it difficult to walk more than 1 kilometre, and hence are less likely to interact with households living further away.

Since one purpose of this chapter is to illustrate the empirical importance of the exclusion bias, it is in the reader's interest to consider peer groups of fixed sizes. As previously shown, the peer group size (k) constitutes an important determinant of the magnitude of the exclusion bias. If I were to allow peer group sizes to vary across households, for instance by using the distance bands approach (Van den Broeck and Dercon, 2011), this would make the demonstration of the properties of the exclusion bias unnecessarily complex. For the purpose of clarity, I therefore opt for the *k nearest neighbours approach* in this chapter.

One reason for choosing the 10 nearest neighbours, instead of, for instance, the five, 15 or 20 nearest neighbours, is that it minimises the sample size reduction. In my analysis, I will need to drop households for which not all k nearest neighbours live within the 1-kilometre range. Moreover, the 2SLS estimation strategy that I will use requires households with no 'excluded neighbours' to be dropped (see below). The extent to which these restrictions lead to a reduction in the total sample size depends on the number of nearest neighbours considered. When I consider the 10 nearest neighbours, the sample reduction is minimal – less than 3%.[47]

In the choice of k in my base model, I also took into consideration the fact that I will control for hamlet fixed effects in my main regression models. Below, I will discuss more extensively why it is important to do this. For now, note that a hamlet is the smallest administrative unit in Tanzania. The average hamlet in the sample contains 28 elderly and disabled households, ranging from 1 to 135 (see Appendix 4.A.1). When controlling for hamlet fixed effects, we are looking at within-hamlet variation in TASAF awareness. It is therefore in our interests to keep k small.

[47] This compares with a drop of 6% and 11% of sampled households for, respectively, the 15 and 20 nearest neighbours set peer group definitions. For the five nearest neighbours set, the drop is 2%, but, as I will highlight below, the exclusion restriction of the instruments I use is unlikely to be satisfied for such a small number of households.

Finally, it is important to note that only a random subset of neighbours is captured in the sample, not all actual neighbours. The nearest sampled tenth neighbour lives, on average, 310 metres away, which is approximately the same distance as the average nearest actual 20th neighbour lives in the true population, that is, 335 metres away (see Appendix 4.A.1). This means that the set of 10 *sampled* neighbours is in fact a proxy for the set of 20 *actual* neighbours (this will be important when interpreting the results). Also note that 306 metres is similar to the 300-metre band limit that Van den Broeck and Dercon (2011) consider in their neighbours set definition using another Tanzanian dataset.

4.3.3. Data description

This study is concerned with some of the most vulnerable groups in Tanzania, namely the elderly and the disabled. Development programmes such as TASAF II aim to empower such households in the improvement of their living standards. To describe the context in which the VG component of TASAF II was launched, Table 4.1 compares elderly and disabled households with other households, in terms of baseline characteristics.[48] The elderly constitute 35% of the population, whereas households with disabled individuals make up 6% of the population.[49] Overall, elderly and disabled households are significantly worse off than non-vulnerable households. Although their housing conditions are similar, these VGs own significantly fewer assets (radio, phone, vehicle and watch or clock), eat meat and eggs on a less frequent basis, are significantly less educated, and the elderly are less likely to have heard about TASAF.

[48] Note that the 'other households' include other types of vulnerable households, such as widow(ers).
[49] Note that these vulnerable groups are not mutually exclusive. That is, the elderly group contains disabled households and the disabled group contains elderly households.

Table 4.1 Elderly/disabled versus other households: a comparison in characteristics

	Other households	Elderly households	Disabled households
Proportion of population (%)	54.9	35.1	6.0
Percentage that owns:			
Radio	66.7	55.7***	52.5***
Phone	36.4	23.6***	24.1***
Vehicle	49.0	38.0***	40.3***
Watch/clock	40.8	37.6**	34.4***
Sofa	21.0	15.9*	14.5***
Average HH size	4.9	5.3***	6.1***
Percentage that is female headed	12.1	31.9***	28.7***
Average age of HH head (years)	36.9	64.4***	56.5***
Percentage with head with primary education	85.8	56.2***	62.6***
Percentage with head = member main tribe of village	78.4	85.2**	84.3***
Percentage with head = member main religion of village	53.9	61.6***	56.8
Percentage of dwellings with improved:			
Floor	29.5	22.7*	21.5**
Roof	63.2	62.4	59.1*
Walls	26.0	20.6	20.9
Water source	44.9	44.0	46.8
Latrine	62.0	60.0	57.0**
Energy source	11.3	8.0	7.1*
Average number of days that the HH ate [ITEM] during the previous week:			
Meat	1.0	0.7***	0.7***
Eggs	0.3	0.2***	0.2***
Milk	1.8	1.9	1.7
Percentage who attended village meeting in previous 3 months	46.7	47.2	47.6
Percentage who attended village meeting in the previous 6 months	68.7	68.5	69.5
Percentage with blood relative of village elite	6.8	7.0	7.0
Percentage with a member who holds any village office	5.4	5.4	5.2
Percentage TASAF informed	59.2	52.6***	57.4
N	6474	18,762	3303

Notes: Observations are weighted by inverse sampling probability. Significance of difference from non-vulnerable group is indicated by asterisks. Vehicle includes bicycle, car, motorbike or truck. See section 4.4.1 for a description of the other variables. ***$p<0.01$; **$p<0.05$; *$p<0.10$.
HH, household; TASAF, Tanzania Social Action Fund.

Almost half of the elderly/disabled households attended a village meeting at least once in the 3 months prior to the survey.[50] This number is not significantly different from that of the non-vulnerable group. This may be surprising given that we expect the elderly and disabled to be relatively less mobile. However, as mentioned above, village meeting attendance is not necessarily voluntarily in Tanzania. Moreover, the data were collected a few months after TASAF VG beneficiaries had been selected. As explained in section 4.2.1, the beneficiary selection could only go ahead if at least 50% of the VG community members were present at the village meeting. Qualitative discussions with villagers indicate that village officials made significant efforts (sometimes using force) to mobilise the VG population in order to reach that quota. As we will see next, despite the relatively high number of elderly/disabled households that attended village meetings, many of those were not informed about TASAF, probably as a result of socioeconomic constraints.

At the time of the survey, only 53% of elderly households and 57% of disabled households were informed about TASAF (Table 4.1). On average, village meetings do seem to be effective in informing people. Whereas 60% of the elderly/disabled households that attended the meetings were informed, only 47% of the absentees were informed (not reported). I make two observations here, as follows. (i) Village meetings are not the only means of informing households, given that almost half of the households that did not attend were also informed. This finding provides some scope for social interactions, with people attending village meetings passing on information to the absentees. (ii) Village meetings are not entirely successful in informing households, as 40% of those attending were not informed. This is possibly the result of a lack of

[50] Most village meetings concerning TASAF II took place within the 6-month period prior to the survey visit.

education and related languages problems.[51] This means that, even for those attending village meetings, there is some scope for regular neighbourhood interactions to enhance the transmission of information.

Table 4.2 compares the characteristics of informed and uninformed elderly/disabled households. The first group is, on average, wealthier in terms of assets, amenities and consumption patterns. A higher proportion of them have an educated head (71% versus 42%); they have more household members; and they are less likely to be female headed (25% versus 39%). They are almost twice as likely to be blood related to the village elite and more than four times as likely to hold political office. On average, compared with their uninformed counterparts, informed households live significantly closer to the village leader (by almost 400 metres; note from Appendix 4.A.1 that the average distance between sampled households in the village is 1800 metres), and they have a higher proportion of informed neighbours (six out of 10 compared with five out of 10).

4.4. Empirical strategy

4.4.1. Empirical specification

This chapter aims to identify the role of endogenous peer effects in the development programme awareness of the elderly and disabled in rural Tanzania. Besides the difficulty in defining relevant peer groups (see section 4.3.2), the identification of social interactions encounters other well-known challenges extensively discussed in the literature (Manski, 1993; Brock and Durlauf, 2001, 2007; Moffitt, 2001; Graham and Hahn, 2005; Soetevent, 2006; Graham, 2008). It is difficult to distinguish social learning from other mechanisms that make individuals' outcomes look similar.

[51] When I look at the proportion of informed village meeting attenders by education category, I find, indeed, that the uneducated households are least likely to benefit from attendance at village meetings. Other reasons provided by the villagers during qualitative discussions include reporting bias, that is, some people may lie about village meeting attendance because they are afraid of being punished if they are not attending. Another reason provided is that some of them are not at all interested in the meetings, and only attend because attendance is compulsory. As a consequence, 'they are drunk and/or they fall asleep and/or chat with their friends during the meetings' (according to respondents to qualitative surveys).

Table 4.2 Characteristics of informed and uninformed elderly/disabled households

	TASAF-uninformed elderly/disabled households	TASAF-informed elderly/disabled households	Statistical significance level of difference
Proportion of sample (%)	46.73	53.27	–
Percentage that owns:			
Radio	45.9	64.4	***
Phone	16.6	30.1	***
Vehicle	39.1	37.6	
Watch/clock	28.3	45.4	***
Sofa	10.6	20.5	***
Average HH size	5.1	5.7	***
Percentage that is female headed	39.0	24.7	***
Average age of HH head (years)	64.9	61.4	***
Percentage with head with primary education	42.3	70.8	***
Percentage with head = member main tribe of village	83.3	86.6	***
Percentage with head = member main religion of village	60.7	61.4	
Percentage of dwellings with improved:			
Floor	16.4	28.1	***
Roof	53.1	70.0	***
Walls	15.8	24.8	***
Water source	39.4	48.2	***
Latrine	54.9	63.2	***
Energy source	5.7	10.0	***
Average number of days that the HH ate [ITEM] during the previous week:			
Meat	0.6	0.8	***
Eggs	0.2	0.3	***
Milk	1.5	2.2	***
Percentage that attended village meeting in previous 3 months	40.1	53.5	***
Percentage that attended village meeting in the previous 6 months	60.4	76.0	***
Percentage with blood relative of village elite	4.7	9.1	***
Percentage with a member who holds any village office	1.7	8.6	***
Percentage of nearest 10 neighbours who are TASAF informed	47.6	64.0	***
Average distance from village leader (metres)	1938	1547	***
N	9306	10,610	

Notes: Significance level of difference is indicated by asterisks. ***$p<0.01$; **$p<0.05$; *$p<0.10$.
HH, household; TASAF, Tanzania Social Action Fund.

If two neighbours are informed about TASAF, one cannot immediately conclude that this is a result of social interactions between the two. Manski (1993) provides three reasons for individuals having similar outcomes to their peers:

1 *Endogenous social effects*, that is, the effects of the outcome of one's peers on one's own outcome. In our context, this refers to the influence of one household's TASAF information set on its neighbouring household's TASAF awareness, through word of mouth. This is the effect we aim to identify in this chapter.

2 *Contextual social effects*, that is, the effects of one's peers' exogenous characteristics on one's outcome. In our study, this refers to the direct effect of neighbours' characteristics on a household's probability of being TASAF informed. For instance, if neighbours have a radio that broadcasts information about TASAF, and if neighbours listen to that radio together, then the correlation in TASAF awareness would be the result of contextual social effects rather than of being driven by endogenous peer effects.

3 *Correlated effects,* which capture the possibility that households living in the same neighbourhood may behave in the same way simply because they have similar characteristics, or because they face similar conditions. Such correlated effects include sorting of similar households into specific neighbourhoods. I will discuss correlated effects more extensively below.

Since I expect each of these effects to be present in the data, I will follow Manski's (1993) popular linear-in-means model of social interactions, incorporating these effects into one empirical framework.

Specifically, I will consider the following linear probability model for whether or not a household is informed about TASAF:

$$I_{lji} = \alpha + \beta \bar{I}_{P_{lji}} + x_{lji}\gamma + \bar{x}_{lji}\delta + \varepsilon_{lji} \qquad (4.1)$$

where I_{lji} is a dummy for whether or not an elderly/disabled household *i* in hamlet *j* in village *l* had ever heard of TASAF at the time of the survey. Assume that each household has a neighbours set P_{lji}, defined by the set of neighbours potentially affecting household *i*'s TASAF awareness, with $i \notin P_{lji}$. The neighbours set P_{lji} may include both vulnerable and non-vulnerable households. The term $\bar{I}_{P_{lji}}$ is the weighted average information set of the households in P_{lji}.[52] Its coefficient β measures the endogenous peer effect of interest in this study. Vector x_{lki} contains a set of socioeconomic individual household characteristics, described below. Vector $\bar{x}_{lji}$ is the weighted average of *x* over the neighbours in P_{lki}. Its coefficient δ measures the contextual peer effects. Finally, parameter ε_{lji} captures the unobserved correlated effects, which I will discuss in more detail below.

Vector x_{lki} controls for the following household characteristics: asset ownership (radio, phone, vehicle, watch/clock, sofa); amenities (improved roof,[53] improved latrine[54]); basic consumption (whether or not the household had eaten meat or drunk milk in the 7 days prior to the survey); whether or not the household head had had primary education; the household's degree of political connectedness (whether or not any household member held political office in the village, and whether or not any household member was related to the political elite); its degree of political activity (two dummies for whether or not any household member attended a village meeting in the

[52] As noted in section 4.3.2, the observations are weighted by the probability of being sampled.
[53] Improved roofs are made of iron, concrete, cement, roofing tiles or asbestos.
[54] Improved latrines include covered or ventilated latrines.

previous 6 months and previous 3 months prior to the survey); its family ties in the village (whether or not the household had any relatives living in the village); and a set of household demographics (household size, age and gender of the household head, a dummy for whether the head belonged to the main tribe and main religion of the village, and a dummy for whether the main occupation of the household head was in the agricultural sector). Given that the household information was collected during a listing exercise, in which households were visited with no prior notification, in some incidences information was provided by a household member other than the household head, if the latter was not present during the interview. The other member was either the spouse of the head or another adult household member. This happened for 40% of the elderly/disabled households. Given the possibility that the household head is better informed about welfare programmes, I will control for whether or not the survey respondent was the household head.

4.4.2. Empirical challenges and identification strategy

Reflection problem

The challenge of disentangling contextual and endogenous effects in a linear-in-means model of social interactions was first highlighted by Manski (1993), who gave it the now commonly used term 'reflection problem'. Moffitt (2001) and Blume et al. (2010) provide further discussions of this problem, which is essentially a problem of identification in a simultaneous equation model (SEM). The problem arises because a household and its peers are likely to simultaneously affect each other. Given that outcomes are observed in equilibrium, it is difficult to know whether an individual's outcome is the cause or the effect of the peer's outcome.

Bramouille et al. (2009) and De Giorgi et al. (2010) developed a strategy that allows one to estimate endogenous and contextual peer effects in the context of 'partially overlapping reference groups'. Since its development, this strategy has been frequently

used in the non-experimental empirical literature on peer effects (Guiso and Schivardi, 2007; Bayer et al., 2009; Calvo-Armengol et al., 2009; Mas and Moretti, 2009; Helmers and Patnam, 2012; Collin, 2013). Peer groups of two observations are partially overlapping if they do not perfectly coincide. Such social networks ensure the occurrence of *intransitive triads,* which are sets of three households *i*, *j* and *k*, such that *i* is affected by *j* and *j* is affected by *k* but *i* is not affected by *k*. Bramouille et al. (2009) show how the presence of such intransitive triads allows one to deal with the reflection problem: the characteristics of 'excluded peers', that is, peers of peers who are neither one's own peers nor oneself, can be used as instruments for the average outcome of the peers in the SEM equation. In the example of the intransitive triad, the characteristics of *k* can be used as instruments for the outcome of *j*, because they will only have an indirect effect on the outcome of *i* through the effect on the outcome of *j*.

In our context, the set of nearest neighbours of household A is unlikely to perfectly coincide with the set of nearest neighbours of neighbouring household B. The presence of such intransitive triads allows us to use the method suggested by Bramouille et al. (2009) and De Giorgi et al. (2010). Specifically, I will use the average exogenous characteristics of a household's neighbours' neighbours set as identifying instruments for $\bar{I}_{P_{ij}}$.[55] The validity of the exclusion restrictions will be discussed below.

[55] Note that, in contrast to the claim made by De Giorgi et al. (2010), even when the SEM is identifiable thanks to the presence of excluded peers, it is still necessary to instrument for the endogenous effect, even in the absence of correlated effects. De Giorgi et al. (2010, page 18) state that '[…] I will discuss how to exploit such excluded peers in an IV strategy that addresses the potential endogeneity due to correlated effects, but it is important to clarify now that […] we would not need to instrument if the only identification issue were reflection'. This is because, even if the model is identifiable using excluded peers as instruments, it can easily be shown by writing out the reduced forms that the error term in the structural equation is correlated with the endogenous explanatory variable in SEM and that OLS would yield biased and inconsistent estimates.

Correlated effects

We can rewrite equation (4.1) to specify the different types of unobserved effects that are likely to be present in our study:

$$I_{lji} = \alpha + \beta \bar{I}_{P_{lji}} + x_{lji}\gamma + \bar{x}_{lji}\delta + \eta_l + \chi_{lj} + \theta_{P_{lji}} + \mu_{lji} + \omega_{lji} \qquad (4.2)$$

where ω_{lji} is assumed to be a random independently and identically distributed component, that is, $E\left[\omega \mid \bar{I}_P, x, \bar{x}\right] = 0$. It is reasonable to assume that unobserved covariates at the household level, μ_{lji} (such as a household's interest in politics), are not correlated with observed outcomes at the neighbours set level, and therefore I do not expect the presence of μ_{lji} in equation (4.2) to bias the results. However, error components η_l, χ_{lj} and $\theta_{P_{lji}}$ capture three types of unobserved covariates that are potentially correlated with the peer group outcome $\bar{I}_{P_{lji}}$:

1 η_l: unobserved village covariates – residents of a specific village are expected to be jointly affected by various village-specific factors that contribute to TASAF awareness. As highlighted in section 4.2.2, village leaders played a crucial role in the dissemination of information about TASAF procedures. Some village officials may have been more effective in organising village meetings than others. Moreover, the village officials were informed by district officials, whose work may have been more productive in some villages than in others. Such intravillage differences are unobserved to the researcher and are captured by η_l.

2 χ_{lj}: unobserved hamlet covariates – within each village, households living in the same hamlet may be subject to similar information shocks. There are 705 hamlets in total across the 100 sampled villages, with an average of seven hamlets per village. The average hamlet in the sample contains 28 elderly/disabled households (see Appendix 4.A.1). Some hamlets are more

remote than others, making it more difficult for their inhabitants to attend village meetings. Also, most hamlets in Tanzania have a hamlet leader, who is empowered to organise meetings at the hamlet level, or to go from door to door to inform people within his/her hamlet. Moreover, it is possible that households with similar unobserved characteristics sort themselves into specific hamlets. Such common unobserved characteristics would also be captured by χ_{lj}. We asked all households in the sample whether or not they were blood related to specific village elite. These data allow me to look at the spatial spread of blood relatives of the village elite across the village.[56] Analysing these data, I find evidence in support of residential sorting based on family ties: 50% of the local blood relatives of a village elite kin live within the same hamlet as that kin. Given that there are on average seven hamlets per village, this is a remarkable result.

3 $\theta_{P_{lji}}$: unobserved neighbours set covariates – households may have characteristics in common with the households in their neighbours set that affect both their own degree of TASAF information and that of their neighbours. For instance, households living near a noticeboard may each be more likely to be informed about TASAF, independently of the presence of peer effects. Moreover, to the extent that households sort themselves into neighbours sets *within* the hamlet, vector $\theta_{P_{lji}}$ will also include common characteristics that are driven by residential sorting. In the context of Tanzania's villagisation movement in the 1970s, most rural land was assigned to households by the government. Since then, plots have been passed on from generation to

[56] On average in a village, there are four (sampled) households blood related to the village leader and two (sampled) households blood related to the Village Executive Officer. On average per hamlet, there are three (sampled) households blood related to the hamlet leader.

generation, usually from father to son. Most of the older rural villagers therefore have little choice as to who their neighbours are, except by donating parts of their plots to their children. This means that if at all households sort themselves within hamlets, this is likely to be at the nearest neighbours set level. Using again the GPS data on the dwellings of the blood relatives of the village elite, I find that on average two out of 10 of their nearest (sampled) neighbours are, indeed, blood related to themselves.

To deal with correlated effects at the village level (η_l) I will include village fixed effects in my empirical specifications. The unique size of the dataset also allows me to deal with correlated effects at the hamlet level, χ_{lj}, by controlling for hamlet fixed effects (in specifications in which I drop the village fixed effects), something that is rarely feasible in other studies owing to insufficient variation within hamlets.[57]

Focusing on within-hamlet variation in households' TASAF awareness also deals with the bias caused by residential sorting into hamlets. Assuming that controlling for hamlet fixed effects captures all shocks simultaneously faced by a household and its excluded peers, instrumenting $\bar{I}_{P_{lji}}$ by the average characteristics of the excluded peers also deals with correlated effects at the neighbours set level. These latter effects include those resulting from possible residential sorting *within* hamlets that is driven by passing on land from generation to generation.

My 2SLS identification strategy crucially relies on the validity of the exclusion restrictions. Specifically, it relies on the assumption that excluded neighbours do not affect an elderly/disabled household's awareness of TASAF other than through passing such information on to their own neighbours. Whether or not this identifying

[57] Controlling for hamlet fixed effects is more effective in my study than adding component fixed effects (which is commonly done in the literature) since the average component in my study is much larger than the average hamlet.

assumption is valid will depend on whether or not the peer group is appropriately defined. I refer back to section 4.3.2 for a more detailed motivation for the choice of the k nearest neighbours as the elderly/disabled households' relevant peers for effective TASAF information exchange. In sum, I argued that effective transmission of information about uncommon activities such as new CDD programmes (as opposed to more familiar matters such as family or agriculture) requires frequent interactions with well-informed individuals, whom one can trust. Constrained by their physical and socioeconomic conditions, most elderly and disabled in rural Tanzania rely on their geographically nearest neighbours for repeated information exchange. This is especially true in the context of rural Tanzania, where people's blood relatives tend to live nearby. Although VGs surely also engage in interactions of some sort with other villagers, such as chats on their monthly visits to the village markets, I do not expect such infrequent interactions to significantly affect their knowledge about TASAF.

Of course, the validity of these assumptions depends on the actual degree of mobility of the elderly and disabled observations in our sample. In section 4.6.2 I will test the robustness of my results to the dropping of observations that are possibly less geographically restricted. Moreover, in section 4.5.2 I will test the results against various neighbours sets sizes. This is important, since the validity of the exclusion restrictions will depend on the choice of k: defining the neighbours set as the 10 nearest neighbours when the 15 nearest neighbours set is in fact the relevant neighbours set would invalidate the identifying assumptions. Section 4.6.2 will address a few other identification concerns.

Exclusion bias

In the previous chapter, I derived a set of theoretical predictions about the magnitude of the exclusion bias, which affects OLS estimation in empirical social interaction models. In my current study on peer effects in development programme awareness, a household

is systematically excluded from its peer group's average outcome calculations. In view of the discussion in Chapter 3, I therefore expect the downward exclusion bias to manifest itself in my OLS estimation results. Furthermore, as peer groups in this application are defined based on geographic distance, their formation is bound to be correlated with hamlet clustering. Therefore, for reasons set out in Chapter 3, I expect the exclusion bias to be more severe in OLS estimation models adding hamlet fixed effects than OLS estimation models adding village fixed effects.

We also know, from Chapter 3, that my 2SLS estimation strategy (using exogenous characteristics of excluded peers as instruments while controlling for the household's own exogenous characteristics) eliminates the exclusion bias. However, in order to better interpret the results, it is still worth deriving the magnitude of the exclusion bias and correct the OLS estimates accordingly. Doing so would allow me to distinguish the exclusion bias from other sources of bias, which would improve my understanding of the extent to which unobserved heterogeneity is affecting my results. The results of this exercise are discussed in Appendix 4.A.3. .

Spatial dependence of observations

Basic results of spatial econometrics encourage us to adjust standard errors for spatial dependence of observations. However, generalized method of moments (GMM) estimations such as Conley's routine (1999), which allows for arbitrary spatial correlation between observations, are computationally very cumbersome for datasets with 19,392 observations. Therefore, in my main specifications, I ignore spatial dependence within hamlets and report standard errors that are clustered at the hamlet level. In section 4.6.1 I test the robustness of my results to the controlling for spatial dependence within hamlets, by randomly selecting 20 out of the 100 villages in the sample and running regressions using Conley's (1999) GMM estimation on the reduced sample. I refer to section 4.6.1 for more details about this method.

4.5. Results

4.5.1. Average peer effects

Table 4.3 presents the OLS estimation results of the basic model (equation (4.1)), in which I define the peer group as the set of 10 nearest neighbours. Column (1) shows the results of a regression without any controls. In each of the remaining columns (2)–(4), the regressions include the full set of a household's own characteristics, vector *x*. Column (2) shows the OLS estimation results of a village fixed effects regression excluding the contextual effects. All coefficient estimates of the household's own characteristics have the expected signs. Households that attended at least one village meeting in the previous 6 months are 8 percentage points (pp) more likely to be informed. Households that attended a village meeting in the previous 3 months are an additional 4% more likely to be informed. Moreover, educated households are, *ceteris paribus*, 13% more likely to have heard about TASAF, and those that hold political office are 16% more likely to be informed. Female-headed households are 8 pp less likely to be informed about TASAF. Interestingly, survey respondents who are head of the household, as opposed to the head's spouse or any other adult household member, are 14 pp more likely to have heard about TASAF. This confirms our earlier conjecture that household heads are generally more informed about development programmes than any other adult members of the household. Intra-household communication about development programmes clearly is imperfect.

The OLS estimate of the endogenous peer effect in column (2) suggests that, for each additional informed neighbour in its set of 10 nearest sampled neighbours, a household is approximately 3.9 pp more likely to be informed about TASAF. Controlling for (observable) contextual social interaction effects in column (3) slightly increases the OLS estimate of the endogenous social interaction effect to 4.2 pp.

Table 4.3 Determinants of being informed: 10 nearest neighbours specification (OLS)

Dependent variable = 1 if informed	(1)	(2)	(3)	(4)
Endogenous peer effect: number of nearest 10 neighbours who are informed	0.045*** (0.027)	0.039*** (0.003)	0.042*** (0.003)	0.007* (0.004)
HH's own characteristics:				
1 = owns a radio		0.058*** (0.009)	0.058*** (0.009)	0.057*** (0.009)
1 = owns a mobile phone		0.042*** (0.009)	0.044*** (0.009)	0.044*** (0.009)
1 = owns a vehicle		–0.005 (0.010)	–0.005 (0.010)	–0.004 (0.011)
1 = owns a watch or clock		0.031*** (0.007)	0.032*** (0.007)	0.035*** (0.007)
1 = owns a sofa		0.031*** (0.009)	0.032*** (0.009)	0.028*** (0.009)
1 = dwelling has improved roof		0.011 (0.012)	0.004 (0.013)	0.005 (0.013)
1 = dwelling has improved latrine		0.028*** (0.008)	0.032*** (0.009)	0.035*** (0.009)
Household size		0.009*** (0.001)	0.009*** (0.001)	0.009*** (0.001)
Age head		–0.003*** (0.000)	–0.003*** (0.000)	–0.003*** (0.001)
1 = employed		0.040*** (0.012)	0.040*** (0.012)	0.041*** (0.012)
1 = member main tribe village		–0.020* (0.010)	–0.024** (0.012)	–0.022* (0.012)
1 = member main religion village		–0.007 (0.009)	–0.009 (0.009)	–0.008 (0.009)
1 = has eaten meat in past 7 days		0.031*** (0.009)	0.028*** (0.009)	0.026*** (0.009)
1 = has drank milk in past 7 days		0.013 (0.008)	0.014* (0.008)	0.011 (0.008)
1 = head with primary education		0.131*** (0.008)	0.130*** (0.009)	0.126*** (0.009)
1 = female headed		–0.084*** (0.010)	–0.082*** (0.010)	–0.085*** (0.010)
1 = attended village meeting in previous 3 months		0.040*** (0.010)	0.055*** (0.011)	0.061*** (0.011)
1 = attended village meeting in previous 6 months		0.082*** (0.013)	0.089*** (0.013)	0.085*** (0.014)
1 = blood related to village elite		0.045*** (0.012)	0.045*** (0.013)	0.046*** (0.014)
1 = holds political office		0.156*** (0.015)	0.154*** (0.015)	0.152*** (0.015)
1 = has relatives in village		0.017** (0.007)	0.016** (0.007)	0.020** (0.007)
1 = respondent household head		0.137*** (0.011)	0.138*** (0.011)	0.141*** (0.011)
Contextual peer effects	No	No	Yes	Yes
Hamlet fixed effects	No	No	No	Yes
N	19,392	19,392	19,392	19,392
Adjusted R^2	0.122	0.244	0.246	0.143

Notes: Dependent variable is a dummy = 1 if HH was informed. Standard errors (shown in parentheses) are adjusted for clustering of observations at village level (as some networks extend to individuals outside the hamlet). (1)-(3) include village fixed effects. Regressions with contextual effects include the full set of neighbours set characteristics. Although the regressions include the proportion of 10 nearest neighbours being informed as a measure of the endogenous social interaction effect, the table presents the marginal endogenous social interaction effect of one additional neighbour being informed. HH, household. ***p<0.01, **p<0.05, *p<0.1.

Controlling for hamlet fixed effects in column (4), in addition to controlling for contextual effects, significantly reduces the endogenous peer effect to approximately 1 pp per additional informed sampled neighbour.

As discussed in section 4.4, we have good reasons to believe that the OLS estimates are biased and inconsistent. Table 4.5 below addresses this bias by instrumenting the endogenous average information set of the peers by the exogenous characteristics of the excluded neighbours. Although the results remain similar when I include the complete set of characteristics of excluded neighbours as instruments (not reported), the table shows the results for regressions only including the instruments that are significant at the 5% level in the 2SLS first-stage regression. The resulting set of four instruments includes: a dummy for whether the household had an improved roof; whether the household was female headed; whether the household had relatives in the village; and whether the respondent to the survey was the household head.

Table 4.4 shows the first-stage regression results for the 2SLS regression controlling for hamlet fixed effects. Each instrument is significant at either the 1% or 5% significance level. The current literature lacks concrete results for testing for weak instruments when errors are not identically and independently distributed: on the one hand, the Craig–Donald (1993) (C-D) Wald F-statistic is not appropriate, as it relies on the assumption of homoscedastic errors; on the other hand, when using the Kleibergen–Paap (2006) (K-P) rank Wald F-statistic, although it does adjust for heteroscedasticity, autocorrelation and clustering of standard errors, it is difficult to interpret its magnitude, as its critical values have not yet been defined.[58] For these reasons, Table 4.4 shows both the C-D and the K-P test statistics. The C–D F-statistic is 26.54, which is above the Stock–Yogo (2005) critical value for 5% maximal instrumental variables (IV) bias

[58] Its critical values have not been computed as they depend on the type of violation of the independent and identically distributed assumption, which in turn depends on the nature of the study (Bazzi and Clemens, 2009).

Table 4.4 Two-stage least squares first-stage regression

	Proportion of 10 nearest neighbours who are informed
Excluded peers' characteristics:	
1 = dwelling has improved roof	0.030*** (0.011)
1 = female headed	0.024** (0.011)
1 = has relatives in village	0.024** (0.009)
1 = respondent household head	0.031*** (0.010)
Household's own characteristics	Yes
Contextual peer effects	Yes
Hamlet fixed effects	Yes
N	19,392
Joint F-statistic	83.48
Adjusted R^2	0.16
Craig–Donald Wald F-statistic	26.54
Kleibergen–Paap Wald F-statistic	8.10
Kleibergen–Paap LM statistic	29.03
p-value for null of underidentification	0.00
Hansen J *p*-value	0.23

Notes: Standard errors (shown in parentheses) are adjusted for clustering of observations at the village level. Contextual effects include the full set of neighbours set characteristics.
TASAF, Tanzania Social Action Fund.
*p<0.10; **p<0.05; ***p<0.01.

toleration relative to OLS and above the Stock–Yogo (2005) critical value for 10% maximal IV size distortion. Using the same Stock–Yogo (2005) critical values, the K-P statistic is above the 20% maximal IV relative bias and just below the 25% maximal IV size threshold. A *p*-value of 0.00 associated with a Kleibergen–Paap rank LM statistic of 29.03 implies a strong rejection of the null of underidentification of the model.

In my preferred specification, column (2) in Table 4.5, the estimated endogenous peer effect indicates an increase of 7.7 pp in the probability of being TASAF informed, per additional informed sampled neighbour in the neighbours set. This means that an elderly/disabled household that has five out of 10 sampled neighbours who are informed (e.g. because they were able to attend village meetings) is, *ceteris paribus*, 38 pp more likely to be informed itself than households living among uninformed households. This endogenous peer effect is quite significant in magnitude, especially when compared with the effect of a household's own characteristics on TASAF programme awareness.

Table 4.5 Determinants of being informed: 10 nearest neighbours specification (2SLS)

Dependent variable = 1 if informed	(1)		(2)	
Endogenous peer effect: number of nearest 10 neighbours who are informed	0.092***	(0.027)	0.077**	(0.030)
Household's own characteristics:				
1 = owns a radio	0.060***	(0.009)	0.057***	(0.008)
1 = owns a mobile phone	0.047***	(0.009)	0.049***	(0.009)
1 = owns a vehicle	–0.005	(0.010)	–0.005	(0.009)
1 = owns a watch or clock	0.031***	(0.007)	0.034***	(0.008)
1 = owns a sofa	0.031***	(0.010)	0.031***	(0.011)
1 = dwelling has improved roof	–0.000	(0.013)	–0.000	(0.010)
1 = dwelling has improved latrine	0.033***	(0.009)	0.033***	(0.009)
Household size	0.009***	(0.001)	0.009***	(0.001)
Age head	–0.003***	(0.000)	–0.003***	(0.000)
1 = employed	0.042***	(0.012)	0.043***	(0.012)
1 = member main tribe village	–0.026**	(0.012)	–0.025**	(0.011)
1 = member main religion village	–0.011	(0.009)	–0.010	(0.009)
1 = has eaten meat in past 7 days	0.027***	(0.009)	0.027***	(0.008)
1 = has drank milk in past 7 days	0.013*	(0.008)	0.012	(0.008)
1 = head had at least 1 year of primary Education	0.123***	(0.009)	0.122***	(0.009)
1 = female headed	–0.087***	(0.010)	–0.090***	(0.009)
1 = attended village meeting in previous 3 Months	0.058***	(0.011)	0.062***	(0.010)
1 = attended village meeting in previous 6 months	0.085***	(0.013)	0.083***	(0.011)
1 = blood related to village elite	0.040***	(0.013)	0.042***	(0.013)
1 = holds political office	0.153***	(0.015)	0.152***	(0.013)
1 = has relatives in village	0.013*	(0.007)	0.014*	(0.008)
1 = respondent household head	0.137***	(0.011)	0.136***	(0.008)
Contextual peer effects	Yes		Yes	
Hamlet fixed effects	No		Yes	
N	19,392		19,392	
Adjusted R^2	0.217		0.101	

Notes: Dependent variable is a dummy = 1 if household was informed about TASAF. Standard errors (shown in parentheses) are adjusted for clustering of observations at the village level (as some networks extend to individuals outside the hamlet). Regressions include the full set of neighbours set characteristics. Although the regressions include the proportion of 10 nearest neighbours being informed as a measure for the endogenous social interaction effect, the table presents the marginal endogenous social interaction effect of one additional neighbour being informed. The four characteristics of excluded peers used as instruments for 2SLS are: dummy for improved roof, female-headed households, whether the household has relatives in the village and whether the respondent was the household head.
*p<0.10; **p<0.05; ***p<0.01.

For instance, education and village meeting attendance increase the probability of being TASAF informed, *ceteris paribus*, by, respectively, 12 and 14 pp.

Up to now, we have interpreted the social interaction results in terms of the effect of the average characteristics of the 10 nearest *sampled* neighbours. In section 4.3.2, I discussed why the set of 10 nearest *sampled* neighbours is actually a proxy for the 20 *actual* nearest neighbours in the true population. This means that, in terms of the true population, we should actually interpret the results in column (2) of Table 4.5 as follows: each additional informed household in a vulnerable household's 20 nearest neighbours set increases its probability of being informed about TASAF by almost 4%. However, given that the main purpose of this chapter is to identify the importance of neighbourhood peer effects, rather than determining the exact number of nearest neighbours that matters, this difference in interpretation is not all that crucial for the main findings of the study.

Moreover, it should be noted that Table 4.5 presents average peer effects. In spatial social interactions models, however, the individual peer effect is expected to depend on the geographical distance between the residence of the household and that of its peers. In our dataset, this distance ranges from 16 meters up to 969 meters (note that we only consider peers who live within 1 km range in our analysis), with an average distance of 186 meters. In general, we expect the peer effects to die out with distance. In Appendix 4.A.2, I present the results of a regression adding an interacted term between the peer effect of the ten nearest neighbours and the average distance of those ten nearest neighbours. The estimation results confirm, indeed, that the peer effect becomes significantly smaller as the distance of the peers grows larger.

Finally, to illustrate the practical importance of the exclusion bias (see Chapter 3) in the OLS estimates of the endogenous peer effect presented in Table 4.3, in Appendix 4.A.3 I use simulations to derive the magnitude of the exclusion bias in this specific empirical application. Doing so, I show that a significant part of the difference between

the OLS estimates and the 2SLS estimates of the peer effect are driven by the exclusion bias, rather than by unobserved correlated effects.

4.5.2. Varying the size of the neighbours set

Table 4.6 presents the 2SLS coefficient estimates of the endogenous peer effect equivalent to regression (2) in Table 4.5, for the five, 10, 15 and 20 nearest neighbours set specifications. As noted above, we have to drop households for which not all k neighbours live in the 1-kilometre range and also those that do not have any excluded neighbours. For comparison purpose, I run the four different regressions on the same sample, that is, one that contains only households with suitable neighbours sets for each value of k.

Whereas the magnitude of the endogenous peer effect (per neighbour) is fairly similar for the ten, fifteen and twenty nearest neighbours specifications, the effect more than quadruples when we consider the five nearest neighbours. Although this result could possibly indicate that the five nearest neighbours are most effective in transmitting information about development programmes (especially given that they tend to include blood relatives of the household), the Hansen J p-value of 0.075 and the negative adjusted R^2 suggest that the exclusion restrictions may not be valid in the five nearest neighbours specification. As discussed in section 4.4.2, if peer groups are too narrowly defined, the identifying assumptions may be violated. For specifications assuming larger social interaction bands (i.e. k = 10, 15 or 20), the Hansen J p-values are consistent with valid exclusion restrictions.

Unsurprisingly, the larger the neighbours set we consider, the smaller the C-D Wald F-statistic and the K-P Wald F statistic.[59] Weak instruments lead to lower precision of

[59] The larger the neighbours sets, the higher the probability that the instruments rely on excluded neighbours for which the information transmission effect is relatively weak.

coefficient estimates, which may explain the finding that the significance level decreases for higher values of k, despite the magnitude of the effect remaining similar.

Table 4.6 Endogenous peer effect for varying neighbours set sizes

Dependent variable = 1 if informed	$k = 5$ (1)	$k = 10$ (2)	$k = 15$ (3)	$k = 20$ (4)
Endogenous social interaction effect: number of nearest k neighbours who are informed about TASAF	0.284*** (0.086)	0.064** (0.032)	0.053* (0.028)	0.051 (0.032)
Household's own characteristics	Yes	Yes	Yes	Yes
Contextual peer effects	Yes	Yes	Yes	Yes
Hamlet fixed effects	Yes	Yes	Yes	Yes
N	17,495	17,495	17,495	17,495
Adjusted R^2	–0.168	0.117	0.114	0.100
Craig–Donald Wald F-statistic	7.23	24.44	15.18	11.27
Kleibergen–Paap Wald F-statistic	3.89	10.01	3.63	2.24
Hansen J p-value	0.075	0.635	0.448	0.298

Notes: Dependent variable is a dummy = 1 if household was informed about the Tanzania Social Action Fund (TASAF). Standard errors (shown in parentheses) are adjusted for clustering of observations at the village level. Contextual effects include the full set of neighbours set characteristics. Although the regressions include the proportion of 10 nearest neighbours being informed as a measure for the endogenous social interaction effect, the table presents the marginal endogenous social interaction effect of one additional neighbour being informed. The four characteristics of excluded peers used as instruments for two-stage least squares are: dummy for improved roof, female-headed households, whether the household has relatives in the village and whether the respondent was the household head. *p<0.10; **p<0.05; ***p<0.01.

4.5.3. Heterogeneous peer effects

This section examines the presence of heterogeneity in endogenous peer effects on programme awareness. Specifically, the aim is to identify (i) the type of disabled/elderly households that are most and least responsive to information in the neighbourhood and (ii) the type of well-informed neighbours who are most effective in transmitting information to disabled/elderly households.

The answers to both these questions will possibly depend on the degree of socioeconomic similarity between a household and its neighbours. The literature on peer group formation has identified gender, age, tribe, religious affiliation, wealth, occupation, education level and kinship as characteristics other than proximity that matter for the formation of information links (Conley and Topa, 2002; Jackson, 2008;

Conley and Udry, 2010; Van den Broeck and Dercon, 2011). Although this literature has focused on other types of information links, such as relationships among African farmers sharing information about agricultural technologies, it is useful to start from the same set of characteristics and investigate which ones also matter in the diffusion of information about development programmes. In addition, I will consider the extent to which the holding of political positions in the village matters for information diffusion through word of mouth.

Table 4.7 reports on the estimation results when interaction terms are added to the main specification in column (2) of Table 4.5.[60] In panel A, I interact the endogenous peer effect with disabled/elderly household *i*'s *own* socioeconomic characteristics. Panel B shows the results of regressions adding interactions of the endogenous peer effect with the same set of socioeconomic characteristics but now with those that apply to the average of the 10 nearest neighbours. In Panel C, I add a triple interaction term, in addition to the bivariate interactions added in panels A and B (controlling for the interaction between the household's own characteristic and that of the average neighbour; not reported).

[60] Instruments for the interaction terms are computed by taking the product of each of the main instruments and the interaction term of interest. Table 4.7 shows the results when the interaction terms are added individually to the base regression. The results remain similar, however, in a model jointly adding all interaction terms (not reported).

Table 4.7 Heterogeneous peer effects in the probability of being informed (2SLS)

Panel A: Interactions with household's *own characteristics*										
	Female headed (1)	Age head (continuous) (2)	Main religion (3)	Main tribe (4)	Primary education (5)	Head is employed (6)	Eaten meat in past 7 days (7)	Owns phone (8)	Relatives in village (9)	Political office (10)
Endogenous peer effect	0.070*** (0.030)	0.159*** (0.033)	0.070** (0.029)	0.078*** (0.033)	0.064** (0.029)	0.030 (0.034)	0.086*** (0.029)	0.078*** (0.029)	0.063*** (0.030)	0.079*** (030)
Interaction effect	–0.009 (0.009)	–0.001*** (0.000)	0.027*** (0.010)	0.005 (0.012)	0.024** (0.010)	0.036** (0.015)	–0.019** (0.009)	–0.011 (0.009)	0.020** (0.008)	–0.027 (0.018)
N	19,392	19,392	19,392	19,392	19,392	19,392	19,392	19,392	19,392	19,392

Panel B: Interactions with household's *neighbours set characteristics*										
	Female headed (1)	Dummy age head >55 (2)	Main religion (3)	Main tribe (4)	Primary education (5)	Head is employed (6)	Eaten meat in past 7 days (7)	Owns phone (8)	Relatives in village (9)	Political office (10)
Endogenous peer effect	0.085*** (0.026)	0.063** (0.027)	0.080*** (0.031)	0.085*** (0.045)	0.070** (0.032)	0.148** (0.070)	0.087*** (0.028)	0.079*** (0.028)	0.055* (0.029)	0.072** (0.030)
Interaction effect	–0.012 (0.026)	0.012 (0.008)	0.019 (0.020)	–0.002 (0.031)	0.003 (0.025)	–0.077 (0.059)	0.055 (0.022)	0.008 (0.022)	0.042** (0.018)	0.123** (0.054)
N	19,392	19,392	19,392	19,392	19,392	19,392	19,392	19,392	19,392	19,392

Notes: Each regression controls for a household's own characteristics, household's neighbours set characteristics and hamlet fixed effects. Although the regressions include the proportion of 10 nearest neighbours being informed as a measure for the endogenous social interaction effect, the table presents the marginal endogenous social interaction effect of one additional neighbour being informed. Standard errors (shown in parentheses) are adjusted for clustering of observations at the village level. In all columns except for (1) and (6) the interacted variable is binary (0 or 1). The four characteristics of excluded peers used as instruments for two-stage least squares are: dummy for improved roof, female-headed households, whether the household has relatives in the village and whether the respondent was the household head. TASAF, Tanzania Social Action Fund.

*p<0.10; **p<0.05; ***p<0.01.

Table 4.7 *Continued*

Panel C: Triple interactions with household's own and household's neighbours set characteristics										
	Female headed (1)	Age head (2)	Main religion (3)	Main tribe (4)	Primary education (5)	Head is employed (6)	Eaten meat in past 7 days (7)	Owns phone (8)	Relatives in village (9)	Political office (10)
Endogenous peer effect	0.092*** (0.027)	–	0.064** (0.030)	0.070 (0.057)	0.062* (0.037)	0.094 (0.229)	0.095*** (0.029)	0.082*** (0.028)	0.063** (0.029)	0.068** (0.029)
Interaction with HH's own characteristic	–0.033* (0.012)	–	0.019 (0.025)	–0.005 (0.041)	0.007 (0.031)	0.036 (0.242)	–0.037* (0.020)	–0.046*** (0.016)	–0.017 (0.024)	–0.023 (0.024)
Interaction with HH's neighbours set characteristic	–0.043 (0.032)	–	–0.051 (0.028)	–0.004 (0.056)	–0.020 (0.036)	–0.062 (0.239)	–0.005 (0.027)	–0.017 (0.025)	–0.003 (0.030)	0.139** (0.059)
Triple interaction (own + neighbours set)	0.078 (0.051)	–	0.020 (0.035)	0.018 (0.054)	0.065 (0.044)	–0.004 (0.253)	0.028 (0.037)	0.098** (0.040)	0.057 (0.037)	–0.118 (0.164)
Significance level of sum of three interaction terms	>10%	–	>10%	>10%	>10%	>10%	>10%	>10%	10%	>10%
N	19,392	19,392	19,392	19,392	19,392	19,392	19,392	19,392	19,392	19,392

Notes: Each regression controls for a household's own characteristics, household's neighbours set characteristics and hamlet fixed effects. In addition, regressions in panel C control for the interaction between a household's own and a household's neighbours set characteristics. Although the regressions include the proportion of 10 nearest neighbours being informed as a measure for the endogenous social interaction effect, the table presents the marginal endogenous social interaction effect of one additional neighbour being informed. Standard errors (shown in parentheses) are adjusted for clustering of observations at the village level. In all columns except for (1) and (6) the interacted variable is binary (0 or 1). The four characteristics of excluded peers used as instruments for two-stage least squares are: dummy for improved roof, female-headed households, whether the household has relatives in the village and whether the respondent was the household head.
*p<0.10; **p<0.05; ***p<0.01.

Households most responsive to information in the neighbourhood

The results in panel A of Table 4.7 suggest that gender, tribe, phone ownership and the holding of political positions in the village do not matter for the extent to which a household benefits from the availability of information in the neighbourhood. However, the results in panel C show that gender does matter when you take the characteristics of the neighbourhood into account: female-headed households benefited more from the presence of well-informed neighbours if they lived in an area with a relatively high proportion of female-headed households.

The results also indicate that elderly/disabled households that are headed by very old members are less likely to effectively capture information about TASAF through endogenous peer effects.[61] This may be due to memory problems related to very old age, or it may be due to very old people being less likely to go out and interact with other people in the neighbourhood.[62] Moreover, the results show that, in poorer neighbourhoods, information on TASAF was less likely to reach wealthier households through social interactions than their non-wealthy counterparts (with wealth being proxied by phone ownership and the number of times a household had eaten meat).[63] It is possible that wealthy individuals living in economically deprived areas prefer to obtain information through other means (e.g. by attending village meetings or reading newspapers). In predominantly wealthy neighbourhoods, on the other hand, wealthy

[61] The interaction term in this column makes use of the continuous age of the household head. Recall that the sample mainly consists of households with heads who are older than 55 years; therefore, this result applies to the oldest of the elderly.

[62] Whereas the sampled units of interest mainly consist of elderly households only, their neighbours may include both young and elderly headed households. This means that an interaction with a dummy for whether or not the neighbours are also headed by elderly households, the results of which are presented in column (2) of panel B in Table 4.7 and discussed below, are more meaningful than adding triple interactions. I therefore omit the respective triple interactions in panel C of Table 4.7.

[63] Similar results apply when I use sofa ownership or an improved roof as proxies for wealth (not reported).

households seem to have been as likely to obtain information on TASAF through interactions with neighbours as their more deprived counterparts.[64]

Households that belong to the main religious group in the village and/or that are headed by educated and/or employed individuals were significantly more likely to benefit from the presence of TASAF-informed households in the neighbourhood. This result is independent of the area in which these households lived. Similarly, households with blood relatives living in the village were more likely to be informed through neighbourhood effects. The latter finding may be a result of residential sorting based on family ties (see above). That is, people who have relatives living in the village are more likely to have relatives among their neighbours, and, as a result, are more likely to interact with them. If this interpretation is correct, we would expect this result to apply only to households whose neighbours *also* claimed to have blood relatives living in the village. This can be verified by adding triple interactions, which I do in panel C of Table 4.7. The results, shown in column (9), are, indeed, consistent with this hypothesis.

Neighbours most effective in the transmission of information

The results in column (9) in panel B of Table 4.7 show that households that have blood relatives living in the same village were more effective in disseminating TASAF information to their neighbours. As discussed above, this result is conditional on the household itself having blood relatives living in the village. As expected, households holding a political position in the village were significantly more likely to spread information in their neighbourhood (column (10)). Each additional political official in the neighbours set adds 1 pp to the endogenous social interaction effect.[65] However, the results in column (13) in panel C show that this only holds for households that do not

[64] This is tested by using a joint F-test on the sum of the estimates of the three interaction terms.
[65] In the data, the number of political neighbours within the 10 nearest neighbours set ranges from zero (57% of the households) to six (less than 1% of the households), with 10% of the households having two political neighbours.

hold a political office themselves (possibly because those households have other means of being informed and do not need to rely on local social interactions).

4.6. Robustness tests

4.6.1. Correcting for spatial dependence of standard errors within hamlets

In section 4.4, I explained that the use of estimation methods that control for spatial dependence within hamlets, such as the routine suggested by Conley (1999), becomes computationally cumbersome when the sample size is very large. However, the presence of such spatial dependence of standard errors may affect my inference results. To address this concern, I randomly select 20 out of 100 villages in the sample and run regressions using Conley's (1999) GMM estimation to adjust for spatial dependence in the reduced sample of 3685 observations. The Conley (1999) method allows for arbitrary spatial correlation between observations that is decreasing with distance and is zero beyond a specified cut-off point. I use the average distance to the tenth neighbour as the cut-off value.

Table 4.8 Robustness to Conley's (1999) correction of standard errors

Dependent variable = 1 if informed	2SLS (1)	2SLS (2)
Endogenous peer effect: number of nearest *k* neighbours who are informed	0.119*** (0.046)	0.119*** (0.047)
Adjusting for spatial dependence of standard errors within hamlet (using Conley's method, 1999)	No	Yes
N (reduced to 20 randomly sampled villages out of 100)	3685	3685

Notes: Dependent variable is a dummy = 1 if household was informed about the Tanzania Social Action Fund (TASAF). Standard errors are shown in parentheses. Standard errors in (1) are adjusted for clustering of observations at the village level. Each regression controls for contextual effects, individual household characteristics and hamlet fixed effects. Although the regressions include the proportion of 10 nearest neighbours being informed as a measure for the endogenous social interaction effect, the table presents the marginal endogenous social interaction effect of one additional neighbour being informed. The four characteristics of excluded peers used as instruments for 2SLS are: dummy for improved roof, female-headed households, whether the household has relatives in the village, whether the respondent was the household head.

2SLS, two-stage least squares.

*p<0.10; **p<0.05; ***p<0.01.

Using the reduced sample, Table 4.8 compares the results of a model in which I do correct for spatial dependence in standard errors within hamlets using the Conley (1999) method with a model in which I do not. In the latter model, I adjust the standard errors for clustering at the hamlet level. I find that the adjusted standard error of the coefficient estimate of the endogenous peer effect is almost identical in both cases. This suggests that, by clustering standard errors at the hamlet level in my main specifications, I am dealing with most of the spatial dependence of observations.

4.6.2. Other identification challenges

As discussed in section 4.4, for my 2SLS strategy to be valid, excluded neighbours should not affect a household's information set about TASAF other than through passing information on to their own neighbours. However, one might consider that not all elderly and disabled households are constrained to their nearest k neighbours for effective information exchange about development programmes such as TASAF. Some of them may be more mobile and frequently interact with people in the village living further away. Such observations would invalidate my IV instruments and consequently bias my 2SLS estimates upwards.

To address this concern, I test whether my results hold up to the dropping of observations for which the exclusion restriction is less likely to hold. As mentioned above, village meeting attendance in most villages in rural Tanzania is compulsory for all those older than 18 years. Elderly and disabled individuals who are physically too weak to attend are usually exempted. This means that restricting the sample to those households that did not attend village meetings is expected to retain the least mobile households. Similarly, the exclusion restriction is more likely to hold for the elderly/disabled households that do not own a vehicle (bicycle, car or truck) and those whose head is relatively old (e.g. older than 65 years). Moreover, as discussed in section 4.3.1, the data for this chapter come from a listing exercise in which we visited all

households in 100 villages in Tanzania. Because these listing visits were generally not announced to villagers, in 40% of the elderly and disabled households the household head was not at home at our survey visit (instead, we interviewed the spouse or another adult household member). Heads who were not at home during the surprise visit are expected to be relatively more mobile.

Column (1) in Table 4.9 copies the results from the unrestricted sample given in column (2) in Table 4.5. Columns (2)–(4) show the results for the set of samples restricted to those households we expect to be least mobile for reasons described in the previous paragraph. If the failure of the exclusion restriction, owing to the presence of mobile households, led to an upward bias in my estimate of the endogenous peer effect in Table 4.5, I would expect the estimate of the endogenous peer effect in Table 4.9 to

Table 4.9 Samples restricted to the least mobile households (HHs) (2SLS)

Dependent variable	= 1 if informed about TASAF					= 1 if informed about VG group
Sample restricted to:	All HHs	Not attending village meetings	Survey respondent was head	Does not own a vehicle	Head aged >65 years	All households (no restrictions)
	(1)	(2)	(3)	(4)	(5)	(6)
Endogenous peer effect	0.077**	0.074**	0.098***	0.115***	0.087**	0.157***
	(0.030)	(0.044)	(0.035)	(0.041)	(0.042)	
N	19,392	10,228	11,591	12,099	9,438	19,392
Kleibergen–Paap Wald F-statistic	8.095	5.397	6.772	5.333	5.899	3.811

Notes: Endogenous peer effect = number of nearest *k* neighbours who are informed. All regressions control for a household's own characteristics, contextual effects and hamlet fixed effects. Standard errors (shown in parentheses) are adjusted for clustering of observations at the village level. Although the regressions include the proportion of 10 nearest neighbours being informed as a measure for the endogenous social interaction effect, the table presents the marginal endogenous social interaction effect of one additional neighbour being informed. The four characteristics of excluded peers used as instruments for two-stage least squares (2SLS) are: dummy for improved roof, female-headed households, whether the household has relatives in the village and whether the respondent was the household head.
TASAF, Tanzania Social Action Fund.
*p<0.10; **p<0.05; ***p<0.01.

go down. Instead, I find that the estimates slightly increase, which is consistent with less mobile households being more reliant on neighbourhood interactions for access to information than mobile households.

Another concern is whether the answer to the question 'Have you personally ever heard of a development programme called TASAF?' is a good proxy for a household's degree of TASAF awareness. It could be argued that people, although knowing the activities of a programme, may have forgotten its name.[66] To address this worry, I use an alternative measure of TASAF awareness that is more specific about the activities that TASAF VGs supported. Specifically, I use as a dependent variable the answer to the question of whether the VG respondent had ever heard of any of the TASAF VGs in the village, where we listed the names under which the groups had been registered when asking this question. The names of these groups usually referred to the activity (such as 'pig keeping for the elderly'). Column (6) in Table 4.9 shows that the estimate of the peer effect increases to almost 16 percentage points when using this alternative measure.

Lastly, my base measure of TASAF awareness is predicated on whether the respondents themselves had personally ever heard about the programme. If the respondent had not heard about TASAF, this does not mean that no one else in the household had, assuming imperfect intra-household transmission of information. As mentioned earlier, the household head is usually the best informed person in the household, and his/her information set is therefore the best proxy for the household's neighbours set. We already noted that the estimated endogenous peer effect is not

[66] Our survey visit took place immediately after TASAF groups had been formed, when the programme was still a very hot topic. When people discussed the programme, they would always use its name. It is therefore very unlikely that respondents who had never heard about the programme would not know the programme's name. Probing more about the programme in qualitative discussions with respondents who had never heard about the programme suggested that they did, indeed, not know anything about the programme.

reduced when I restrict the sample to those households for which the head was the survey respondent (see column (3) in Table 4.9).

4.7. Conclusion

For targeting of CBD programmes to be effective, it is crucial that information about the programmes reaches all members of the eligible communities. In practice, however, many development programmes suffer from low public awareness. In rural Tanzania, VGs such as the disabled and the elderly often have limited access to primary sources of information (such as village meetings and newspapers) as a result of their physical and socioeconomic conditions. The aim of this chapter was to analyse the extent to which the presence of well-informed neighbours may compensate for this lack of direct access to information. I find that, on average, for each additional informed neighbour in a set of 10 sampled nearest neighbours, a disabled or elderly household was approximately 8 percentage points more likely to be informed about TASAF. To put this effect into context, it is useful to compare it with the impact of attending village meetings (i.e. the medium most heavily relied on by TASAF to sensitise the population): having two informed neighbours has a similar impact on the probability of being informed as has attending a village meeting (15% and 14%, respectively).

The finding of such significant endogenous peer effects implies the presence of large social multiplier effects in TASAF programme awareness. This result is important for anyone involved in development programme sensitisation campaigns. To illustrate, consider the effectiveness of a sensitisation intervention that distributes radios to a small subset of households in a village. In the absence of any peer effects, this policy would only affect the awareness of the beneficiary households. In the presence of contextual peer effects but no endogenous peer effects, the policy would not only affect the direct beneficiaries but also their direct peers, who benefit by listening to the radio of their

peers. In the presence of an endogenous peer effect in addition to the contextual peer effects, the intervention would reach not only the beneficiaries and their direct peers but also the peers of these peers through word of mouth. This example illustrates that, given the presence of endogenous peer effects, it becomes a more reasonable expectation for relatively cheap interventions, targeted only at subsets of a community, to substantially improve the awareness level of the entire population.

This chapter not only identified the presence of endogenous peer effects. Through the unique size of the dataset, it also allowed me to gauge the profile of elderly and disabled households that are most and least responsive to information present in the neighbourhood. I found that, on average, the economically worse-off households seem to be those that benefit most from the presence of well-informed neighbours. This is possibly because better off households have better access to primary sources of information, such as radio and newsletters and, therefore, rely less on peers for information. The chapter also identified specific groups that are less likely to be informed through social interactions and therefore should be targeted separately by complementary – specifically designed – sensitisation strategies. These groups include the oldest of the elderly members, female-headed households living in a male-dominated neighbourhood, those who do not have blood relatives living in the village, those who are uneducated and/or unemployed and/or those who do not belong to the main religion of the village. On the other hand, the study identified households with members holding political positions as being relatively more effective in the information diffusion process.

I should note that the external validity of the findings in this chapter is likely to be restricted to those CBD programmes that are targeted at elderly and disabled individuals. The pattern of social interactions may be different in the context of other development programmes. More generally, in addition to informing those in charge of

designing sensitisation campaigns, the study may also be of interest to those involved with programme monitoring. It highlights the importance of regularly sampling households from different parts of a community during the programme sensitisation phase, to evaluate the extent to which information effectively spreads to the various sections of the community and to adjust sensitisation programmes accordingly.

4.A. Chapter 4 Appendix

4.A.1. Population and sample information

Number of sampled districts		5	
Number of sampled villages per district		20	
Total number of sampled villages		100	
Total number of hamlets in sampled villages		705	
Number of hamlets per sampled village	Min.	2	
	Max.	15	
	Mean	7	

					True population	Sample
Total number of households (number in parentheses shows fraction of true population)	All				61,610	30,339 (49%)
	Elderly/disabled				22,957	19,916 (87%)
Number of households per village	All			Min.	116	72
				Max.	3761	964
				Mean	616	303
	Elderly/disabled			Min.	42	27
				Max.	641	580
				Mean	230	199
Number of households per hamlet	All			Min.	17	3
				Max.	655	196
				Mean	114	43
	Elderly/disabled			Min.	8	1
				Max.	114	135
				Mean	42	28
Distance between households within village (metres) (table shows *average* min., max. and mean of 100 villages)				Min.	2	4
				Max.	7662	8900
				Mean	1867	1800
Distance to the *k*th nearest household within village (table shows *average* min., max. and mean of 100 villages)		$k = 1$		Min.	2	4
				Max.	2471	922
				Mean	54	72
		$k = 5$		Min.	26	45
				Max.	2839	1296
				Mean	141	200
		$k = 10$		Min.	55	92
				Max.	2976	1513
				Mean	216	310
		$k = 15$		Min.	81	134
				Max.	3083	1681
				Mean	278	402
		$k = 20$		Min.	106	173
				Max.	3188	1852
				Mean	335	488

4.A.2. Interaction between peer effect and geographical distance of peers

Dependent variable = 1 if informed		
Endogenous peer effect: number of nearest 10 neighbours who are informed	0.086***	(0.028)
Interaction of peer effect with average geographical distance of ten peers (in meters)	-0.00002***	(0.0001)
Household's own characteristics:		
1 = owns a radio	0.057***	(0.008)
1 = owns a mobile phone	0.046***	(0.009)
1 = owns a vehicle	–0.005	(0.011)
1 = owns a watch or clock	0.034***	(0.007)
1 = owns a sofa	0.028***	(0.011)
1 = dwelling has improved roof	–0.003	(0.013)
1 = dwelling has improved latrine	0.032***	(0.009)
Household size	0.010***	(0.001)
Age head	–0.003***	(0.001)
1 = employed	0.043***	(0.012)
1 = member main tribe village	–0.024**	(0.012)
1 = member main religion village	–0.010	(0.009)
1 = has eaten meat in past 7 days	0.027***	(0.009)
1 = has drank milk in past 7 days	0.013	(0.008)
1 = head had at least 1 year of primary Education	0.122***	(0.009)
1 = female headed	–0.089***	(0.009)
1 = attended village meeting in previous 3 Months	0.062***	(0.011)
1 = attended village meeting in previous 6 months	0.084***	(0.014)
1 = blood related to village elite	0.040***	(0.014)
1 = holds political office	0.152***	(0.015)
1 = has relatives in village	0.013*	(0.007)
1 = respondent household head	0.138***	(0.011)
Contextual peer effects	Yes	
Hamlet fixed effects	Yes	
N	19,392	

Notes: Dependent variable is a dummy = 1 if household was informed about TASAF. Standard errors (shown in parentheses) are adjusted for clustering of observations at the village level (as some networks extend to individuals outside the hamlet). Regressions include the full set of neighbours set characteristics. Although the regressions include the proportion of 10 nearest neighbours being informed as a measure for the endogenous social interaction effect, the table presents the marginal endogenous social interaction effect of one additional neighbour being informed. The four characteristics of excluded peers used as instruments for 2SLS are: dummy for improved roof, female-headed households, whether the household has relatives in the village and whether the respondent was the household head.

*p<0.10; **p<0.05; ***p<0.01.

4.A.3. Evidence of exclusion bias in OLS estimation

Given that we consider a linear-in-means model in which households are excluded from their own peer group, we would expect the exclusion bias to affect the results presented in Table 4.3. In this Appendix, I derive the magnitude of the bias, which allows me to better interpret the differences in estimation results (i) between the OLS and 2SLS model and (ii) between the OLS model that adds village fixed effects and the OLS model that adds hamlet fixed effects.

Chapter 3 revealed that the magnitude of the exclusion bias depends on the peer group size and the size of the pool from which peers are selected. In the specifications underlying the results in Table 4.3, we considered peer groups of size 10. If the relevant pool of potential peers had simply been all the households in the hamlet, excluding the household itself, and if peers had been randomly drawn from within those hamlets, it would have been straightforward to calculate the magnitude of the exclusion bias, using the theoretical results from Chapter 3. However, in the current study, although peer group formation is highly correlated with hamlet formation, peers are not restricted to the hamlet.[67] Moreover, since we specify the peers to be the 10 nearest neighbours, and given residential sorting, peer selection is not random within the hamlet. In this Appendix I demonstrate how one can rely on simulations to estimate the magnitude of the exclusion bias in the case in which theory becomes significantly more complex.

Consider a simulation in which we generate a dataset of 30,339 observations (conform the sample size in our study) and randomly assign one normally distributed

[67] For 30% of the elderly/disabled households in the sample, a positive proportion of their 10 nearest neighbours lives in a different hamlet from the household itself. For 80% of the elderly/disabled households, less than two out of 10 peers live outside their own hamlet; For 95% of the households, less than five out of 10 peers live outside their own hamlet.

characteristic *y* to each observation. The underlying data-generating process (DGP) is the following:[68]

$$y_{li} = 10 + \varepsilon_{li},\ \varepsilon_{li} \sim N(0,1) \tag{4.3}$$

To distinguish exclusion bias from other sources of bias, I assume zero peer effects in the true population (i.e. $\beta_1 = 0$), no residential sorting and zero correlated effects. I randomly assign the 30,339 households to different locations in a geographic structure that is identical to the one in this study (i.e. with 100 villages and 705 hamlets of varying sizes). A household's peers are defined by the set of 10 nearest neighbours. Once the dataset is compiled, I run both village fixed effects and hamlet fixed effects OLS regressions on a model defined by:

$$y_{li} = \beta_0 + \beta_1 \bar{y}_{k_i} + \varepsilon_{li} \tag{4.4}$$

where y_{li} is the outcome of household *i* in hamlet *l* and $\bar{y}_{k_i}$ is the average outcome of that household's peer group of 10 nearest neighbours. As in Table 4.5, the regressions are run on the subset of 19,392 disabled/elderly households. In each simulation, I store the coefficient estimate $\hat{\beta}_1$ and repeat this procedure 10,000 times.

As a result of these simulations, I obtain an average estimate of the endogenous peer effect of –0.03 for the village fixed effects specification and –0.26 for the hamlet fixed effects specification. Since we know that in the true DGP $\beta_1 = 0$, we can conclude that

[68] The simulation results remain similar if we consider a binary characteristic (0 or 1) instead of a normally distributed variable. The reason for reporting on a normally distributed variable is to be consistent with the simulation framework considered in the previous section. Moreover, note that, whereas model (4.3) represents a simple bivariate model, the models we considered above add a set of 22 household control variables. In Chapter 3, I explained that the addition of control variables may reduce the exclusion bias. However, I do not expect this reduction to be significant in the current application, given the relatively low coefficient estimates of the control variables in regression (4.2) in Table 4.5 (with a maximum value of 0.152 for the coefficient estimate of political position holdings). Running simulations adding a set of 22 control variables to the DGP given in equation (4.3), with the coefficient values being the respective coefficient estimates of the 22 households' characteristics in column (2) of Table 5.4, I obtain estimates of the exclusion bias that are, indeed, very similar to those I estimate in this section.

the exclusion bias *per neighbour*, for the village fixed effects and the hamlet fixed effects specifications, respectively, is as follows:

$$\begin{cases} |Bias|^{Village} = 0.003 / neighbour \\ |Bias|^{Hamlet} = 0.026 / neighbour \end{cases} \quad (4.5)$$

Table 4.10 copies the estimation results from columns (3) and (4) in Table 4.3 and columns (1) and (2) in Table 4.5, and combines these with the simulation results presented in equation (4.5).

A naive comparison between the OLS estimation results in columns (1) and (2) in Table 4.10 would yield the conclusion that correlated effects at the hamlet level account for 3.5 pp (= 4.2–0.7) of the (biased) estimate of 4.2 pp of the endogenous peer effect, in a model ignoring clustering at the hamlet level. However, it is difficult to understand how this reconciles with the much smaller difference between the 2SLS estimation results of the models with and without hamlet fixed effects: comparing column (4) with column (3), we obtain a difference of only 1.5 pp. Moreover, if one were to solely rely on the two-stage least squares local average treatment effects (LATE) interpretation and measurement error as explanations for obtaining 2SLS estimates that are larger than OLS estimates, it would be difficult to understand the larger difference between 2SLS and OLS in models adding hamlet fixed effects (difference of 5 pp) compared with the equivalent difference in models not adding hamlet fixed effects (difference of 7 pp).

Table 4.10 shows how these inconsistencies disappear when we correct the OLS estimates for the estimated exclusion bias. The difference in (adjusted) estimates between hamlet and village fixed effect models becomes 1.2 pp in OLS, which now compares better with the equivalent difference of 1.5 pp in 2SLS. Moreover, the difference between the adjusted OLS and 2SLS estimates compare better between models adding hamlet fixed effects (difference of 4.4 pp) and those that do not

(difference of 4.7 pp). Note that there is still scope for the LATE interpretation of 2SLS and measurement error as possible explanations for obtaining 2SLS estimates that are higher than OLS estimates, as, even after adjusting the estimates for exclusion bias, a positive difference remains.

Table 4.10 Evidence of exclusion bias in ordinary least squares (OLS) estimation

Dependent variable = 1 if informed	OLS (1)	OLS (2)	2SLS (3)	2SLS (4)
'Naïve' estimation results:				
Endogenous peer effect: number of nearest 10 neighbours who are informed	0.042*** (0.003)	0.007** (0.003)	0.092*** (0.027)	0.077** (0.030)
Household's own characteristics	Yes	Yes	Yes	Yes
Contextual peer effects	Yes	Yes	Yes	Yes
Hamlet fixed effects	No	Yes	No	Yes
Correction using simulation results:				
Exclusion $\lvert bias \rvert$ per peer	0.003	0.026	0.000	0.000
Estimate endogenous peer effect corrected for bias = endogenous peer effect + exclusion $\lvert bias \rvert$	0.045	0.033	0.092	0.077
N	19,392	19,392	19,392	19,392

Notes: Dependent variable is a dummy = 1 if household was informed about the Tanzania Social Action Fund (TASAF). Standard errors (shown in parentheses) are adjusted for clustering of observations at the village level. Contextual effects include the full set of neighbours set characteristics.
*p<0.10; **p<0.05; ***p<0.01.

Chapter 5

Political connections and social networks in targeted transfer programmes: evidence from rural Ethiopia[69,70,71]

5.1. Introduction

Increasingly large transfer programmes are in operation across the developing world amid growing consensus that such programmes can form a key part of poverty reduction strategies (World Bank, 2006). Targeting is a central feature of these programmes, but concerns remain about the effectiveness of reaching the intended recipients. This chapter investigates how targeted food transfers are allocated in Ethiopia, one of the poorest settings in the world.

There is considerable evidence on possible biases in food transfer delivery against the poor in Ethiopia in terms of geography, demographics, gender or assets (Sharp, 1997; Clay et al., 1999; Jayne et al., 2001, 2002; Coady et al., 2004; Asfaw, 2006; Gilligan and Hoddinott, 2007). Clay et al. (1999), for instance, find that the primary beneficiaries of food aid programmes are those at both extremes of the need distribution: those with the least and those with the most food available. Much less is known about how transfers are actually reaching the final recipients. Existing studies generally had to make conclusions on targeting based on a relatively small information set on household and community characteristics. Using micro-level data from the Ethiopian Rural Household Survey (ERHS), this chapter augments the analysis by

[69] Previously published as: Caeyers, B. and Dercon, S., 2012. Political connections and social networks in targeted transfer programmes: evidence from rural Ethiopia. *Economic Development and Cultural Change* 60 (4), 639–675.

[70] Co-author: Stefan Dercon, University of Oxford; Department for International Development, UK.

[71] The data used in this article were collected as part of a collaboration among the Economics Department, Addis Ababa University, Ethiopia; the International Food Research Institute, Washington, DC; and the Centre for the Study of African Economies, University of Oxford, UK. The research benefited from support from iiG (Research Programme Consortium on Improving Institutions for Pro-Poor Growth), which was funded by the Department for International Development, UK. We are grateful for excellent comments by two referees and the editors of *Economic Development and Cultural Change*. We also gratefully acknowledge the feedback received from various academics at the University of Oxford during a Gorman Workshop presentation and my DPhil transfer interview.

studying the local political economy of transfers, focusing on social and political correlates of food aid access.[72]

Food aid delivery in Ethiopia has long taken two main forms: food for work (FFW) and free food distribution (FFD). Food aid is delivered using administrative targeting, with extensive decentralisation. When only limited systematic information for targeting is available to central bureaucracies, decentralised community-based systems for transfer delivery may offer a mechanism to ensure that the poor are served, as local government officials may have superior access to information about households in their community (e.g. Alderman, 2002). Nevertheless, it should not be assumed that using community-based institutions will deliver effective targeting (Conning and Kevane, 2002). Decentralising the power to allocate free goods and services to local political organisations may offer voice to the poor, but it could also strengthen local elites and networks by offering scope for rent seeking and patronage (Bardhan and Mookherjee, 2000; Conning and Kevane, 2002). Whether this decentralisation better represents the needs of the local population is dependent on the heterogeneity of preferences in the local population (Besley and Coate, 2003). It offers opportunities for targeting to reflect local conceptions of need and deprivation, but these conceptions are not necessarily pro-poor (Conning and Kevane, 2002). The evidence available, not least from South Asia, points to the crucial role of the local political economy in service delivery and decentralisation (Besley et al., 2005; Galasso and Ravallion, 2005; Pande, 2007). The evidence from Africa is very limited.

Studying these processes in Africa, and not least in Ethiopia, is important. In the last 10 years, Ethiopia has been the largest recipient of food aid just after North Korea

[72] The chapter focuses exclusively on targeting issues. Other papers have addressed the analysis of the impact of the transfer programmes involved (Yamano et al., 2005; Gilligan and Hoddinott, 2007). Using the same data as used in this chapter and based on a matched difference-in-difference estimator, Gilligan and Hoddinott (2007) found that both food for work and free food aid programmes raised consumption growth considerably.

(World Food Program, 2006). Each accounts for approximately 10% of total global food aid flows. Over this decade, typically approximately 5–14 million people were considered 'at risk' as part of international appeals, and food aid was widely distributed via FFW and food aid programmes. In 2002–3, a severe drought struck the country that affected approximately 12.6 million people – almost one-fifth of the population – and making it the most serious natural disaster in Ethiopia in recent decades, well beyond the 1984–5 and 1973–4 famines in terms of people affected (data from Emergency Events Database (2007)). One of the largest recent relief operations is credited for avoiding widespread famine and crisis (World Food Program, 2006). In its aftermath, the Ethiopian government has started to roll out a massive safety net scheme – the Productive Safety Net Programme (PSNP) – that combines FFW and food and cash transfers as part of a programme to rebuild the assets of the poor, with the aim of reaching 8 million people across the country.

The current chapter focuses on the targeting of food aid, through FFD and FFW, in the aftermath of the 2002–3 drought. Even though, in some of the localities involved, the situation was more severe as a result of the drought, the transfer delivery system and local targeting were very similar to the system that has been in operation for many years and, indeed, to the local delivery system as part of the current PSNP. In order to investigate whether social networks and/or political connections are of relatively greater significance in the period immediately after the drought when the need is acute or in the period sometime after the drought when the need is felt by only some very deprived people, the analysis will be disaggregated into two periods, with period 1 denoting the first 7 months immediately after the drought (September 2002–March 2003) and period 2 denoting the 12 months following period 1 (April 2003–March 2004). Besides looking at the selection criteria, this chapter also investigates whether social and political

networks matter for the level of FFD and FFW rewards per recipient household and, for FFW, per day worked.

Although the main focus of this chapter will be on the role of political connections in access to food aid, we will also consider correlations of food aid receipt with other social network types. We distinguish between horizontal and vertical networks. The former refer to social contacts with similar degrees of power, whereas the latter refer to links between citizens and the political elite. Furthermore, the role of the informal social safety net (ISSN) is considered separately. The conjectures are that vertical networks in which a person is connected to the local political elite may well matter, given that the entire allocation process relies on the local *kebele* committee making lists of eligible households. The *kebele* is a locally elected administrative unit, with close links to local, district and national political processes. The role of horizontal networks may relate to being connected to others to obtain information or, even, as in the case of the informal insurance network, to be able to share any transfers (as would be predicted by risk-sharing models; Attanasio and Rios-Rull, 2000; Dercon and Krishnan, 2003). Furthermore, the official targeting rules for food aid (which are currently in operation for the PSNP) explicitly state that those households with family support or other means of social protection and support should not receive food aid (Ministry of Agriculture and Rural Development, 2004).

We find that political connections seem to matter in the free food-allocation process, although only clearly in the period immediately after the drought. In this period, targeting based on need seems to be only weakly significant. Sometime after the drought, however, targeting moves very significantly to those in need, and social and political network variables are no longer significant. Vertical connections play the most important role, but also the ISSN of a household seems to be influencing the FFD process. Horizontal social networks in general, however, do not seem to have a

significant impact on how free food is allocated among the households. As for access to FFW, we find that only labour supply characteristics such as ability to work matter significantly, and we do not find any evidence of selection based on economic need. Overall, social and political networks do not seem to matter for participation selection or days worked. However, we do find evidence that households with local political connections are significantly better rewarded in terms of cash or food receipts per working day and in total than households without such connections.

The chapter is organised as follows. In section 5.2, we discuss very briefly how food aid is targeted in Ethiopia. In section 5.3, we set up a conceptual framework in which we discuss the potential role of social and political networks in the FFD process and the FFW scheme. Section 5.4 offers a description of the data and some relevant summary statistics. Section 5.5 contains the empirical specifications, a discussion of the main problems potentially involved and the estimation results. Section 5.6 describes some suggestive evidence on the channels through which political connections affect food aid targeting in Ethiopia. Finally, a discussion of the results and some concluding remarks are provided in section 5.7.

5.2. Food aid targeting in Ethiopia

Food aid delivery in Ethiopia has historically taken two main forms: FFW and FFD; only in the last few years, in the context of the PSNP, have cash transfers, mainly via cash for work (CFW), also begun to be used more systematically. There is a longstanding commitment by the government of Ethiopia to distribute the majority of food aid via FFW programmes related to public works. As FFW is only effective for those able to work, it is recognised that free transfers will remain an important part of any safety net system, not least during crisis situations.

The drought of 2002–3 triggered such a potential crisis. A failure of the main rains in several crop-dependent areas resulted in food production estimates approximately 18% below normal levels (Famine Early Warning System Network, 2002–3). Large amounts of food aid were pledged and delivered – nearly 1.5 million tons – and its distribution was managed via the Disaster Prevention and Preparedness Commission (DPPC). However, ultimately, neither international partners nor the DPPC determine how food is distributed among households within the villages in Ethiopia, and this is a direct consequence of the hierarchic nature of the food aid distribution process.

Free food allocations are typically made in three stages. First, the DPPC allocates food to each district (*wereda*). Then, for rural distribution, *wereda* committees assign allocations to individual peasant associations (PAs) within their *wereda*. In rural areas, the PA is the local *kebele*, the lowest level of political administration, usually consisting of a relatively small number of villages. Finally, the *kebele* leaders prepare a list of beneficiary households, and distribution is carried out by members of the PA. A critical element of this process is that, although the amount of food to be allocated to each *wereda* is determined at the federal level, the actual beneficiary households are designated at the local community (PA) level (Jayne et al., 2001).

In principle, the *kebele* leadership is elected using some official procedures for free and fair elections, although in practice political manipulation and historical political control by those in power at the national or regional level may cast doubt on the extent of local political competition (Pausewang and Aalen, 2002). In particular, voters tend to perceive the ruling party as agents of the state, and would be reluctant to upset those agents as voters are dependent on the state for the receipt of many benefits, such as access to health cards (offering entitlements to health services) or modern inputs. Although its evidence gathering can be questioned on methodological grounds, Human Rights Watch (2010) argued that safety net and other allocations are increasingly used

for political purposes. Lefort (2007) argued that this leads the rural population to be first and foremost concerned to vote for the winning side, since to do otherwise carries intense risks to their welfare and even survival. There was a significant increase in electoral competition only by the 2005 elections. Its dramatic consequences led to serious repression afterwards. More recently, electoral competition has again been minimal. In the 2001 *kebele* elections, there was no significant national electoral competition, and in rural areas voters largely supported the ruling national coalition – the Ethiopian People's Revolutionary Democratic Front.

The official goal has long been that work-based allocation (as in FFW or CFW) programmes account for 80% of all distributions (World Food Program, 1995). These programmes are used to build community assets, such as dams and roads, while alleviating hunger. The selection of FFW participants in Ethiopia has followed widely different rules (Sharp, 1997). In some regions, the distribution process is similar to that for FFD, where local community groups (e.g. PAs) choose households eligible for participation based on some underlying criteria such as land size, work ability and asset ownership. Programmes in other areas have occasionally used self-targeting schemes, which do not make use of administrative restrictions on participation and where households self-select into the programme. In principle, self-selection programmes are designed in such a way that the programme is unattractive to any but the poorest. Commonly used self-selection mechanisms of FFW programmes have been to give out relatively low-quality food, to require queuing to receive the transfers and to pay out low wages (cash or in kind) relative to the prevailing market wages (Sharp, 1997). However, although some of the empirical evidence finds that FFW effectively reaches the intended beneficiaries (Von Braun et al., 1998; Gebremedhin and Swinton, 2000), several studies have found evidence that the non-poor also participate in FFW programmes (Sharp, 1998; Clay et al., 1999; Gebremedhin and Swinton, 2000; Jayne et

al., 2002). Sharp (1997) concludes that 'a careful literature search produced no Ethiopian examples of self-targeting employment schemes successfully selecting the poorest and excluding the better-off'. Several explanations for this have been given in the literature. Barrett and Clay (2003) argue that imperfect or missing factor markets, such as labour, land and finance markets, lead to the poor opting out of FFW while the rich self-select into them. The most common explanation, however, is that FFW wages are set too high relative to the prevailing market wages, often on purpose, to attract enough labour and to make the programme successful (Sharp, 1997). When wages are set too high, and given restrictions on funds, excess labour supply occurs, including labour from the non-needy group. FFW employers will then have to ration participation in some way, potentially based on criteria other than objective need. In any case, as FFW employers are typically linked to the local *kebele*, processes similar to FFD are ultimately used in practice.

5.3. The role of social and political networks in food transfers

The role of social capital in affecting the well-being of households and the development of societies has received much attention in the literature (Coleman, 1988; Putnam et al., 1993; Narayan and Pritchett, 1999; Grootaert et al., 2002; Durlauf and Fafchamps, 2008). Its meaning, however, remains highly imprecise. Durlauf and Fafchamps (2008) conclude that the most successful theoretical studies of social capital and development are those in which the focus is not on social capital per se, but where social capital is modelled as a specific form of social network structure that affects individual outcomes. This contrasts with most early studies, in which social capital is defined as a set of social resources of a community, such as trust and control (Coleman, 1988; Putnam et al., 1993).

To investigate the importance of social contacts for households within each community in the last stage of the food distribution process (through either FFD or FFW), the analysis in this chapter concentrates on social networks at the household level instead of aggregate social capital at the community level. In the remainder of this chapter, the term *social networks* is used to refer to social networks at this level.[73] We will, however, go a step further by highlighting the nature of the network relations within the network, focusing on the degree of political connectedness of the network as a means of accessing benefits.

The literature on social networks outlines various links between social networks and economic outcomes. Here, a distinction is made between the *horizontal* and *vertical* social network of a household (Putnam et al., 1993). The former refers to social contacts with similar degrees of power, whereas the latter refers to links between the political elite (the PA leaders) and the household. As part of either the horizontal social network and/or the vertical social network, we distinguish a subset denoted by the *informal social safety net*. This includes only those other households that a household can actually rely on in times of need. These distinctions matter for the discussion, since the roles of these networks in the food distribution process may differ.

Three main micro-level properties of networks are considered here: their joint value as informal social insurance, their ability to improve information flows and their function as a source of favouritism. Informal insurance networks refer to the group of associates (friends, family, neighbours or others) on which one can rely in times of need. There is a large literature documenting their relevance in poor settings (Townsend, 1995; Dercon, 2002). Given that in closely knit societies, such as the

[73] The importance of social capital and political connectivity at the community level in explaining differences between communities in their food aid-targeting processes is of course also of interest. It may be, for instance, that the degree of successful food aid targeting of a community is related to its degree of political connectedness to the district administration, as well as its social cohesion, trust and control. The available data are unlikely to be most suitable to analyse this, as we have only a limited number of PAs (nine) in the sample.

communities studied in this chapter, information on the presence of these networks may be relatively high, it may also feature in food aid targeting. In fact, one of the explicit targeting rules intended to be used for assessment at the community level is whether a person can rely on family support or other forms of social protection (Ministry of Agriculture and Rural Development, 2004). The *kebele* leaders may, therefore, be less likely to offer food aid to those well connected in terms of informal insurance networks. However, the analysis will have to take into account that formal food aid targeting may well crowd out informal networks. Standard informal insurance models under enforcement constraints would directly predict this: food aid targeted to a particular person will change the outside options available and therefore increase incentives to leave the informal risk-sharing groups (Cox and Jimenez, 1992; Attanasio and Rios-Rull, 2000; Cox et al., 2004). In line with these models, there is some evidence from Ethiopia using earlier rounds of the dataset used in this chapter that the presence of food aid in the community crowds out informal insurance (Dercon and Krishnan, 2003). In contrast, Lentz and Barrett (2005) do not find any evidence of such effects in Ethiopian food aid targeting.

Horizontal social networks in the community may also affect targeting via their role in information transmission. In a world of imperfect information, social ties can provide an individual with useful information about opportunities and rights otherwise not available. This information advantage of social relations has typically been used to illustrate the importance of social networks in the job market (Granovetter, 1975). Its relevance for food aid targeting is obvious. Households that have a larger network have a higher probability of receiving information about potential food aid programmes than more excluded households. Belonging to a larger network may also allow more effective lobbying for support.

Vertical connections, that is, connections to those in power, amplify considerably the potential of these networks to acquire goods and services, as they provide more ready access to information on rights to support and availability of food aid. Moreover, vertical connections also improve upward information flows, in terms of signalling of need. This is likely to matter as the official allocation criteria remain vague, and *kebele* leaders are likely to need to supplement observable household characteristics with other criteria. An improved information flow linked to vertical connections cannot easily be distinguished from another intuitive link: political connections as a source of favouritism. It may well be that the PA leaders use their political power to manipulate actions in favour of households that they are vertically connected to, irrespective of whether these households need the food aid or not.[74]

This creates the potential for a whole series of other linked behaviours, documented elsewhere in rural settings, from vote buying by local leaders to repression (Pande, 2007; Robinson and Baland, 2008). Political processes at the PA (and any other) level in Ethiopia are definitely not transparent, lacking clear checks and balances. The dependence of rural households on the *kebele* committees for different benefits and services, such as health services and access to modern inputs and credit, implies that not just voting for the winning side but also connectivity to those in power may well be crucial.

A priori, there are several reasons why the effects of social and political networks may be relatively more important during crisis times, and this will be formally tested further below. First, PA leaders may need time to properly assess objective need-based eligibility of all households. As long as such information is not yet available, the PA leaders will have to rely on information that is more readily available. Households that

[74] This property of social networks is, for obvious reasons, often also used in the job search literature (Lin, 2001).

are politically connected will be able to signal their need more quickly. Second, it may take time for information on FFD and FFW to reach all needy households in the PA. Households with political connections will have more direct access to such information and hence have comparative informational advantage over those households that are not vertically connected. Third, during crisis time, relatively more people are likely to be in need of food aid, and the demand for FFW opportunities is likely to be relatively high. Under such circumstances, local officials will have to ration participation in some way, using criteria other than objective need, such as social and political networks. PA leaders may argue that a needy household that can rely on other households through an ISSN needs public food aid less than a needy household without such social networks. Moreover, if it is necessary to choose between two equally needy households, local officials may prefer to give food aid to a household they are socially connected to.

5.4. Data source and descriptive statistics

Our analysis uses data from the ERHS, a longitudinal household dataset collected as part of collaboration between the International Food Policy Research Institute, Addis Ababa University, Ethiopia, and the Centre for the Study of African Economies, University of Oxford, UK. The dataset was collected in seven rounds from 1994 to 2009 in 15 rural Ethiopian PAs. In nine of these PAs, free food was distributed and FFW programmes were organised during the drought crisis in September 2002–March 2004. As we focus on this drought episode, we use the data from these nine PAs and obtain a sample of 602 households in total.[75] Rounds 5 and 6 were collected in 1999 and 2004, respectively, and contain detailed information on the households' pre-drought characteristics useful for targeting analysis and their participation in the food aid

[75] Using the same dataset as the one we use in our study, Gilligan and Hodinott (2007) conclude that 'Among households in villages receiving food aid, the overall attrition rate was low, 6.5 percent [over 1999–2004], or 1.3 percent per year. There is no correlation between observable household characteristics and attrition'. We are therefore confident that attrition will not significantly influence our results.

programmes during and after the 2002 drought. They form the basis for the analysis. Across these PAs, approximately 60% of households received free food between September 2002 and March 2004; approximately 62% of households participated in FFW programmes during this period; and approximately 39% benefited from both programmes. Splitting this period into two, 42% received free food aid during the 7-month period immediately after the drought and 44% during the subsequent 1-year period. The share of households participating in FFW programmes amounts to 57% in both period 1 and period 2. Approximately 25% received free food in both periods, and approximately 53% of the households participated in FFW programmes in both periods, implying a larger re-allocation of free food aid between the two periods than in the FFW programme, which seems to have been largely to the same target group.

Table 5.1 shows selected summary statistics for the characteristics of the households in the sample, distinguishing between those that received free food in 2002–4 and those that did not, as well as those that participated in FFW programmes and those that did not. As part of the community survey in each PA, PA leaders were asked to specify the criteria they used to allocate food aid in their community. We obtain seven different criteria, which are ranked from most to least frequently listed by the different PA leaders: people unable to work, old people, poor people, landless people, large families, people with limited livestock, and female-headed households. Broadly speaking, these criteria could be viewed as consistent with the 'official' guidelines, which state that food aid should be given to those who are facing serious food shortages (defined as deficiencies in food availability for 3 months or longer), those who experienced a serious loss of assets without means to support themselves, or those without family support or other means of protection (Ministry of Agriculture and Rural Development, 2004). Ethiopian food aid policy states that only those persons who are unable to work are eligible for free food (Disaster Prevention and Preparedness Commission, 2000).

Table 5.1 Characteristics of free food distribution (FFD)/food for work (FFW) beneficiary households (2002–4) and non-FFD/FFW beneficiary households: based on peasant association criteria (1999)

		Average work ability score of the household head (1)	Average proportion of elderly in household (2)	Average household consumption per capita (3)	Average household land ownership per capita (ha) (4)	Average household size (5)	Average household livestock units/capita (6)	Share of female-headed HHs (%) (7)
FFD	Group NOT receiving free food (237 observations)	3.82 (0.48)	0.03 (0.11)	74.60 (58.49)	0.22 (0.23)	6.21 (2.78)	0.41 (0.40)	0.25 (0.19)
	Group receiving free food (365 observations)	3.62 (0.70)	0.05 (0.14)	76.22 (75.46)	0.24 (0.25)	5.68 (2.65)	0.44 (0.42)	0.31 (0.21)
	Test statistic difference in means between 2 groups	3.95***	–1.84*	–0.28	–1.12	2.33**	–0.88	–1.54
FFW	Group NOT participating in FFW (229 observations)	3.53 (0.77)	0.06 (0.17)	76.88 (74.06)	0.20 (0.18)	5.56 (2.97)	0.42 (0.41)	0.35 (0.26)
	Group participating in FFW (373 observations)	3.80 (0.50)	0.04 (0.10)	74.79 (66.17)	0.25 (0.27)	6.09 (2.52)	0.44 (0.42)	0.24 (0.18)
	Test statistic difference in means between 2 groups	–5.07***	1.85*	0.36	–2.63***	–2.35**	–0.48	2.78***
	Full sample (602 observations)	3.70 (0.63)	0.05 (0.13)	75.58 (69.23)	0.23 (0.24)	5.89 (2.71)	0.43 (0.41)	0.28 (0.20)

Notes: Numbers in parenthesis are standard deviations. Standard *t*-tests for first six columns and test of equality of proportions for last column. Average work ability of head is evaluated in 2004 instead of 1999 because of data limitations. The score ranges from 1 to 4, where 1 = not able at all and 4 = perfectly able. Livestock units are scaled using standard tropical units. For the binary variable 'Share of female-headed households', the statistic used to test equality of proportions adjusted standard deviations appropriately. HH, household. *significant at 10%, **significant at 5%, ***significant at 1%.

Needy persons who are able to work are required to participate in FFW in order to obtain food aid.

Most of the significant differences in characteristics between the beneficiaries and non-beneficiaries of FFD and FFW relate to differences in labour supply characteristics (Table 5.1). Except for land ownership, there are no significant differences between any of the characteristics related to economic vulnerability (consumption and livestock ownership). For land size, the difference is only significant for FFW, and the difference is in the opposite direction from what we would expect, with FFW participants having on average more land than non-FFW participants.

The most significant difference in characteristics between beneficiary and non-beneficiary groups is found in the average work ability score of the household head. This measure is an average score based on five different questions concerning the household head's ability to carry out daily activities. Each question has four possible answers from not being able to perform the task at all (value 1) to being able to do it easily (value 4).[76] The average score across the five questions offers an index of ability to work, ranging from 1 to 4, where 1 is interpreted as not being able to work at all and 4 is interpreted as perfectly able to work. The results show that this score is on average significantly lower for free-food-receiving households than for non-free-food-receiving households and significantly higher for FFW participants' households than for non-FFW participants. This seems to be the most important criterion for food aid targeting among the criteria provided by the PA leaders. This result is not surprising, given that disability is also the most emphasised criterion for FFD by the DPPC (Disaster Prevention and Preparedness Commission, 2000).

[76] These five questions were: 'Can this person: (1) Stand up after sitting down? (2) Sweep the floor? (3) Walk for 5 km? (4) Carry 20 litres of water for 20 m? (5) Hoe a field for a morning?'

Free-food-receiving households have, on average, a smaller household size than non-free-food-receiving households, whereas FFW-participating households have, on average, more members than non-FFW-participating households. A priori, the expected bivariate correlation with household size is ambiguous, as larger households have more mouths to feed, but at the same time they are likely to have more able-bodied members. Similar results are obtained for the average proportion of the elderly in the household, with FFD- and FFW-participating households having, on average, a higher and lower proportion of elderly, respectively. Finally, FFW-participating households have, on average, a significantly lower share of female-headed households than non-participating households. This difference is not significant for FFD, however.

Table 5.2 shows other characteristics of the households in the survey sample, again distinguishing between beneficiary and non-beneficiary households, but this time based on different ways of conceptualising social networks and political connections. We use three different measures, each capturing different types of networks, with different predictions about their role in food transfer allocation. First, we use a self-reported measure of the size of the ISSN. In round 6 of the survey, we asked how many other households or individuals the household could rely on in times of need. This provides a direct measure of the informal insurance network. Because it focuses on potential support, it is superior to measures based on who actually received transfers from others, as required in most risk-sharing analyses within networks investigating whether consumption is smooth because of transfers (Fafchamps and Lund, 2003). Controlling for household resources, it is also likely to be a relevant indicator that the *kebele* committee may have to look at in view of the official targeting rules (i.e. food aid should not be given to those able to rely on other sources of support).[77]

[77] The measure is only available for round 6.The potential endogeneity with food aid is explored further below.

Table 5.2 Characteristics of free-food-receiving (2002–4) and non-free-food-receiving households: based on social networks[a]

		Average number of persons the household can rely on in times of need (1)	Size of the largest *iddir* the household belonged to prior to drought[a] (2)	Share of households that have close associates holding official positions (%) (3)
FFD	Group NOT receiving free food (237 observations)	9.49 (16.54)	82.72 (154.77)	0.29 (0.21)
	Group receiving free food (365 observations)	6.64 (6.15)	94.85 (165.68)	0.45 (0.25)
	Test statistic difference in means between 2 groups	2.98***	–0.90	–3.89***
FFW	Group NOT participating in FFW (229 observations)	7.61 (15.04)	80.07 (149.04)	0.30 (0.21)
	Group participating in FFW (373 observations)	7.86 (8.65)	96.21 (168.52)	0.44 (0.25)
	Test statistic difference in means between 2 groups	–0.25	–1.19	–3.49***
	Full sample (602 observations)	7.76 (11.50)	90.07 (161.45)	0.39 (0.24)

Notes: Numbers in parentheses are standard deviations. The size of the *iddir* is expressed as the number of its members. If the household belonged to several *iddirs*, only the largest *iddir* was considered. For the binary variables the statistic used to test equality of proportions adjusted standard deviations appropriately.
FFD, free food distribution; FFW, food for work. *significant at 10%, **significant at 5%, ***significant at 1%.

Column 1 in Table 5.2 indicates that this may indeed be the case for FFD, as free-food-receiving households claim on average to have significantly more people to rely on than non-free-food-receiving households. On average, a non-free-food receiver claims to have 10 individuals to rely on, whereas a free-food receiver claims to only have seven individuals to rely on. Of course, other evidence on risk-sharing networks has shown that wealthier households may have larger networks (with the causality possibly going in either direction; De Weerdt, 2004; Fafchamps and Gubert, 2007). Hence, we have to be careful in interpreting these correlations from this bivariate analysis. For FFW,

however, there does not seem to be any significant difference between participating and non-participating households in terms of the number of people the household can rely on in times of need.

To measure the broader horizontal social network of the households, we use the size of the largest *iddir* that the household belonged to prior to the drought. The *iddir* is a funeral society. Members pay a contribution, in many cases monthly, and its benefits are that the group pays for the costs of a funeral of a member and any of the member's close relatives. Virtually all people in the sample are members of such a group, although in each PA many different groups exist. Dercon et al. (2006) discuss details of their functioning as a funeral group. Hoddinott et al. (2005) and Krishnan and Sciubba (2009) document how members of *iddirs* also typically are linked in other ways, such as in the form of labour-sharing groups (to carry out particular agricultural tasks together). It is the most ubiquitous, and a relatively egalitarian, social institution in Ethiopia (Dercon et al., 2006). As virtually everybody is a member of at least one *iddir*, we use the size of the largest *iddir* that a person belongs to as our proxy for the network strength.[78] Table 5.2 shows that the average size of the *iddir* that the household belonged to prior to the drought was 83 for the non-free-food-receiving group and 95 for the free-food-receiving group. However, the differences are not statistically significant. Similarly, the differences in terms of network strength are not significant between FFW-participating households and non-participating households.

Finally, we explore a measure of political (vertical) connectivity. The most direct measure collected for our purposes was the simple question 'Do any of your close associates (relatives [not parents], friends, patrons) hold an official position in this

[78] We explored other measures for the strength of the *iddir*, but the data on the value of benefits of the *iddir* proved too incomplete to pursue across all the villages. Bold and Dercon (2009) use more data collected on all the *iddirs* for a subsample, but this subsample proved too small for the purposes of this chapter. Data limitations also prohibited us from using the sum of the members of all *iddirs* that the household belonged to, instead of taking the size of the largest *iddir*.

kebele?'[79] Note that all positions in the *kebele* are effectively 'political', based on appointments by the political leadership. Table 5.2 indicates that vertical social networks do, indeed, matter in the food allocation process, in both FFD and FFW. Differences between beneficiaries and non-beneficiaries in terms of political connections are strongly significant, with approximately 30% of the non-beneficiary households being politically connected and approximately 45% of the beneficiary households.

In Tables 5.1 and 5.2 we compared the average characteristics of FFD beneficiaries and FFW participants with the characteristics of FFD non-beneficiaries and FFW non-participants, respectively. Pooling the FFD and FFW beneficiaries together, we do not find any significant differences between the average beneficiary (either FFD and/or FFW) and the average non-beneficiary (neither FFD nor FFW) in terms of the characteristics in Table 5.1 (not reported). This is not surprising, given that most of the characteristics for which we found significant differences in Table 5.1 relate to labour supply characteristics, for which the official targeting rules are symmetric for FFD and FFW.

5.5. Empirical analysis

5.5.1. Model specification

Although we will model and discuss FFD and FFW participation separately, we will use similar estimation equations for both of them. First, we explore the probability of obtaining free food or participating in FFW programmes as a function of household characteristics based on the PA criteria, social and political networks, and other exogenous attributes that we believe to be important. In particular, we will estimate the following equation:

[79] The data on these political connections were collected in round 6, although political office had not changed since the elections of 2001, before the drought crisis; any consequences for the analysis are discussed further below.

$$F_{ij} = f(PAC_{ij}, HSN_{ij}, VSN_{ij}, ISSN_{ij}, X_{ij}, Z_j) \quad (5.1)$$

We will estimate a number of versions of this model, each with a different dependent variable and time period. In particular, F_{ij} is a dummy equal to 1 if household *i* in PA *j* received free food (for the FFD model) or participated in public works (for the FFW model). We first consider the probability of obtaining aid in the total period of 19 months after Ethiopia was struck by the drought in 2002, that is, the period September 2002–March 2004. In order to analyse whether the effect of social and political networks differs between the period immediately after the drought, when the need is more acute, and the period sometime after the drought, when only the most seriously affected households are likely to be most needy, we also split this period into two periods: the probability of obtaining food aid in the 7 months immediately after the drought struck Ethiopia (September 2002–March 2003) and the probability of obtaining food aid in the 12 months after the initial period (April 2003–March 2004).

Vector PAC_{ij} contains household characteristics based on the PA criteria discussed in section 5.4; HSN_{ij}, VSN_{ij} and $ISSN_{ij}$ represent the measures of the household endowment of, respectively, horizontal social networks (the *iddir*), vertical social networks (political connectivity) and the ISSN; X_{ij} is a vector of other household characteristics; and Z_j is a vector of community characteristics. We will control for all these community-wide effects using PA fixed effects, capturing any between-PA variation in food aid allocation, and identifying all targeting effects via within-PA variation. As the dependent variable is binary, the model is estimated using a probit model.[80]

[80] Similar results are obtained when a linear probability model is used.

Based on our earlier discussion, the vector PAC_{ij} in the model consists of the following characteristics of the household: its work ability (proxied by the average work ability score of the household head); its proportion of elderly; its income level (proxied by its real consumption per capita[81]); its land ownership (measured by hectares of owned land per capita, including crops, grazing and garden); log of the household size; and a dummy for whether the household is female headed or not.[82]

The vector X_{ij} consists of other household characteristics that have not been mentioned by the PA leaders, but that we expect to play a role in the food distribution process. While consumption measured in 1999 has value as a proxy for pre-crisis poverty, it may not capture immediate needs during the crisis of 2002–4. We therefore also add some measures of idiosyncratic shocks, including a dummy for whether the household was affected by the drought in 2002–3. Other broader measures of 'need' may include the proportion of children and elderly in the household. Furthermore, we include a dummy for whether the household head is literate (which could capture wealth or income-earning potential, although also the ability to enforce access to food aid) and a dummy for whether the household belongs to the majority religion group of its PA (offering a further measure of horizontal networks). A further control relates to inertia in food aid distribution, in the form of a dummy for whether the household received food aid in the past (at any time between 1982 and 1999).[83] Jayne et al. (2002) found that, in

[81] We also explored models adding non-linear terms of consumption per capita, but, since these do not change the results on the social network variables, they are dropped from the results reported in this chapter.

[82] Note that, in the reported regressions, we did not include livestock ownership, although it does reflect a criterion of the PA leaders. Including livestock, together with consumption and land ownership, does not change any of the results and does not show any significant effect of livestock. However, it appears to cause problems of multi-collinearity, affecting significance levels of other variables. Since only one PA claimed to be using livestock ownership as a target criterion and since its coefficient estimate in the regressions is never significant, we believe that avoiding multi-collinearity problems is a preferable choice. We fully recognise that land ownership and real consumption are also highly correlated, but, as we will see in the regressions, land ownership has a very significant impact on the food distribution process. Hence, we decide to include it in the regression model.

[83] The data on past food aid were obtained using previous ERHS data rounds.

the mid-1990s, food aid displayed substantial inertia: *weredas* and households receiving food aid in the past tended to receive food aid again, controlling for a wide variety of characteristics and shocks. To some extent, this should not come as a surprise: one of the official guidelines for food aid delivery uses the criterion of targeting households that have faced continuous food shortages and received food assistance. In short, inertia is part of policy for rather benevolent reasons: to avoid the need to repeat assessment (so the needy can get support when required without delay), although it can of course perpetuate any bias in allocation. Similarly, a *wereda* is currently selected for the PSNP if it is chronically food insecure and has been a recipient of food aid for a significant period in the past (Sharp et al., 2006).

All of the control variables, except for the average score of ability to work of the household head, are evaluated in 1999.[84] This avoids potential simultaneity problems, since we are interested in food allocation between 2002 and 2004, a time period after 1999. One could rightly point out that assets in 1999 could still be endogenous, since both food aid receipts in 2002–4 and assets in 1999 are likely to be correlated with past food aid. However, since we include a dummy for past food aid in the regression (related to 1999), we can credibly avoid this problem.

As has already been discussed in section 5.4, HSN, VSN and ISSN will be measured, respectively, by the size of the largest *iddir* that the household was member of prior to the drought, a dummy for whether the household has close associates holding official positions and the number of persons the household can actually rely on in times of need.

Besides modelling the selection criteria through a probit model, we also investigate whether social networks affect the amount of food aid received through FFD and FFW.

[84] The reason why we use work ability measured in 2004 is because of data limitations, as explained in the previous section.

We use the same specification as in (5.1), but now considering as the dependent variable the value (in birr) of the quantity of food and cash received in the second period (April 2003–March 2004). Because of data limitations, we cannot model quantities received during the first 7 months immediately after the drought. We use an ordinary least squares (OLS) model for those receiving aid in either form. One alternative would be to use a Tobit model for the entire sample, but it imposes strong restrictions on the role of explanatory variables, and it requires strong distributional assumptions on the errors for consistency.[85] An alternative would be a Heckman selection model, but, as there is no obvious source of identification of the first stage, identification occurs via nonlinearities in the model, which again is not ideal. In any case, using a Tobit or Heckman selection model does not affect the findings.

For FFW, we also investigate whether differences in the value of food/cash received through FFW are driven by differences in the total number of days worked in FFW programmes, or by differences in the daily wage (cash/food) received. We explore this further, with the same specification, using OLS regressions, limited to those households that participated in FFW in this second period.

There are a number of potential problems with our estimation equations, especially in terms of identifying the impact of the network variables on access to food aid. A key issue is the potential endogeneity of these measures. For example, there may be simultaneity problems: networks and political connections may be affected by the process of food aid delivery. Before addressing such identification concerns in section 5.5.4, the next two sections first describe the (possibly naive) OLS estimation results for FFD and FFW, respectively.

[85] A Tobit model would restrict the impact of the explanatory variables to be exactly the same for both selection of beneficiaries and the determination of the levels of the rewards. Existing literature (Jayne et al., 2000) suggests that this is not the case.

5.5.2. Free food distribution targeting results (OLS)

Table 5.3 gives the results concerning the determinants of FFD. Column 1 gives the marginal effects of a range of variables on the probability of obtaining food aid at any time between September 2002 and March 2004. Over this aggregated period, there is evidence of means-based targeting within PAs. Evaluated at the mean of all other variables, households at the 25th percentile of log per capita consumption had an average probability of 65% of receiving free food. The probability declines to 58% and 52% at the 75th and 95th percentiles, respectively. Note, however, that the 'richest' households still have more than a 50% probability of obtaining food despite substantial numbers, including some of the poorest, not getting any aid. While FFD seems to be correlated with poverty measures such as consumption in 1999, it does not seem to be correlated with measures reflecting immediate post-famine needs, such as whether the household was affected by the drought. Another significant determinant is the average work ability score of the household head. A 1-point increase in this score (scores range from 1 to 4) decreases the probability of obtaining free food by 12 percentage points (pp).

Other strongly significant determinants are land ownership and log of household size, both having a large negative effect on the probability of obtaining free food. The negative effect of household size may be explained by the fact that, for households with a given proportion of children and elderly, the absolute number of active members increases with household size and, therefore, decreases the probability of obtaining free food. In short, food aid allocation is correlated with official targeting rules related to underlying poverty, as well as the relative exclusion of able-bodied workers. Still, the marginal effects show that, even with higher levels of consumption, one could still obtain food aid in this period.

Table 5.3 Determinants of free food distribution (FFD)

Dependent variable	FFD = 1 if HH received free food 2002–4		FFD = 1 if HH received free food, period 1		FFD = 1 if HH received free food, period 2		Value of food/cash received, period 2	
Estimation equation	(1) Probit		(2) Probit		(3) Probit		(4) OLS	
VSN†	0.119***	(0.045)	0.103**	(0.045)	0.070	(0.047)	–3.227	(22.744)
ISSN	–0.007**	(0.003)	–0.007**	(0.003)	–0.001	(0.003)	–1.152	(1.222)
HSN	–0.003	(0.009)	–0.014	(0.009)	0.006	(0.009)	3.524	(2.982)
Ln (consumption/capita)	–0.079**	(0.037)	–0.027	(0.036)	–0.114***	(0.040)	–18.991	(12.796)
Work ability score head	–0.121***	(0.042)	–0.066*	(0.038)	–0.104***	(0.039)	6.219	(15.878)
Proportion elderly (>65 years)	0.051	(0.191)	0.218	(0.181)	–0.017	(0.173)	–45.613	(54.550)
Land area owned/capita	–0.359**	(0.141)	–0.230*	(0.131)	–0.318**	(0.137)	–34.897	(76.112)
Ln (HH size)	–0.184***	(0.060)	–0.011	(0.054)	–0.200***	(0.058)	8.025	(21.639)
HH head is female†	–0.022	(0.056)	0.057	(0.054)	–0.009	(0.056)	–43.443**	(21.012)
Proportion children (<6 years)	–0.078	(0.146)	–0.064	(0.148)	0.048	(0.151)	–10.633	(55.312)
HH head is literate†	0.017	(0.051)	0.044	(0.050)	–0.037	(0.052)	–6.384	(21.551)
Major religious group †	–0.036	(0.059)	–0.035	(0.061)	–0.011	(0.064)	34.649	(26.279)
FF (1982–99)†	0.139*	(0.077)	0.050	(0.073)	0.133*	(0.074)	–20.109	(28.979)
Drought shock (2000–2)†	0.040	(0.057)	0.165***	(0.051)	–0.063	(0.058)	17.967	(17.737)
Ill male (1999–2002)†	0.060	(0.083)	0.005	(0.087)	0.048	(0.091)	16.239	(29.740)
Ill female (1999–2002)†	0.055	(0.081)	0.081	(0.085)	–0.027	(0.089)	36.238	(29.320)
Member died (1999–2002)†	0.042	(0.056)	0.112**	(0.055)	–0.082	(0.056)	22.452	(24.089)
N	602		602		602		263	
Pseudo R^2	0.187		0.105		0.175		0.305	

Notes: All models in this table include peasant association dummies (not reported). Period 1, September 2002–March 2003; period 2, April 2003–March 2004. Reported coefficients are marginal effects. Equations (1)–(3): changes in marginal probabilities for continuous variables and discrete changes in the probability from changing the value from 0 to 1 for dummy variables (marked with †). Consumption per capita is measured by total consumption minus transfers in 1999. *Iddir* size is scaled (divided by 50). Numbers in parentheses are robust standard errors. Last column is conditional on receiving free food. Mean value of food/cash received by those receiving in this period is 176.9 birr (176.8 standard error). Birr exchange rate at the time of the survey: 8.5 birr per US dollar.

FF, free food; HH, household; HSN, size of the largest *iddir* the HH belongs to; ISSN, number of persons a HH can rely on in times of need; OLS, ordinary least squares; PA, peasant association; VSN, HH has close associates holding official position; *significant at 10%, **significant at 5%, ***significant at 1%.

Another significant determinant is whether the household received free food at any time in the past between 1982 and 1999. Households that did receive free food in the past had 14 pp higher probability of receiving free food between 2002 and 2004. Again, this is consistent with official targeting rules that suggest taking into account earlier food aid receipts to determine eligibility, even if it reflects inertia, and the potential for serious targeting errors (Jayne et al., 2002). Jayne et al. found inertia at the community level – and our results show that this finding carries through to the household level. This finding may also be influenced by unobserved omitted variables: for example, these households may have a time-invariant, but to the researcher unobserved, need for continuous food aid, while it may be observable for the *kebele* administration.

The results suggest that social networks in general are related to the food distribution process in September 2002–March 2004. The Wald statistic for joint significance of the social network variables is 11.89, which is equivalent to a *p*-value of 0.01. But the results also show that not all types of social networks matter equally. Households with political connections (vertical networks) had 12 pp higher probability of obtaining food aid in 2002–4 than households that did not have associates holding official positions. The magnitude of this effect may seem small at first, but the practical importance of this effect becomes clear if it is compared with the effect of the other significant determinants of food aid targeting. Compare this effect with the effect of consumption, for instance. A 1% increase in consumption per capita is associated with a decrease in the probability of obtaining food aid of 0.08 pp. This implies that, *ceteris paribus*, someone can be 150% richer but have political connections and still receive food aid with probability equal to that of someone else with no political connections.

The number of people that the household can rely on in times of need is significantly negatively related to food aid. For each additional person the household could rely on in times of need, the probability of obtaining food aid in 2002–4 decreased by 0.7 pp. The

significance level does not decrease if it is included separately in the regression (not reported). This result implies that households with, for instance, 20 people to rely on have, *ceteris paribus*, 14% less probability of obtaining food aid. Given that the number of people to rely on ranges from 0 to 205 in our sample, with an overall average of 7.8 and more than 14% of the sample having more than 15 persons to rely on, the magnitude of this effect can become quite significant for those persons with a large ISSN. The third social network type, horizontal connections measured by the size of the largest *iddir* that the household belonged to prior to the drought, does not seem to be related to the food distribution process.

In order to explore the hypothesis that the effects of social networks are relatively more important in the period immediately after the drought and become less important sometime after the drought, estimations in Table 5.3, columns 2 and 3, disaggregate the analysis into two different time periods: period 1 denotes the period including the first 7 months immediately after the drought (September 2002–March 2003) and period 2 denotes the 12 months immediately after period 1 (April 2003–March 2004). The results are strikingly consistent with our predictions outlined in section 5.3. In period 1, means-based targeting seems to be very weak, with the consumption effect being insignificant and land ownership and work ability of the household being only weakly significant. Evaluated at the mean value of all other characteristics, households at the 25th, 75th and 95th percentiles of consumption levels have, respectively, 42%, 39% and 37% probability of obtaining free food. As predicted, targeting in period 1 seems to be based on information more readily available, such as whether the household experienced a drought and whether a household member died in the period preceding the drought. Moreover, social and political networks seem to play a significant role in period 1. As expected, in period 2 the targeting pattern seems to change in favour of the most vulnerable households. Social and political network effects are no longer significant.

Poorer households have a significantly higher probability of obtaining public food aid than non-vulnerable households in period 2, irrespective of their political connections and their ISSN. Again evaluated at the mean values of all other characteristics, households at the 25th, 75th and 95th consumption percentiles have, respectively, 46%, 35% and 28% probability of obtaining free aid.[86]

So far we have looked only at the probability of being selected for free food aid receipts, without considering potential differences in the amount of food aid received. Table 5.3, column 4, shows the OLS results, explaining the amount of food aid received in period 2 for those actually receiving food aid. Unfortunately, we do not have data available on the amount of food aid received in period 1, exactly the period when social network effects turned out to be highly significant, as shown in column 2. For period 2, we find that allocation of the level of aid is just noisy around the mean, or at least we find that the amount of food aid is unrelated to social networks and economic needs, with the exception that female-headed households are receiving approximately one-quarter less than the mean payment.

Finally, to investigate whether social network effects differ across welfare groups, owing to, for example, differences in network strength, we added interactions of the social network variables with log of consumption per capita. The results (not reported) indicate no significant differences in social network effects in consumption between poor and richer households.

5.5.3. Food for work targeting results (OLS)

Table 5.4 shows the estimation results of the same probit models as in Table 5.3 but now with the probability of participating in FFW as the dependent variable. The targeting mechanism used in FFW programmes seems to be rather different from that

[86] Nevertheless, these effects are still estimated with a high standard error, and a full pooling test between the two periods suggests that, overall, the differences are not significant, despite the significance of individual results.

used in FFD. For FFW, there does not seem to be any significant difference between period 1 and period 2 targeting patterns (as the discussion of the descriptive statistics already alluded to: most people who receive FFW do so in both periods). Participation seems to be exclusively based on labour supply characteristics, such as work ability, household size and the health of male household members. For FFW we find no evidence whatsoever of targeting based on economic need. This is consistent with the findings of other studies that analyse the targeting outcomes of FFW programmes in Ethiopia (Gebremedhin and Swinton, 2000). None of the social networks seem to have any significant effect on the probability of participating in FFW. This is rather surprising, given that we would expect, in line with Granovetter's (1995) early findings, networks to be generally connected with job acquisitions. Overall, these results are consistent with a view of the operation of FFW in which self-selection on the basis of labour supply characteristics is central but at wages that are likely to be too high for any meaningful self-selection on the basis of need. The higher visibility of FFW compared with FFD (with people having to work to gain rations) may make selection based on political and social patronage harder to sustain.

Table 5.4 Determinants of food for work (FFW) participation

Dependent variable	FFW = 1 if HH participated in FFW 2002–4		FFW = 1 if HH participated in FFW, period 1		FFW = 1 if HH participated in FFW, period 2	
Estimation equation	(1) Probit		(2) Probit		(3) Probit	
VSN†	0.029	(0.046)	0.019	(0.048)	0.061	(0.048)
ISSN	–0.001	(0.002)	–0.001	(0.002)	–0.001	(0.002)
HSN	–0.006	(0.009)	–0.001	(0.009)	–0.003	(0.009)
Ln (consumption/capita)	0.028	(0.034)	0.058	(0.037)	0.022	(0.036)
Work ability score head	0.175***	(0.039)	0.185***	(0.042)	0.178***	(0.041)
Proportion elderly (>65 years)	0.164	(0.179)	0.046	(0.188)	0.214	(0.188)
Land area owned/capita	0.156	(0.134)	0.271*	(0.148)	0.185	(0.139)
Ln (HH size)	0.117**	(0.056)	0.130**	(0.059)	0.095	(0.059)
HH head is female †	–0.101*	(0.053)	–0.081	(0.055)	–0.130**	(0.056)
Proportion children (<6 years)	0.066	(0.148)	0.127	(0.154)	0.038	(0.154)
HH head is literate†	0.030	(0.050)	–0.005	(0.052)	0.057	(0.051)
Major religious group †	0.005	(0.058)	–0.016	(0.060)	–0.019	(0.060)
FF (1982–99)†	–0.013	(0.071)	–0.044	(0.073)	0.031	(0.075)
Drought shock (2000–2)†	0.067	(0.057)	0.102*	(0.058)	0.076	(0.058)
Ill male (1999–2002)†	–0.201**	(0.094)	–0.249***	(0.091)	–0.238***	(0.090)
Ill female (1999–2002)†	–0.004	(0.088)	–0.052	(0.092)	0.019	(0.089)
Member died (1999–2002)†	–0.009	(0.053)	–0.014	(0.055)	–0.015	(0.055)
N	602		602		602	
Pseudo R^2	0.172		0.189		0.193	

Notes: All models in this table include peasant association dummies (not reported). Period 1, September 2002–March 2003; period 2, April 2003–March 2004. Reported coefficients are marginal effects. Changes in marginal probabilities for continuous variables and discrete changes in the probability from changing the value from 0 to 1 for dummy variables (marked with †). Consumption per capita is measured by total consumption minus transfers in 1999. *Iddir* size is scaled (divided by 50). Numbers in parentheses are robust standard errors.

FF, free food; HH, household; HSN, size of the largest *iddir* the HH belongs to; ISSN, number of persons a HH can rely on in times of need; PA, peasant association; VSN, HH has close associates holding official position; *significant at 10%; **significant at 5%; ***significant at 1%.

The results in Table 5.4 are restricted to explaining the probability of participating in FFW. They do not tell us anything about the total number of days worked in FFW, nor about the amount of food or cash received as rewards under the FFW scheme. Table 5.5 allows us to investigate this further. The OLS estimation results of a model using the total value of payment for FFW of those who actually participated in the programme suggest that political connections have a significantly positive effect on the total value of food/cash received in period 2, with households that are politically connected having received on average 96 birr (or approximately one-quarter of a standard deviation, and 28% of the mean) more than non-politically connected households. This effect may be driven by two different sources: (i) households that are politically connected worked at the same daily wage rate as households that are not politically connected, but with the former group working relatively more in terms of the number of days worked; (ii) households that are politically connected perhaps worked the same number of days as those that are not politically connected, but with the former working at a daily wage rate that was higher than that paid to the latter. It is also possible that both effects occurred simultaneously. Exploring this further, Table 5.5, columns 2 and 3, indicates that the second of these two effects dominated in period 2. Households with at least one close associate holding an official position received on average 1 birr (or approximately one-third of both the mean and the standard deviation) more per day worked on FFW than households that did not have such connections. We do not find evidence that politically connected households worked significantly more in FFW than households without such connections.

Again, regressions adding interactions of the social network variables to wealth variables indicate no heterogeneity in effects across different welfare groups (not reported).

Table 5.5 Determinants of food for work (FFW) rewards among FFW-participating households (OLS)

Dependent variable	Total value of food/cash received period 2, OLS		Total number of days worked in FFW, period 2 OLS		Daily wage rate FFW (value in birr), period 2 OLS	
Estimation equation	(1)		(2)		(3)	
VSN	96.052*	(53.084)	4.512	(7.255)	0.943***	(0.330)
ISSN	–4.186**	(1.875)	–0.707**	(0.288)	–0.011	(0.016)
HSN	–3.258	(4.386)	0.687	(0.695)	–0.042	(0.040)
Ln (consumption/capita)	–11.734	(24.003)	–1.406	(4.565)	–0.105	(0.201)
Work ability score head	22.857	(27.234)	3.103	(4.477)	–0.003	(0.233)
Proportion elderly (>65 years)	92.865	(198.795)	–2.968	(29.517)	0.652	(1.392)
Land area owned/capita	–9.530	(114.982)	35.762	(27.333)	–0.326	(0.741)
Ln (HH size)	84.680**	(37.613)	27.017***	(7.033)	0.137	(0.324)
HH head is female	55.710	(43.021)	1.933	(7.597)	0.210	(0.305)
Proportion children (<6 years)	–70.811	(103.319)	0.873	(18.643)	–0.373	(0.847)
HH head is literate	31.590	(43.003)	5.056	(6.659)	0.190	(0.342)
Major religious group	–82.859	(74.232)	–18.707	(13.730)	0.022	(0.363)
FF (1982–99)	–25.451	(107.057)	9.080	(12.719)	0.036	(0.477)
Drought shock (2000–2)	–115.559*	(69.551)	10.409	(8.305)	–0.891	(0.585)
Ill male (1999–2002)	123.616	(106.066)	–5.599	(11.294)	1.044	(0.723)
Ill female, 1999–2002	–5.474	(61.922)	14.442	(14.114)	–0.108	(0.447)
Member died (1999–2002)	13.911	(50.905)	18.542**	(9.145)	–0.293	(0.365)
N (FFW participants only)	344		334		334	
Adjusted R^2	0.340		0.410		0.384	

Notes: All models in this table include peasant association dummies (not reported). Period 2, April 2003–March 2004. Income per capita is measured by total income minus transfers in 1999. *Iddir* size is scaled (divided by 50). Nine missing and one outlier observations are dropped in regressions (2) and (3). Numbers in parentheses are robust standard errors. Mean values (standard errors in brackets): total value 341 (431); days worked 78 (73); and daily payment 3.80 (3.33).
FF, free food; HH, household; HSN, size of the largest *iddir* the HH belongs to; ISSN, number of persons a HH can rely on in times of need; PA, peasant association; VSN, HH has close associates holding official position; *significant at 10%; **significant at 5%; ***significant at 1%.

5.5.4. Identification concerns and 2SLS

It is plausible that social networks and political connections are affected by the food aid distribution process, making a causal interpretation of the results above problematic. For example, political leaders may use food aid distribution to forge connections and for different forms of vote buying. This is additionally complicated in the data as those on political connections were collected in 2004, after the food aid distribution under consideration here had taken place. This is actually much less of a problem than it may seem at first. The political leadership was in fact elected in 2001; it was simply in power throughout this period, and the subsequent election took place only in 2005. As a result, we can be certain that the connections refer to the same leadership as that during the 2002 drought and its aftermath.

Still, it could be the case that the leadership forged additional connections during and after the 2002–3 drought by using food aid (e.g. to build up a support base in view of later elections). In itself, this is still a sign that the local political economy matters for food distribution, even though the effects in Table 5.3 can then not be simply interpreted as meaning that households with better political connections are managing to obtain food transfers, but it is possibly evidence of attempted vote buying. In other words, it refers to the interpretation of the observed correlation – and it may be a sign of reverse causality. A first step to explore the relevance of such reverse causality in the form of food aid distribution affecting political connections (and, more generally, all network variables) exploits the information on past food aid (pre-2000). In particular, note that, if there were reverse causality, we would expect not only current food aid programmes but also past food aid programmes to have influenced social networks in 2004. More specifically, we would expect the conditional correlation between social networks and past (pre-2000) food aid also to be significantly positive. As we found a positive conditional correlation between past food aid and the receiving food aid in

Table 5.3, the hypothesis of reverse causality would imply that omitting past food aid from the regression equation would introduce an upward bias in the coefficient estimates of the social network variables. When running a regression similar to (5.1) in Table 5.3 but now omitting past food aid (1982–99), we find that the coefficient estimates for all variables remain close to identical (not reported). In particular, the network coefficients are identical. Against this, one could argue that past food aid programmes since 1982 are less relevant for current vertical connections than recent food aid programmes.[87] But even if we restrict the analysis to past food aid programmes between 1997 and 1999, we obtain similar results (not reported). Given that past FFD does not seem to be linked to current FFW participation (as shown in Table 5.4), we do not expect running a model for FFW without past food aid to alter the results. Running the regression shows that, indeed, this is not the case (not reported).

However, this does not show that other sources of unobserved heterogeneity do not affect our inference on the impact of social network and political connections variables. It might be that social networks, being the result of behavioural decisions, are correlated with unobservable time-variant household characteristics that, in turn, are correlated with the probability of obtaining food aid.[88] For example, if wealth is correlated with political connections, and the current observable wealth variables imperfectly capture wealth, then the correlation between political connections and food aid access may be a reflection of unobserved heterogeneity in wealth. Given the recent political history, this is not necessarily what one would expect, with two relatively total reversals of economic and political power in rural Ethiopia since the 1970s. First, after the revolution in 1974, a total land reform saw the government taking full control of all

[87] For example, because those who were in charge of food distribution in the 1980s are no longer likely to be in charge in 2002.
[88] Note that time-invariant unobservable characteristics will be captured by our past food aid control to the extent that past food aid was targeted based on similar criteria.

land. The vast redistribution of land meant that rich and powerful families lost most of their land, and political power came into the hands of a new group of cadres. However, in 1991, the government was defeated by rebel forces; across the country political power came into the hands of a new political leadership, and many members of the previous political leadership across the country lost dramatically also in economic terms. While the leadership may be in an (economically) advantaged position, local political affiliation (in the form of being connected to the leadership) is not simply translated in wealth, not least since power has changed repeatedly in recent times.

Whether the link between political connections and access to food aid is largely via political processes in the community, rather than simply via unobservables, can be explored further. In particular, acknowledging that there may be a problem of unobserved heterogeneity, we use an instrumental variable approach to estimate the relationship in equation (5.1) to understand the relationship between political connections and food aid. We use a dummy for whether either parent of the household head has held an official position in the *kebele* or elsewhere. Since this variable relates to past political connections, it is unlikely to have an impact on the probability of obtaining food aid other than through current political connections inherited from the parents.[89]

Note that both our dependent variable (whether food aid is received) and our endogenous variable (whether the household has political connections) are binary

[89] In response to a suggestion made by one examiner in my DPhil transfer interview, we conducted the following informal check for the exclusion restriction of our instrument. We ran regressions (not reported) similar to regression (1) in Table 5.3, but where we replaced our measure for political connections, i.e. the dummy for whether or not the household has any close associates holding official positions, directly by our instrument, i.e. the dummy for whether any parent of the head held an official position. We find that this measure has a significantly positive effect on the probability of obtaining food aid similar in magnitude to the effect previously estimated, although at a lower significance level. Once controlling for the dummy for whether or not the household has any close associates holding official positions, however, the effect of parental political affiliation turns insignificant. This suggests that parental political affiliation only has an *indirect* effect on the probability of obtaining food aid for the children, i.e. through affecting the political connections of future generations.

variables. In the case of binary regressor(s), standard instrumental variable probit estimation leads to inconsistent estimates for reasons associated with 'forbidden regressions' (Wooldridge, 2002). There exist two common approaches to estimating causal effects in such models, each with its own potential drawbacks (Angrist, 2001; Bhattacharya et al., 2006; World Bank, 2011): two-stage least squares (2SLS) estimation disregards the binary structure of the variables and presents linear instrumental variable estimates. The advantage of the 2SLS estimator is its computational simplicity and the ability to use post-estimation tests for weakness of the instruments. The second approach, that is, the bivariate probit model, estimates both binary variables jointly using a standard maximum likelihood procedure. Both approaches yield very similar results for our analysis. To make use of standard post-estimation tests, we decide to focus on the 2SLS results in this chapter.

Columns 1 and 2–3 in Table 5.6 present the estimation results for, respectively, the first- and second-stage regressions of the 2SLS estimation process. We find that the instrument is statistically significant at 1%. With an implied *F*-statistic of 28.81, this does not seem to be a weak instrument (Staiger and Stock, 1997). Moreover, based on the Durbin–Wu–Hausman test we cannot reject the null hypothesis that the political connection variable is exogenous in both models. Households in which the parents of the head held official positions have almost 30 pp higher probability of having political connections than other households in the PA. As anticipated, the first-stage regression also shows how political connections appear to have only limited correlation with wealth variables or demographic variables, although female-headed households are clearly much less connected whereas households headed by a literate person are significantly more connected to the political elite.

Table 5.6 Instrumental variable (IV) estimations

Dependent variable	VSN		FFD = 1 if HH received free food, 2002–4		FFW = 1 if HH participated in FFW, 2002–4	
Estimation equation	(1) 2SLS: 1st stage		(2) 2SLS: 2nd stage FFD		(3) 2SLS: 2nd stage FFW	
Either of head's parents has held an official position†	0.277***	(0.053)				
VSN†			0.302*	(0.177)	0.066	(0.181)
ISSN	0.003	(0.002)	–0.003**	(0.001)	–0.001	(0.002)
HSN	0.009	(0.007)	–0.005	(0.008)	–0.005	(0.008)
Ln (consumption/capita)	–0.020	(0.030)	–0.062**	(0.030)	0.022	(0.029)
Work ability score HH head	0.045	(0.035)	–0.104***	(0.032)	0.149***	(0.035)
Proportion elderly (>65 years)	0.034	(0.167)	–0.004	(0.131)	0.110	(0.162)
Land area owned/capita	–0.132	(0.121)	–0.237**	(0.113)	0.123	(0.096)
Ln (HH size)	–0.045	(0.048)	–0.137***	(0.050)	0.099**	(0.048)
HH head is female†	–0.117**	(0.047)	0.016	(0.051)	–0.086*	(0.048)
Proportion children (<6 years)	–0.122	(0.125)	–0.037	(0.127)	0.072	(0.128)
HH head is literate†	0.111**	(0.045)	–0.019	(0.050)	0.013	(0.048)
Major religious group †	–0.022	(0.055)	–0.027	(0.052)	0.005	(0.055)
Received FF (1982–99)†	–0.059	(0.063)	0.134*	(0.073)	–0.011	(0.067)
Drought shock (2000–2)†	0.062	(0.049)	0.004	(0.051)	0.053	(0.050)
Ill male (1999–2002)†	–0.107	(0.082)	0.070	(0.076)	–0.165**	(0.084)
Ill female, 1999–2002†	0.043	(0.080)	0.042	(0.072)	–0.005	(0.077)
Member died (1999–2002)†	0.005	(0.047)	0.026	(0.048)	–0.001	(0.045)
N	602		602		602	
Pseudo R^2	0.189		0.187		0.202	
F-statistic instrument	28.81					

Notes: All models in this table include peasant association dummies (not reported). Consumption per capita is measured by total consumption minus transfers in 1999. *Iddir* size is scaled (divided by 50). Changes in marginal probabilities for continuous variables and discrete changes in the probability from changing the value from 0 to 1 for dummy variables (marked with †). Numbers in parentheses are robust standard errors. FF, free food; FFD, free food distribution; FFW, food for work; HH, household; HSN, size of the largest *iddir* the HH belongs to; ISSN, number of persons a HH can rely on in times of need; PA, peasant association; 2SLS, two-stage least squares; VSN, HH has close associates holding official position. *significant at 10%; **significant at 5%; *** significant at 1%.

The last two columns in Table 5.6 present the results for the 2SLS second-stage regressions for FFD and FFW.[90] The results concerning the network variables remain similar, as before, with political networks and the ISSN having a significant effect on the probability of obtaining free food, while having no significant effect on the probability of participating in FFW. The 2SLS point estimates are larger than implied by the probits in Tables 5.3 and 5.4, or the linear probability model without instrumenting (not reported), but also have larger standard errors. Given the size of these errors, the reduction in significance level seems to be due to the loss in estimation precision, but, in general, they broadly confirm our earlier findings. In short, there is no clear evidence that suggests that unobserved heterogeneity or other endogeneity problems affect our conclusions derived earlier.

It is also possible to explore further whether there is any evidence against reverse causality with respect to the link between obtaining food aid and the ISSN measure. In particular, in line with concerns about the crowding out of informal security by formal transfer schemes (Dercon and Krishnan, 2003), it may well be that the negative relationship between the size of the network in the data (measured in 2004) and receiving food aid is linked to the fact that those receiving food aid reduced their network size, so the correlation would then be just spurious. The data provide us with more direct evidence against this interpretation and, therefore, against reverse causality. In particular, the survey asked the households about not only the exact number of people they could rely on in 2004 but also whether their informal security network had remained the same, increased or decreased compared with 1999. Table 5.7 shows the frequency of the answers for the food-receiving and free-food-receiving household groups and the FFW participants and non-participants separately.

[90] The linear probability model is used in the first-stage regression, in line with the arguments in Angrist and Krueger (2001).

Table 5.7 Change in informal social safety net (ISSN), 1999–2004

		Share of households with no change in ISSN (%)	Share of households with increase in ISSN (%)	Share of households with decrease in ISSN (%)
FFD	Non-free-food-receiving group (232 observations)	30.60	34.05	35.34
	Free-food-receiving group (356 observations)	35.67	30.34	33.99
	t-statistic difference in shares between 2 groups	–1.27	0.94	0.34
FFW	Group NOT participating in FFW (223 observations)	30.94	37.22	31.84
	Group participating in FFW (365 observations)	35.34	28.49	36.16
	t-statistic difference in shares between 2 groups	–1.10	2.21**	–1.07
	Full sample (588)	33.67	31.80	34.52

Notes: FFD, free food distribution; FFW, food for work.
*significant at 10%; **significant at 5%; ***significant at 1%.

The results show that there is no statistically significant relationship between participating in the free food programme and the change in the household's support network. This suggests that participating in the free food aid programme has not systematically influenced the ISSN. As for FFW, there seems to be a significant difference in the share of households that claimed to have experienced an increase in ISSN since 1999. However, the difference seems to be in the opposite direction from what we would expect, as more households not participating in FFW seem to have increased their ISSN than those participating in FFW. Hence, this would not suggest that this has led to an overstatement of our results.

A final potential identification problem stems from possible measurement error. Given that our social network measures are based on data self-reported by the household, they are possibly affected by recall bias and/or bias due to cognitive

dissonance (i.e. when respondents make different survey answers agree). If this bias is related to specific household characteristics that affect the food aid distribution process, then our social network results may be spurious. For instance, if poor households tend to under-report the size of their ISSN (e.g. in the hope of obtaining more food aid in the future), then a spurious negative relationship between the probability of obtaining food and the size of the ISSN would arise if local officials target poor households irrespective of their ISSN size. Similar reasoning applies to the political network effect. However, the analysis in this section suggests that measured social networks are only weakly correlated with observable demographic and wealth characteristics of the households. This suggests that any measurement error is unlikely to bias our estimation results upwards.

5.6. Channels: improved information or favouritism?

Our empirical analysis has revealed a strong correlation between political affiliation and the probability of obtaining food aid in Ethiopia. As argued in section 5.3, political connections may influence the food allocation process through different pathways. On the one hand they may improve information flows: vertical connections may provide households with direct access to information about beneficial services and may give signals to local officials about the household's need. On the other hand, political connections can be a source of favouritism. Distinguishing between these explanations is not self-evident, and our dataset does not allow us to properly identify the relevant channels. Instead, this section offers some suggestive evidence, which turns out in favour of the favouritism story.

First, our analysis has shown that households with political connections receive, *ceteris paribus*, significantly higher daily salaries for FFW projects than households without any connections to the political elite. It would be hard to justify this difference solely on the basis of improved information flows for politically connected households.

Second, if connections were used by PA officials to better inform them about the eligibility status of the villagers in their communities, we would expect the impact of political affiliations to be positive for eligible households and negative for ineligible households. This pattern would not arise with favouritism. Lacking objective eligibility data, I proxy eligibility by economic status and look at estimation results of regressions adding interactions between the log of ex-ante consumption per capita and political connections (not reported). I do not find any significant differences across welfare groups in terms of the benefits of being connected to the political elite. Recognising that economic status is only an imperfect proxy for eligibility, I remain careful in deriving firm conclusions from these results. However, the evidence available is more consistent with favouritism rather than with the information sharing interpretation of the results.

Third, we collected some qualitative data on the sampled households' perception of the food allocation process during the period 2002–4. We considered the answers to the question in which the respondents were asked, in 2004, whether they would describe the recent food aid allocation process as 'very fair', 'somewhat fair', 'just ok', 'somewhat unfair' or 'very unfair'. We computed the total share of households responding either 'somewhat unfair' or 'very unfair'. On average, 44% of all sampled households described the food aid allocation process as unfair. Although we do not find any significant differences in answers between the households with and those without any political affiliations, we do find a significant difference between the non-food-receiving (52%) and food-receiving (38%) groups in this share, consistent with perceptions of favouritism and capture in FFD. However, as it can just as well be evidence of unhappiness with outcomes, we need to be careful in interpreting this result. We also asked the respondents how they would describe the information they had about how food assistance would be allocated. On average, 27% of all sampled households reported not to have had any idea and 38% reported to have had some idea. The

remaining 35% of the households reported to have been well informed. We do not find significant differences between politically connected and non-connected households in the share of households that reported not to have had any idea. Of the remaining households, a slightly *lower* share of households in the well-connected cohort reported to have been well informed as opposed to 'having some idea' than in the non-connected cohort. So, if anything, the politically connected households seemed to have been *less* informed than the households without any political connections.

Finally, one plausible hypothesis could be that, if connections matter for information, they are likely to be more important in larger communities, where information flows may be harder. In order to explore this, we considered a further specification, in which we added interaction terms between our political connection variable and the size of the community (PA) where the household lives.[91] While all other effects remain very similar, we find that the political connection effect decreases with the size of the PA: the larger the PA, the smaller the positive effect of political connections on the probability of obtaining food aid (not reported). Although statistical identification of the interaction terms is expected to be weak given the small number of PAs in our dataset, this evidence, together with the evidence described above, makes the favouritism interpretation harder to dismiss.

5.7. Conclusion

A large part of the Ethiopian population continues to regularly rely on transfers in the form of food aid, delivered via donors and the government. There is little known about how the local political economy and networks affect their allocation, even though targeting is largely administrative based and implemented by the local political leadership. This chapter offered evidence on these processes.

[91] Note that PA size as a level effect is captured by the community fixed effects.

Official rules and commitments make it clear that free food aid is targeted to those at risk of hunger and poverty with limited access to alternative means of support, including via working on FFW programmes or other forms of support. Most evidence suggests that targeting is correlated with observable characteristics reflecting needs, but that it is imperfect (Jayne et al., 2001; Gilligan and Hoddinott, 2007). In practice, targeting of food aid distribution in Ethiopia depends on local political leaders within the *kebele* administration; in principle, these leaders are elected but in reality they are typically closely linked to those in power regionally and nationally. Because of a lack of objectively verifiable information on need, effective targeting also depends on the needy being able to effectively communicate this need or to enforce their rights.

We studied the role of social networks and political connections in the allocation of food aid in rural Ethiopia, focusing on food aid distributed in the aftermath of one of the largest relief operations in recent years. We used data from a household panel data survey that collected information on food aid transfers before and after a serious drought in 2002. Controlling for baseline characteristics typically correlated with the categories of people that rules suggest should be targeted, we find that those with larger social networks for informal insurance have less access to food aid. On average, for each additional person the household can rely on, the probability of obtaining food aid decreases by almost 1 percentage point. More strikingly, we find that households in 'vertical' networks, measured by households having close associates who hold official positions, have 12 percentage points higher probability of obtaining food aid than other households in the PA that are not vertically connected. We find strong differences between the immediate post-drought period, in which political connections are strong and significant, as well as directly observable factors related to temporary need, such as having suffered from the drought, and the subsequent 12 months, in which more general 'need' variables dominate. Nevertheless, even in this later period, political connections

appear to matter in terms of the size of the transfers received. The allocation of FFW appears to be based largely on labour supply characteristics linked to ability to work and not on need or political or social connections. However, the level of support received, in terms of total or daily payment received, appears to be linked to political networks.

There are many ways in which this relationship can be understood. It could be informational: being connected with those holding official positions may improve information flows, for example, in terms of signalling need or availability of support. This idea is not supported by subjective data that we have available. As these PAs comprise, usually, not more than approximately 500 households, and also given some of the tests conducted in this chapter, improved information is unlikely to be the full story. Favouritism is a plausible alternative explanation, irrespective of whether those households actually need the food aid or not. This is a striking result, leaving less well-connected households with real needs without support.

Regarding the policy implications of these results, it is important to recognise that social networks may be an essential channel of access to a particular service. Considering connections to the political elite as a requisite for one to be able to express one's needs or to benefit from preferential treatment, community members may purposefully create social networks to maximise the chances of benefitting from a service. Although this potential source of endogenous link creation would not alter the conclusion that political connections matter in the distribution of food aid – as it is exactly the fact that it matters that would be driving the link in the first place - the policy implications would be different. If social networks indeed form an essential channel of access to food aid in Ethiopia, it would probably be desirable to design alternative procedures – easily accessible to all - through which needy community members can rapidly express their needs in disaster situations such as droughts.

Finally, it is important to note that finding evidence of the impact of political affiliations on food aid receipt does not mean that political connections make food aid targeting less pro-poor. Whether this link is legitimate depends on the distribution of political connections in the economy. In fact, as we argued in this chapter, at the time of the 2002–4 droughts local political affiliation in Ethiopia was unlikely to be correlated with wealth. Nevertheless, while giving food aid to poor households one is politically connected to may be considered favourable (e.g. on the ground of making use of improved information), giving food aid to well-connected wealthy households may be less justifiable in a context where poor households with real needs are left without support. We have to be careful not to overstate these results, as correlations with 'need' variables remain important. But, in future programmes, it will be worth looking deeper into the political economy of these villages. Searching for ways to increase the accountability of local *kebele* committees will be required if improved targeting is to be achieved.

Chapter 6

Conclusion

The aim of this thesis was twofold. First, it aimed to provide further empirical evidence on the role of social networks in the targeting performance of community-based development programmes in sub-Saharan Africa – in view of their potential to improve information flows and/or alter local targeting preferences. Second, it addressed some methodological concerns that currently exist in the empirical literature. This concluding chapter brings together the main results and summaries of this investigation and provides some key lessons for future research and policy-making processes.

Using a randomised survey experiment in Tanzania, Chapter 2 examined the differences between computer-assisted personal interviewing (CAPI) and pen-and-paper interviewing (PAPI) survey applications along various dimensions. We found that the PAPI data contained a relatively high number of errors, which were correlated with household characteristics such as gender of the household head and household size. This correlation is likely to induce sample bias if dubious observations are dropped. Whereas the quality of PAPI data depends on the interviewer characteristics (such as the level of education and survey experience), the various CAPI features make the CAPI data insensitive to differences in enumerator skills. We also showed that interview time is reduced in CAPI by 10%, possibly as a result of the automated routing and the use of drop-down menus. Respondents' perceptions are independent of the survey application.

We further demonstrated a tendency for the spread of total measured consumption to be higher on paper than on CAPI, translating into significantly higher measured inequality and lower poverty estimates. Investigating further the nature of the PAPI technique's measurement error for consumption, we fail to reject the hypothesis that it is classical: It attenuates the coefficient on consumption when used as an explanatory

variable and we find no evidence of bias when consumption is used as the dependent variable.

Given the relatively large fixed upfront costs, CAPI is, budget-wise, probably more viable for larger surveys. However, in light of the results in Chapter 2, and given the current pace at which CAPI costs are decreasing, CAPI is likely to become the more desirable survey option in the future.

Chapter 3 formally assessed a particular ordinary least squares (OLS) bias in linear-in-means peer effects models, the exclusion bias, the intuition for which was recently introduced to the literature by Guryan et al. (2009). We derived an analytical expression for the magnitude of the bias in a model of random peer assignment, which revealed that the bias increases with peer group size and decreases with the size of the pool from which peers are drawn. Moreover, whenever peer group formation is correlated with cluster formation, the exclusion bias becomes more severe in models adding cluster fixed effects (such as models adding classroom dummies when peers are selected from within the class). Importantly, in the latter empirical specification, the magnitude of the bias does not reduce as the sample size increases to infinity.

After discussing various limitations to the solution suggested by Guryan et al. (2009), we show how two-stage least squares (2SLS) estimation strategies are generally more effective in eliminating the exclusion bias. The latter result may explain the common, yet counter-intuitive, trend in the social interaction literature of observing 2SLS estimates that are higher than OLS estimates of the endogenous peer effects. Overall, the results of Chapter 3 call for more recognition of the consequences of the exclusion bias in the social interaction literature. At the minimum, the researcher should be more cautious against naive interpretations and comparisons of estimation results between models that are differently affected by the exclusion bias.

Chapter 4 highlighted that, in rural Tanzania, the elderly and disabled have, as a result of their physical and socioeconomic conditions, only restricted access to primary sources of information about development programmes (such as village meetings and newspapers). Using a unique spatial dataset collected at the baseline of an impact evaluation of the Tanzania Social Action Fund II (TASAF II), we examined the extent to which social interactions with well-informed neighbours may compensate for the lack of direct access to information.

Using a popular 2SLS estimation strategy, controlling for hamlet fixed effects and observable household and neighbours set characteristics, we identified large and significant average endogenous peer effects in TASAF awareness. In my preferred specification, for every additional informed person in the nearest 10 sampled neighbours set, the vulnerable household was almost 8 pp more likely to be informed. These estimated peer effects are strongest for households that have blood relatives living in the village and/or whose head is a member of the main religion and/or educated and/or employed. Unsurprisingly, neighbours who hold political positions in the village are most effective in transmitting information to the elderly and disabled.

However, Chapter 4 also identified several demographic groups that, in addition to having restricted access to traditional information sources, are less sensitive to information in the neighbourhood, such as female-headed households living in male-dominated neighbourhoods. For these households to have any voice in community-based development programmes, it is important that sensitisation agents take these concerns into consideration when designing information campaigns.

Finally, Chapter 5 considered the role of social networks in the community-based food aid allocation process in Ethiopia, in the aftermath of a serious drought (2002–2003). We found that, although targeting is clearly imperfect, free food aid is responsive to need, as well as targeted to households with less access to support from relatives or

friends. Nevertheless, we also demonstrated a strong correlation between food aid percipiency and political connections: households with close associates in official positions have 12 percentage points higher probability of obtaining free food than households that are not well connected. This effect is large as it is equivalent to the impact of a decrease in the initial living standard of 150% in obtaining food aid. The correlation with political connections is specifically strong in the immediate aftermath of a drought. Payment for food for work is also approximately one-third higher for those with political connections. Although the food aid programmes in Ethiopia appear to be responsive to need, Chapter 5 warns policy-makers that, in future, it is crucial to look more closely at the local political economy of these processes.

References

Aaronson, D., 1998. Using sibling data to estimate the impact of neighborhoods on children's educational outcomes. *Journal of Human Resources* 33 (4), 915–946.

Aizer, A., 2007. Public health insurance, program take-up and child health. *Review of Economics and Statistics* 89 (3), 400–415.

Alderman, H., 2002. Do local officials know something we don't? Decentralization of targeted transfers in Albania. *Journal of Public Economics* 83, 375–404.

Angrist, J.D., 2001. Estimation of limited-dependent variable models with binary endogenous regressors: simple strategies for empirical practice. *Journal of Business and Economic Statistics* 19, 2–28.

Angrist, J.D. and Krueger, A.B., 2001. Instrumental variables and the search for identification: from supply and demand to natural experiments. *Journal of Economic Perspectives* 15, 69–87.

Angrist, J. and Lang, K., 2004. Does school integration generate peer effects? Evidence from Boston's Metco Program. *American Economic Review* 94 (5), 1613–1634.

Araujo, M.C., Ferreira, F.H.G., Lanjouw P. and Ozler, B., 2008. Local inequality and project choice: theory and evidence from Ecuador. *Journal of Public Economics* 92 (5–6), 1022–1046.

Asfaw, S., 2006. Does food aid reach the poor? New evidence from northern Ethiopia. Development and Agricultural Economics, University of Hannover. Available at http://www.tropentag.de/2006/abstracts/full/27.pdf.

Attanasio, O. and Rios-Rull, J.-V., 2000. Consumption smoothing in island economies: can public insurance reduce welfare? *European Economic Review* 44 (7), 1225–1258.

Baird, S., McIntosh, C. and Ozler, B., 2011. The regressive demands of demand-driven development. Policy Research Working Paper no. WPS 5883. World Bank, Washington, DC.

Bandiera, O. and Rasul, I., 2006. Social networks and technology adoption in northern Mozambique. *Economic Journal* 116 (514), 869–902.

Banerjee, A.V., Banerji, R., Duflo, E., Glennerster, R. and Khemani, S., 2010. Pitfalls of participatory programs: evidence from randomized experiments in education in India. *American Economic Journal: Economic Policy* 2 (1): 1-30.

Banks, R. and Laurie, H., 2000. From PAPI to CAPI: the case of the British Household Panel Survey. *Social Science Computer Review* 18 (4), 397–406.

Bardhan, P. and Mookherjee, D., 2000. Capture and governance at local and national levels. *American Economic Review* 90 (2), 135–139.

Bardhan, P. and Mookherjee, D., 2006. Pro-poor targeting and accountability of local governments in west Bengal. *Journal of Development Economics* 79 (2), 303–327.

Barrett, C. and Clay, D., 2003. How Accurate is food-for-work self-targeting in the presence of imperfect factor markets? Evidence from Ethiopia. *Journal of Development Studies* 39 (5), 152–180.

Bayer, P. and Ross, S., 2009. Identifying individual and group effects in the presence of sorting: a neighborhood effects application. Economic Research Initiatives at Duke (ERID) Working Paper No. 51. Duke University, Durham, NC. Available at http://ssrn.com/abstract=1602011 or http://dx.doi.org/10.2139/ssrn.1602011.

Bayer, P., Hjalmarsson, R. and Pozen, D., 2009. Building criminal capital behind bars: peer effects in juvenile corrections. *Quarterly Journal of Economics* 124 (1), 105–147.

Bazzi, S. and Clemens, M., 2009. Blunt instruments: a cautionary note on establishing the causes of economic growth. Working Paper 171. Center for Global Development, Washington, DC.

Berg, E., Ghatak, M., Manjula, R. and Rajasekhar, D., 2013. Motivating knowledge agents: can incentive pay overcome social distance? Working paper. Available at http://erlendberg.info/agentincentives.pdf.

Besley, T. and Coate, S., 2003. Elected versus appointed regulators: theory and evidence. *Journal of the European Economic Association* 1 (5), 1176–1206.

Besley, T., Pande, R. and Rao, V., 2005. Participatory democracy in action: survey evidence from south India. *Journal of the European Economic Association* 3 (2–3), 648–657.

Bhattacharya, J., Goldman, D. and McCaffrey, D., 2006. Estimating probit models with self-selected treatments. *Statistics in Medicine* 25 (3), 389–413.

Bjorkman, M. and Svensson, J., 2007. Power to the people: evidence from a randomized field experiment of a community-based monitoring project in Uganda. Policy Research Working Paper Series 4268, The World Bank.

Black, D.A., Berger, M.C. and Scott, F.A., 2000. Bounding parameter estimates with nonclassical measurement error. *Journal of the American Statistical Association* 95 (451), 739–748.

Blume, L.E., Brock, W.A., Durlauf, S.N. and Ioannides, Y.M., 2010. Identification of social interactions. Economies Series 260. Institute for Advanced Studies, Vienna.

Bold, T. and Dercon, S., 2009. Contract design in insurance groups. Economics Series Working Paper no. 421. Department of Economics, University of Oxford, UK.

Borjas, G.J., 1992. Ethnic capital and intergenerational mobility. *Quarterly Journal of Economics* 107 (1), 123–150.

Borjas, G.J., 1995. Ethnicity, neighborhoods, and human-capital externalities. *American Economic Review* 85 (3), 365–390.

Bound, J. and Krueger, A.B., 1991. The extent of measurement error in longitudinal earnings data: do two wrongs make a right? *Journal of Labor Economics* 9 (1), 1–24.

Bound, J., Brown, C., Duncan, G.J. and Rodgers, W.L., 1994. Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics* 12 (3), 345–368.

Bound, J., Brown, C. and Mathiowetz, N., 2001. Measurement error in survey data. In: Heckman, J. and Leamer, E. (eds), *Handbook of Econometrics*, vol. 5. Elsevier, Amsterdam, The Netherlands, pp. 3705–3843.

Bramouille, Y., Djebbari, H. and Fortin, B., 2009. Identification of peer effects through social networks. *Journal of Econometrics* 150 (1), 41–55.

Brock, W. and Durlauf, S., 2001. *Handbook of Econometrics*. Elsevier Science, Amsterdam, pp. 3299–3380.

Brock, W.A. and Durlauf, S.N., 2007. Identification of binary choice models with social interactions. *Journal of Econometrics* 140 (1), 52–75.

Brown, K.M. and Laschever, R.A., 2012. When they're sixty-four: peer effects and the timing of retirement. *American Economic Journal: Applied Economics* 4 (3), 90–115.

Calvo-Armengol, A., Patacchini E. and Zenou, Y., 2009. Peer effects and social networks in education. *Review of Economic Studies* 76 (4), 1239–1267.

Capéau, B. and Dercon, S., 2006. Prices, unit values and local measurement units in rural surveys: an econometric approach with an application to poverty measurement in Ethiopia. *Journal of African Economies* 15 (2), 181–211.

Case, A., 1992. Neighborhood influence and technological change. *Regional Science and Urban Economics* 22, 491–508.

Chesher, A. and Schluter, C., 2002. Welfare measurement and measurement error. *Review of Economic Studies* 69 (2), 357–378.

Cipollone, P. and Rosolia, A., 2007. Social interactions in high school: lessons from an earthquake. *American Economic Review* 97 (3), 948–965.

Clay, D.C., Molla, D. and Habtewold, D., 1999. Food aid targeting in Ethiopia: a study of who needs it and who gets it. *Food Policy* 24, 391–409.

Coady, D., Grosh, M. and Hoddinott, J., 2004. Targeting outcomes redux. *World Bank Research Observer* 19 (1), 61–85.

Coleman, J.S., 1988. Social capital in the creation of human capital. *American Journal of Sociology* 94, S95–S121.

Collin, M., 2013. Peer effects in the adoption of property rights: experimental evidence from urban Tanzania. Unpublished paper presented at a conference held at the Centre for the Study of African Economies, University of Oxford, UK.

Conley, T.G., 1999. GMM estimation with cross sectional dependence. *Journal of Econometrics* 92 (1), 1–45.

Conley, T.G. and Topa, G., 2002. Socio-economic distance and spatial patterns in unemployment. *Journal of Applied Economics* 17, 303–327.

Conley, T.G. and Udry, C.R., 2010. Learning about a new technology: pineapple in Ghana. *American Economic Review* 100 (1), 35–69.

Conning, J. and Kevane, M., 2002. Community-based targeting mechanisms for social safety nets: a critical review. *World Development* 30 (3), 375–394.

Couper, M.P. and Burt, G., 1994. Interviewer attitudes toward computer-assisted personal interviewing (CAPI). *Social Science Computer Review* 12 (1), 38–54.

Cox, D. and Jimenez, E., 1992. Social security and private transfers in developing countries: the case of Peru. *World Bank Economic Review* 6 (1), 155–169.

Cox, D., Hansen, B.E. and Jimenez, E., 2004. How responsive are private transfers to income? Evidence from a laissez-faire economy. *Journal of Public Economics* 88 (9–10), 2193–2219.

Cragg, J.G. and Donald, S.G., 1993. Testing identifiability and specification in instrumental variable models. *Econometric Theory* 9, 222–240.

Daponte, B., Sanders, S. and Taylor, L., 1999. Why do low-income households not use food stamps? Evidence from an experiment. *Journal of Human Resources* 34 (3), 612–628.

De Giorgi, G., Pellizzari, M. and Redaelli, S., 2010. Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics* 2 (2), 241–275.

de Leeuw, E.D., 2008. The effect of computer-assisted interviewing on data quality: a review of the evidence. Mimeo., Department of Methodology and Statistics, Utrecht University.

de Leeuw, E. and Nicholls II, W.L., 1996. Technological innovations in data collection: acceptance, data quality and costs. *Sociological Research Online* 1 (4).

de Melo, G., 2011. Peer effects identified through social networks. Evidence from Uruguayan schools. Working Paper, Department of Economics, University of Siena, 627, Department of Economics, University of Siena, Italy. Available at http://www.econ-pol.unisi.it/dipartimento/it/node/1606.

Dercon, S., 2002. Income risk, coping strategies and safety nets. *World Bank Research Observer* 17 (2), 141–166.

Dercon, S. and Krishnan, P., 2003. Food aid and informal insurance. *Economic Journal* 113 (486), C86–C94.

Dercon, S., De Weerdt, J., Bold, T. and Pankhurst, A., 2006. Group-based funeral insurance in Ethiopia and Tanzania. *World Development* 34 (4), 685–703.

De Weerdt, J., 2004. Risk-sharing and endogenous network formation. In: Dercon, S. (ed.), *Insurance Against Poverty*. Oxford University Press, Oxford, pp. 197–216.

De Weerdt, J. and Dercon, S., 2006. Risk-sharing networks and insurance against illness. *Journal of Development Economics* 81 (2), 337–356.

Disaster Prevention and Preparedness Commission, 2000. *National Food Aid Targeting Guidelines*. DPPC, Addis Ababa.

Duflo, E. and Saez, E., 2002. Participation and investment decisions in a retirement plan: the influence of colleagues' choices. *Journal of Public Economics* 85 (1), 121–148.

Duflo, E. and Saez, E., 2003. The role of information and social interactions in retirement plan decisions: evidence from a randomized experiment. *Quarterly Journal of Economics* 118 (3), 815–842.

Duflo, E., Dupas, P. and Kremer, M., 2011. Peer effects, teacher incentives, and the impact of tracking: evidence from a randomized evaluation in Kenya. *American Economic Review* 101 (5), 1739–1774.

Durlauf, S.N. and Fafchamps, M., 2008. Social capital. In: Aghion, P. and Durlauf, S. (eds), *Handbook of Economic Growth*. Elsevier, Amsterdam, ch. 26.

Durlauf, S.N. and Ioannides, Y.M., 2010. Social interactions. *Annual Review of Economics* 2, 451–478.

Emergency Events Database, 2007. *CRED International Disaster Database*. University Catholique de Louvain, Brussels. Available at www.emdat.be.

Evans, W.N., Oates, W.E. and Schwab, R.M., 1992. Measuring peer group effects: a study of teenage behavior. *Journal of Political Economy* 100 (5), 966–991.

Fafchamps, M. and Gubert, F., 2007. The formation of risk sharing networks. *Journal of Development Economics* 83 (2), 326–350.

Fafchamps, M. and Lund, S., 2003. Risk-sharing networks in rural Philippines. *Journal of Development Economics* 71 (2), 261–287.

Fafchamps, M. and Minten, B., 2002. Returns to social network capital among traders. *Oxford Economic Papers* 54, 173–206.

Fafchamps, M. and Quinn, S., 2012. Networks and manufacturing firms in Africa: initial results from a randomised experiment. Working Paper, Centre for the Study of African Economies, University of Oxford, UK.

Available at http://users.ox.ac.uk/~econ0087/networks.pdf.

Fafchamps, M., McKenzie, D., Quinn, S. and Woodruff, C., 2012. Using PDA consistency checks to increase the precision of profits and sales measurement in panels. *Journal of Development Economics* 98, 51-57.

Famine Early Warning System Network, 2003. *Estimating Meher Crop Production Using Rainfall in the 'Long Cycle' Region of Ethiopia*. FEWS-NET, USAID, Washington, DC.

Fletcher, J., 2012. Peer influences on adolescent alcohol consumption: evidence using an instrumental variables/fixed effect approach. *Journal of Population Economics* 25 (4), 1265–1286.

Fletcher, J.M. and Ross, S.L., 2012. Estimating the effects of friendship networks on health behaviors of adolescents. Working Paper no. 18253, National Bureau of Economic Research, Cambridge, MA.

Galasso, E. and Ravallion, M., 2005. Decentralized targeting of an antipoverty program. *Journal of Public Economics* 89 (4), 705–727.

Gaviria, A. and Raphael, S., 2001. School-based peer effects and juvenile behavior. *Review of Economics and Statistics* 83 (2), 257–268.

Gebremedhin, B. and Swinton, S.M., 2000. Reconciling food-for-work project feasibility with food aid targeting in Tigray, Ethiopia. *Food Policy* 26 (1), 85–95.

Gibson, J. and Kim, B., 2007. Measurement error in recall surveys and the relationship between household size and food demand. *American Journal of Agricultural Economics* 89 (2), 473–489.

Gilligan, DO. and Hoddinott, J., 2007. Is there persistence in the impact of emergency food aid? Evidence on consumption, food security, and assets in rural Ethiopia. *American Journal of Agricultural Economics* 89 (2), 225–242.

Glaeser, E.L., Sacerdote, B. and Scheinkman, J.A., 1996. Crime and social interactions. *Quarterly Journal of Economics* 111 (2), 507–548.

Glaeser, E.L., Sacerdote, B.I. and Scheinkman, J.A., 2002. The social multiplier. Discussion Paper no. 1968, Harvard Institute of Economic Research, Cambridge, MA. Available at http://post.economics.harvard.edu/hier/2002papers/2002list.html.

Goetz, A.M. and Jenkins, R., 2001. Hybrid forms of accountability: citizen engagement in institutions of public-sector oversight in India. *Public Management Review* 3 (3), 363-383.

Goux, D. and Maurin, E., 2007. Close neighbours matter: neighbourhood effects on early performance at school. *The Economic Journal* 117 (523), 1193–1215.

Graham, B.S., 2008. Identifying social interactions through conditional variance restrictions. *Econometrica* 76 (3), 643–660.

Graham, B.S. and Hahn, J.Y., 2005. Identification and estimation of the linear-in-means model of social interactions. *Economics Letters* 88 (1), 1–6.

Granovetter, M., 1975. *Getting a Job: A Study of Contacts and Careers*, 2nd edn. University of Chicago Press, Chicago, 1995.

Grimard, F., 1997. Household consumption smoothing through ethnic ties: evidence from Cote d'Ivoire. *Journal of Development Economics* 53 (2), 391–422.

Grootaert, C., Oh, G.-T. and Swamy, A., 2002. Social capital, household welfare and poverty in Burkina Faso. *Journal of African Economies* 2 (1), 4–38.

Guiso, L. and Schivardi, F., 2007. Spillovers in industrial districts. *Economic Journal* 117 (516), 68–93.

Guryan, J., Kroft, D. and Notowidigdo, N.J., 2009. Peer effects in the workplace: evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics* 1 (4), 34–68.

Halliday, T.J. and Kwak, S., 2012. What is a peer? The role of network definitions in estimation of endogenous peer effects. *Applied Economics* 44 (3), 289–302.

Hanushek, E.A., Kain, J.F., Markman, J.M. and Rivkin, S.G., 2003. Does peer ability affect student achievement? *Journal of Applied Econometrics* 18 (5), 527–544.

Haughton, J. and Khandker, S.R., 2009. Understanding the determinants of poverty. In *Handbook on Poverty and Inequality*, vol. 1. World Bank, pp. 145–160.

Helmers, C. and Patnam, M. 2012. Does the rotten child spoil his companion? Spatial peer effects among children in rural India. Working Paper WPS/2010-13, Centre for the Study of African Economies, University of Oxford, UK.

Hernanz, V., Malherbet, F. and Pellizzari, M., 2004. Take-up of welfare benefits in OECD countries: a review of the evidence. OECD Social, Employment and Migration Working Papers, 17.

Hoddinott, J., Dercon, S. and Krishnan, P., 2009. Networks and informal mutual support in 15 Ethiopian villages. *Institutional Economics Perspectives on African Agricultural Development* 38, 1–26.

Hoxby, C., 2000. Peer Effects in the classroom: learning from gender and race variation. Working Paper no. 7867, National Bureau of Economic Research, Cambridge, MA.

Human Rights Watch, 2010. *Development without Freedom: How Aid Underwrites Repression in Ethiopia*. Human Rights Watch, New York.

IRIS, 2011. Comparative assessment of software programs for the development of computer-assisted personal interview (CAPI) applications. LSMS-ISA report. University of Maryland, MD.

Jackson, M.O., 2008. *Social and Economic Networks.* Princeton University Press, Princeton, New Jersey.

Jayne, T.S., Strauss, J., Yamano, T. and Molla, D., 2001. Giving to the poor? Targeting of food aid in rural Ethiopia. *World Development* 29 (5), 887–910.

Jayne, T.S., Strauss, J., Yamano, T. and Molla, D., 2002. Targeting of food aid in rural Ethiopia: chronic need or inertia? *Journal of Development Economics* 68, 247–288.

Keefer, P. and Khemani, S., 2004. Why do the poor receive poor services? *Economic and Political Weekly* 39 (9), 935–943.

Khemani, S., 2007. Can information campaigns overcome political obstacles to serving the poor? In: Devarajan, S. and Widlung, I. (eds), *The politics of service delivery in democracies: better access for the poor*. Stockholm, Ministry of Foreign Affairs, Government of Sweden, pp. 56-69.

Kleibergen, F. and Paap, R., 2006. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics* 133 (1), 97–126.

Krishnan, P. and Patnam, M., 2012. Neighbours and extension agents in Ethiopia: who matters more for technology diffusion? Available at http://www.mpatnam.com/research.

Krishnan, P. and Sciubba, E., 2009. Links and architecture in village networks. *Economic Journal* 119 (4), 917–949.

Lalive, R. and Cattaneo, M.A., 2009. Social interactions and school decisions. *The Review of Economics and Statistics* 91 (3), 457–477.

Laschever, R., 2005. The doughboys network: Social interactions and labor market outcomes of world war I veterans, Mimeo, Northwestern University.

Lee, L.F., 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics* 140 (2), 333-374.

Lefort, R., 2007. Powers – *mengist* – and peasants in rural Ethiopia: the May 2005 elections. *Journal of Modern African Studies* 45 (2), 253–274.

Lentz, E.C. and Barrett, C.B., 2005. Food aid targeting shocks and private transfers among East African pastoralists. Working Paper. Cornell University, Ithaca, NY.

Lin, N., 2001. Building a network theory of social capital. In: Lin, N., Cook, K. and Burt, R.S. (eds), *Social Capital: Theory and Research*. Aldine de Gruyter, New York, pp. 3–31.

Lin, X., 2007. Peer effects and student academic achievement: An application of spatial autoregressive model with group unobservables, Mimeo, Department of Economics, Tsinghyua University, Beijing.

Luke, N. and Munshi, K., 2006. New roles for marriage in urban Africa: kinship networks and the labor market in Kenya. *Review of Economics and Statistics* 88 (2), 264–282.

Lyle, D., 2007. Estimating and interpreting peer and role model effects from randomly assigned social groups at West Point. *The Review of Economics and Statistics* 89 (2), 289–299.

Lyle, D., 2009. The effects of peer group heterogeneity on the production of human capital at West Point. *American Economic Journal: Applied Economics* 1 (4), 69–84.

Manski, C., 1993. Identification of endogenous social effects: the reflection problem. *Review of Economic Studies* 60 (3), 531–542.

Manski, C.F., 2000. Economic analysis of social interactions. *Journal of Economic Perspectives* 14 (3), 115–136.

Mansuri, G. and Rao, V., 2013. Localizing development: does participation work? Policy Report. World Bank, Washington, DC.

Mas, A. and Moretti, E., 2009. Peers at work. *American Economic Review* 99 (1), 112–145.

Ministry of Agriculture and Rural Development, 2004. *Productive Safety Net Programme: Programme Implementation Manual*. MOARD, Government of the Federal Democratic Republic of Ethiopia, Addis Ababa.

Moffitt, R.A., 2001. Policy interventions, low level equilibria, and social interactions. In: Durlauf, S.N. and Young, H.P. (eds), *Social Dynamics*. MIT Press, Cambridge, MA, pp. 45–82.

Munshi, K., 2004. Social learning in a heterogeneous population: technology diffusion in the Indian green revolution. *Journal of Development Economics* 73 (1), 185–213.

Munshi, K. and Rosenzweig, M., 2006. Traditional institutions meet the modern world: caste, gender, and schooling choice in a globalizing economy. *American Economic Review* 96 (4), 1225–1252.

Naguib, K., 2012. The effects of social interactions on female genital mutilation: evidence from Egypt. Working Paper.

Available at http://www.mendeley.com/profiles/karim-naguib.

Narayan, D. and Pritchett, L., 1999. Cents and sociability: household income and social capital in rural Tanzania. *Economic Development and Cultural Change* 47 (4), 871–897.

Nicholls II, W. and de Leeuw, E., 1996. Factors in acceptance of computer-assisted interviewing methods: a conceptual and historical review. In *Proceedings of the Survey Research Methods Section, American Statistical Association*. Available at http://www.amstat.org/sections/SRMS/Proceedings/papers/1996_130.pdf.

Office of Chief Government Statistician Zanzibar, 2006. *Zanzibar Household Budget Survey 2004/2005*. Office of Chief Government Statistician Zanzibar, Tanzania.

Olken, B., 2007. Monitoring corruption: evidence from a field experiment in Indonesia. *Journal of Political Economy* 115 (2), 200–249.

Pande, R., 2007. Understanding political corruption in low income countries. CEPR Discussion Paper no. DP6273. Centre for Economic Policy Research, London, UK.

Pankhurst, R., 1986. *The history of famine and epidemics in Ethiopia prior to the twentieth century*, London Central Books (for Relief and Rehabilitation Commission, Addis Ababa), London, UK.

Pausewang, S. and Aalen, L., 2002. Ethiopia 2001: local elections in the southern region. Nordem Report 03/2002, University of Oslo, Norway.

Platteau, J.P. and Gaspart, F., 2003. The risk of resource misappropriation in community-driven development. *World Development* 31 (10), 1687–1703.

Putnam, R., Leonardi, R. and Nanetti, R., 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press, Princeton, NJ.

Raudenbush, S.W. and Bryk, A.S., 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edn. Sage Publications Inc., Thousand Oaks, CA.

Robinson, J.A. and Baland, J.-M., 2008. Land and power: theory and evidence from Chile. *American Economic Review* 98 (5), 1737–1765.

Sacerdote, B., 2001. Peer effects with random assignment: results for Dartmouth roommates. *Quarterly Journal of Economics* 116 (92), 681–704.

Shankar, S., Gaiha, R. and Jha, R., 2010. Information and corruption: the National Rural Employment Guarantee Scheme in India. ASARC Working Paper 32-32, Australia South Asia Research Centre, Australian National University, Canberra.

Sharp, K., 1997. *Targeting Food Aid in Ethiopia*. Save the Children Fund (UK), Addis Ababa.

Sharp, K., 1998. *Between Relief and Development: Targeting Food Aid for Disaster Relief in Ethiopia*. Relief and Rehabilitation Network, Overseas Development Institute, London.

Sharp, K., Brown, T. and Teshome, A., 2006. *Targeting Ethiopia's Productive Safety Net Programme (PSNP)*. Overseas Development Institute (ODI) and the IDL Group, London.

Soetevent, A.R., 2006. Empirics of the identification of social interactions: an evaluation of the approaches and their results. *Journal of Economic Surveys* 20 (2), 193–228.

Staiger, D. and Stock, J.H., 1997. Instrumental variables regression with weak instruments. *Econometrica* 65 (3), 557–586.

Stock, J.H. and Yogo, M., 2005. Testing for weak instruments in linear IV regression. In: Stock, J.H. and Andrews, D.W.K. (eds), *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, Chapter 5. Cambridge University Press, Cambridge.

Tanzania Social Action Fund, 2006. *TASAF Annual Report 2005–2006*. Dar es Salaam, Tanzania.

Tanzania Social Action Fund, 2007. *TASAF Annual Report 2006–2007*. Dar es Salaam, Tanzania.

Tanzania Social Action Fund, 2008. *TASAF Annual Report 2007–2008*. Dar es Salaam, Tanzania.

Tanzania Social Action Fund, 2011a. *TASAF II Operational Manual*. Dar es Salaam, Tanzania.

Tanzania Social Action Fund, 2011b. *TASAF 4th Quarterly Report April–June 2011*. Dar es Salaam, Tanzania.

Taylor, S., 1998. Setting up computer-assisted personal interviewing in the Australian longitudinal study of ageing. *Statistical Science* 13 (1), 14–18.

Townsend, R.M., 1995. Consumption insurance: an evaluation of risk-bearing systems in low-income economies. *Journal of Economic Perspectives* 9 (3), 83–102.

Van den Broeck, K. and Dercon, S., 2011. Information flows and social externalities in a Tanzanian banana growing village. *Journal of Development Studies* 47 (2), 231–252.

Von Braun, J., Teklu, T. and Webb, P., 1998. *Famine in Africa: Causes, Responses, and Prevention*. Johns Hopkins University Press, Baltimore, MD.

Wang, L.C., 2010. Peer effects in the classroom: evidence from a natural experiment in Malaysia. Working Paper, Department of Economics Seminar Series, University of California, San Diego, CA.

Available at http://economics.missouri.edu/seminars/files/2010/100129.pdf.

Wooldridge, J., 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

World Bank, 2006. *World Development Report 2006: Equity and Development*. World Bank, Washington, DC.

World Bank, 2011. A practical comparison of the bivariate probit and linear IV estimators. Policy Research Working Paper WPS5601. World Bank, Washington, DC.

World Bank, 2012. *The World Bank 2012–2022 Social Protection and Labor Strategy. Resilience, Equity, and Opportunity*. World Bank, Washington, DC.

World Food Program, 1995. *World Food Program and Food Aid in Ethiopia.* World Food Program, Addis Ababa.

World Food Program, 2006. *Evaluation of the Response to the 2002–03 Emergency in Ethiopia*. World Food Program, Rome.

World Food Program, 2012. *2011 Food Aid Flows*. World Food Program, Rome.

Wydick, B., Karp Hayes, H. and Hilliker Kempf, S., 2011. Social networks, neighborhood effects, and credit access: evidence from rural Guatemala. *World Development* 39 (6), 974–982.

Yamano, T., Alderman, H. and Christiaensen, L., 2005. Child growth, shocks, and food aid in rural Ethiopia. *American Journal of Agricultural Economics* 87 (2), 273–288.

Zimmerman, D., 2003. Peer Effects in academic outcomes: evidence from a natural experiment. *The Review of Economics and Statistics* 85 (1), 9–23.