Essays in development economics: land rights, ethnicity and birth order

Matthew Collin

New College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Economics

Michaelmas Term 2012

Essays in development economics: land rights, ethnicity and birth order

Matthew Collin

New College
University of Oxford

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

Michaelmas Term 2012

Abstract

Aside from the introduction and conclusion, this thesis comprises four core chapters:

The first chapter investigates the presence of endogenous peer effects in the adoption of formal property rights. Using data from a unique land titling experiment held in an unplanned settlement in Dar es Salaam, Tanzania. I show a strong, positive impact of neighbour adoption on the household's choice to purchase a land title. I also show that this relationship holds in a separate, identical experiment held a year later in a nearby community, as well as in administrative data for approximately 45,000 land parcels in the same city. I also discuss possible channels, including the possibility of complementarities in the reduction in expropriation risk.

The second chapter examines the relationship between ethnic heterogeneity and the demand for formal land tenure. Using a unique census of two highly fractionalised settlements in Dar es Salaam, I show that households located near coethnics are significantly less likely to purchase a limited form of land tenure recently offered by the government. I attempt to address one of the chief concerns, endogenous sorting of households, by conditioning on a households choice of neighbors upon arrival in the neighborhood. These results suggest that close-knit ethnic groups may be less likely to accept state-provided goods if they can generate reasonable substitutes.

The third chapter is a short chapter which presents results from a recent policy experiment in Tanzania where formal land titles were provided to informal settlers at randomised prices. Land owners were also randomly assigned conditional discounts, which could only be applied if a woman was designated as owner or co-owner of the land in question. Results show that conditionality has no adverse effects on demand for land titles, yet drastically increases the probability a woman is included. We discuss the implications of these results for the expected bargaining power impacts of the intervention.

The final chapter investigates birth order effects on both anthropometric and education outcomes in a longitudinal survey of children from the Philippines. Birth order effects are present early in life for both outcomes, but attenuate as children approach adulthood. There is also evidence for nonlinear birth order effects, with both firstborn and lastborn children holding an advantage over middleborn children. These results are at odds with prevalent theories of birth order which predict lasting and monotonic differences in outcomes across children.

Word Count

The approximate number of words in this thesis is 89,790, excluding the bibliography. This is calculated using the number of words on page 206 (410) multiplied by the number of pages (219).

Declaration

Two chapters from this thesis (Chapters 2 and 4) use field experiment data from a project jointly designed and implemented with Daniel Ali Ayalew, Klaus Deininger, Stefan Dercon, Justin Sandefur and Andrew Zeitlin. At least some of this work will likely be turned into articles intended for publication with all or some of these collaborators as co-authors. The current draft version of Chapter 2 presented in the thesis was researched by me using the data and the experimental design. Chapter 4 is submitted as a jointly-authored chapter (with all of the collaborators listed above), although I have also contributed a large part of the work on this paper.

Acknowledgements

There have been innumerable people who have offered me support and help during the research and writing of this doctoral thesis. First and foremost there is Professor Stefan Dercon, who encouraged me to return to Oxford to start the DPhil. Either in his office at Queen Elizabeth House or over coffee in his home, Stefan never failed to provide critical yet optimistic supervision, and I never walked out of a meeting without a good idea of what to do next or the belief that finishing the DPhil would be possible. I am grateful to Stefan for all the time and support he has given me over the past five years, without which the task of writing a DPhil would have been much more arduous.

I am also grateful to both Justin Sandefur and Andrew Zeitlin for recruiting me to work on the land titling project in Dar es Salaam, which itself became the backbone for most of this thesis. Justin and Andy patiently introduced me to the basics of survey design, training and managing enumerators, as well as designing and running field experiments.

They also put up with many a frantic e-mail sent from Dar es Salaam and Justin's reassurances kept me sane during long hours of problem solving in the field. Andy also took time to provide crucial feedback, suggestions and ideas on most of my thesis chapters.

My fellow desk-mates and friends in the department, Bet Caeyers, Sebastian Königs, Stefano Carria, Gerhard Toews, Julien Labonne, Andrew Kerr and many others kept me entertained with stimulating and distracting conversation and also helped me think more critically about what I was working on. Bet especially has given me an incredible amount of support over the past few years, from reading through nearly all of my chapters, to rallying my friends to send me supportive messages and treats during my final DPhil crunch. My classmates and friends Naureen Karachiwalla and Martina Kirchberger were enormously helpful and supportive both before and after the PRS stage of the DPhil. My housemates Ricarda and Alex had to put up with my hermit-like behaviour and thesis-induced grumpiness during the last few months of my thesis, yet still endeavored to keep me in good humour.

Many, many thanks to Helene Bie Lilleør and her colleagues at the Rockwool Foundation for hosting me on numerous occasions, as well as Helene's helpful suggestions during my penultimate year of the DPhil. I am also grateful to Albert Park and Francis Teal for their helpful comments made during my Transfer of Status interview on what is now my fourth chapter of the thesis. Both James Fenske and Simon Quinn also gave me useful comments on my second chapter shortly before submission.

This thesis was supported financially by the Economic and Social Research Council, who generously provided me with several years of funding. I must also thank the department for choosing me as one of the recipients for this studentship.

Finally, I am ever grateful to my parents Richard and Thea Collin for continuously encouraging and supporting me, as well as for putting up with the occasional DPhil grumble. Dad continues to be the best editor I could ever ask for. Mom patiently delivered snack upon snack to me during visits home when I opted to spend most of my time working in my room. What have I have accomplished thus far in life would not have been possible without the love and guidance of these two, and so I dedicate this thesis to them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The mystery of property rights

In his seminal book on the benefits of giving the poor access to functional property rights systems, Peruvian economist Hernando de Soto ends the second chapter with an exuberantly optimistic piece of advice for policymakers:

> "Leaders of the Third World and former communist nations need not wander the world's foreign ministries and international financial institutions seeking their fortune. In the midst of their own poorest neighbourhoods and shanty towns there are - if not acres of diamonds - trillions of dollars, all ready to be put to use if only we can unravel the mystery of how assets are transformed into live capital."

While it difficult to discern whether de Soto actually set the agenda for property rights research or just wrote *The Mystery of Capital* at the right moment, economists have duly responded to his challenge to look for "acres of diamonds" (De Soto 2000). The result has been a research agenda heavily focused on identifying the impact of interventions which formalise informal property, such as land titling schemes. If de Soto's prognosis is correct, bringing property rights to the poor should not only give them an incentive to invest in land without fear of losing it later on, but it should also enable them to sell this newfound asset or use it as collateral to access credit. These three impacts, reducing expropriation risk, unlocking credit access and increasing transferability, have long been among the standard benefits for which researchers have been searching for (Besley 1995).

Informal urban settlements in developing countries, infamous for a unique combination of insecure tenure yet potentially high land values, seem to be an obvious place to go hunting for 'diamonds'. However, the evidence on the impact of land titling efforts has been notably mixed. While recent research has suggested that land titling does induce households to invest more in their property, these same studies invariably fail to find any convincing increase in access to credit, restricting the scope for the the transformative improvements for which many have hoped for (Field 2005; Galiani and Schargrodsky 2010). The evidence does not get substantially more compelling when we also consider rural, agricultural settings where research suggests that perceived, *de facto* property rights matter more than formal ones (Dercon and Ali 2007; Goldstein and Udry 2008). Despite this mixed evidence, the quest to find a substantial impact continues unabated as researchers move from passively taking advantage of natural experiments to actively encouraging governments to induce experimental variation in the rollout of property rights (Ayalew Ali and Goldstein 2011). Indeed, much of the evidence in this thesis was produced from an ongoing field experiment in urban Tanzania which ultimately aims to identify these impacts.

The continued preoccupation with impact analyses of property rights interventions has led to a loss of focus on what factors generate the demand for property formalisation in the first place. The pursuit of individualised tenure is thought to be a strongly endogenous process, with moves away from customary or other forms of collective landholding being driven primarily by land scarcity (Platteau 2000). In many developing countries, this process is most evident by the emergence of informal land contracts and deeds, such as those found in the informal settlements of Tanzania (De Soto and Cheneval 2006). It follows that land titling schemes should not be judged solely on their theoretical benefit, but also on their ability to generate efficiency gains above and beyond what the informal, endogenously-determined system has produced, be it based on customary tenure or a pseudo-formal system often found in urban slums. Local demand for titling plays an important role in this, as low demand for large-scale titling might be an important signal that such a scheme is redundant (Platteau 2000).

This is crucial information for governments embarking on such interventions, given the lack of evidence for large de Soto-style impacts and growing concerns over the cost-effectiveness of titling schemes (Woodruff 2001; Jacoby and Minten 2007). Better knowledge of the *de facto* state of land tenure and the determinants of demand for formal tenure would also allow policymakers to design better tenure instruments to act as attractive substitutes to the existing system.

In light of this, in the next section I will briefly discuss a land registration and titling scheme which was introduced in urban Tanzania in the early 2000s. I will then use several salient observations from this scheme, concerning the demand for property rights and their expected benefit, to frame the first few chapters of the thesis.

Three observations from an urban land titling scheme

Tanzania is one of the many countries grappling with de Soto's challenge to unearth "trillions of dollars" in its urban slums. For most of its post-independence history the government has been fairly hostile towards fully private, individual land tenure, only changing its stance on property rights during the 1990s. This change culminated with two landmark pieces of legislation known as the 1999 Land Act and the Village Land Act, both of which drastically improved the state's recognition of informal land holdings (Sundet 2005). I will discuss both acts in more detail in Chapters 2-4, but for now I will consider one particular endeavour which emerged from the new land policy: the residential license scheme.

In contrast to a full land title, a residential license (*lesini ya makazi* in Swahili) is a short term, renewable lease on urban land from the government, but one which requires very little in the way of established boundaries to obtain. Rather than undertake costly surveying, the Ministry of Lands opted to use areal photography of urban settlements, combined with hand-drawn boundaries by geographic information service (GIS) professionals, to establish approximate boundaries of each plot (Kironde 2006). In collaboration with the three municipal governments of Dar es Salaam, the Ministry also created a land

Figure 1.1: Clustering of residential license take-up in Dar es Salaam

Note: Land parcels in dark grey are those who have adopted residential licenses.

registry comprising each unplanned land parcel, including information on the owner and some basic parcel characteristics. These municipal land registries were completed between 2004 and 2005, after which the government began offering residents in unplanned areas residential licenses.

The municipal government for Kinondoni, the most populous of the three municipalities making up Dar es Salaam, keeps records of all households counted in the property registration exercise, as well as data on which households went on to purchase a residential license by 2008. Combining the municipal records with GIS maps created from the property register[1] reveals several interesting results:

(1) The adoption of residential licenses is highly clustered geographically

Using the GIS data from the Kindoni municipality, I have mapped the location of land-owning households which went on to purchase a residential license. Figure 1.1 displays a cropped snapshot of the GIS map, where parcels which have purchased a residential license are highlighted with dark-grey. What is immediately apparent is the high degree of spatial clustering: adopting households are typically adjacent to other adopting households, forming small 'islands' of registered properties. This visual relationship can be

[1]Both the municipal records and the GIS data were obtained from the Kinondoni Municipality by a lucky researcher with a flash drive.

Table 1.1: Household's residential license adoption and average neighbour adoption

	# of nearest-neighbours considered			
	3	5	10	15
% of neighbours adopting a RL	0.523***	0.684***	0.760***	0.812***
	(0.00699)	(0.00778)	(0.00819)	(0.00844)
Constant	0.219***	0.144***	0.108***	0.0843***
	(0.00363)	(0.00382)	(0.00392)	(0.00398)
Observations	48672	48702	48702	48702

Each column is the result from a bivariate regression of a dummy for a household's RL choice on the average of the household's n nearest-neighbours, where each column indicate a given choice of n. Robust standard errors. $^{*}p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$

verified empirically: Table 1.1 shows the results from a bivariate regression of each household's decision to purchase a residential license on the average of its neighbours, with each column indicating the number of neighbours being considered. In every column, the average rate of neighbour adoption is strongly a positively correlated with the household's decision to adopt, indicating that there is a strong degree of spatial correlation.[2]

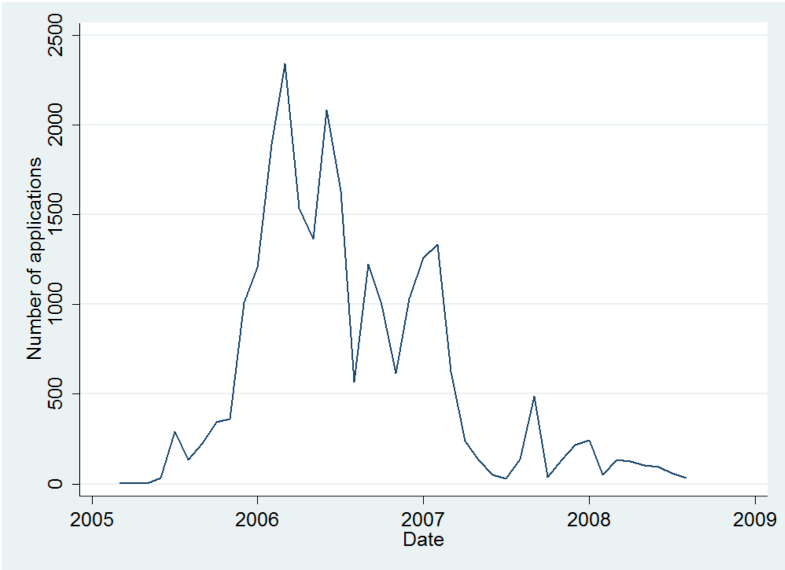
Geography is one plausible explanation, as spatially-proximate households are also likely to share similar characteristics or environments (such as land quality or access to public utilities) which determine their demand for formal land tenure. However, the clustering seen in Figure 1.1 is surprisingly local. Such close grouping in the propensity to obtain a residential license indicate either that these shared characteristics also vary quite strongly over space, or that something else is driving this clustering.

(2) Demand for formal property rights has been fairly muted

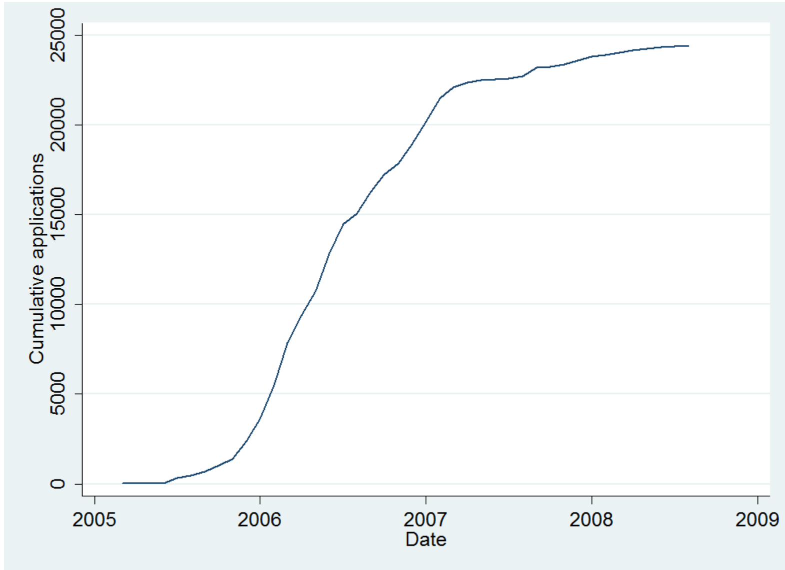
What makes the residential license scheme so interesting is that it represented a real effort by the Tanzanian government to nudge the inhabitants of informal settlements into formal recognition, using fairly well-established methods of land registration and formalisation (Deininger, Augustinus, Enemark, and Munro-Faure 2010). Despite this endeavour, the programme has largely failed in its ultimate goal of approaching full coverage of the city's unplanned settlements. Of Kinondoni's approximately 60,000 informal

[2] There are a number of issues pertaining to both identification and inference using spatial data, which I will discuss in the next few chapters, but for now I wish to focus on the basic spatial correlation.

Figure 1.2: Monthly residential license application rates, Kinondoni Municipality



(a) Applications per month



(b) Cumulative applications

Table 1.2: Gender distribution of property registration

	Property register	RL Database
Male owner listed	31,493 (75%)	13,545 (72%)
Female owner listed	10,329 (25%)	5,147 (28%)
Total	41,822 (100%)	18,792 (100%)

Note: Author's calculations based on land registry maintained by Kinondoni Municipality.

land parcels, just over 20,000 were registered for a residential license between 2005 and 2008. Figure 1.2 shows the application rate per month during this time period, as well as the cumulative number of applications filed. While application rates quickly soared in early 2006, they promptly tapered off by the end of the year. While we are unable to see if there was a recrudescence in residential license registration later on, by most accounts very few, if any, properties were registered after 2008 (Kironde 2011).

Why was demand so low, given that residential licenses are a relatively cheap investment? A single application costs roughly $6, although also makes a landowner liable for property tax and land rent. It is possible that, given their limited scope, residential licenses did not offer a credible alternative to the already-vibrant informal system.

(3) Formal property registration is heavily skewed towards male ownership

Table 1.2 shows the estimated gender breakdown of ownership of land,[3] first from the Kinondoni property registrar, then for the subset of households which elected to purchase a residential license. In both columns, the majority (72-75%) of the listed owners are male, a result in line with estimates by Tanzanian scholars (Kironde 2006). Land titling programmes have historically been skewed heavily towards male ownership, despite the *de jure* recognition of women's right to own land in many countries (Payne, Durand-Lasserve, and Rakodi 2007). Given the status quo, if there are diamonds to be found in formalising property rights, it is possible that the gains will be heavily accrued by men. Even worse, if informal, customary tenure systems afford women some initial protection, this could be stripped away with the introduction of formal titling.

[3]The gender of each landowner has been estimated using the first name of the registered owner. Data which a gender match could not be made has been excluded. Details of this process are given in Chapter 4.

Three papers on property rights, and one on birth order

The next three chapters of my thesis begin to investigate the three observations discussed above. In the second chapter, I and my coauthors investigate whether or not the decision to enter a formal property system is contagious, using experimental data from two unplanned settlements in Dar es Salaam. Rather than attempting to identify the impacts of land titling in a vacuum, this chapter explicitly recognizes that a household's decision to title will have spillover effects on its neighbours, externalities which the literature has largely ignored so far. These effects might explains the heavy clustering seen in Figure 1.1. To identify these spillover effects cleanly, I rely on exogenous variation in land title adoption using an innovative natural field experiment in the slums[4] of Dar es Salaam.

In Chapter 3 I investigate one possible determinant of the muted demand observed above by investigating whether or not households with higher levels of informal tenure security are less likely to adopt a residential license. Using baseline data from the same communities as the field experiment, I compare levels of demand between households living near others of the same ethnic background and those which are ethnically isolated, under the assumption that households living near coethnics already have a strong degree of tenure security and so will be less likely to adopt. To account for endogenous sorting, I rely primarily on changes in the household's group of neighbours which occurred after the household moved into the slum.

Chapter 4 is a short chapter, as it is based on fairly preliminary evidence from the field experiment being used for the results in Chapter 2. This chapter investigates whether or not it is possible to shift land ownership away from the male-dominated regime observed in the residential license database by inducing more women to become formal landowners. This is done by relying on random variation in the incentive to co-title introduced by the same field experiment. Given that the intervention was largely successful in inducing households to include women as formal land owners, the chapter closes with a discussion on the implications for expected bargaining power impacts.

Finally, Chapter 5 pivots away from property rights in East Africa to focus on the issue of intrahousehold allocation of resources in the Philippines. I look at one specific

[4] Throughout this thesis, I will use the term "slum" as shorthand for unplanned, informal settlements.

determinant of child outcomes: a child's birth order. Similar to ethnicity and gender, birth order is a predetermined characteristic which has significant implications for welfare. In this chapter, I use data on a single cohort of children from the island of Cebu in the Philippines to investigate the persistence of birth order effect across time and the potential for nonlinear birth order effects.

In closing, this thesis attempts a substantial contribution to the literature on the development of urban property rights, as well as a small contribution to the growing literature on intrahousehold distribution. Along the way, it also makes minor contributions to the literature of peer effects, the interaction between ethnicity and the provision of public goods, and child outcomes. In Chapter 6, I present concluding remarks, discussing the results of each chapter in turn, and their implications for future research.

Chapter 2

Peer effects and the rise of property rights in the slums of Dar es Salaam

2.1 Introduction

Throughout its history, Dar es Salaam has been shaped by a constant battle between authorities desperate to maintain control over the city's development and the ongoing pressure of informal growth and migration from rural areas. This struggle has roots as far back as the times of British colonial rule, when the colonial authorities tried, but largely failed to introduce a formal land title system to help contain the expansion of a growing Indian population (Brennan 2007). Despite half a century of of large-scale urban planning and 'strict' government control over the allocation of land, Dar es Salaam remains largely informal today, with over 80% of land belonging to to informal, unrecognized settlements (Kombe 2005). It is hardly surprising then that the Tanzanian government, like many others dealing with rapid urban growth, is keen to find innovative ways to sustainably propagate a formal tenure system.

One facet of tenure adoption which is often overlooked is how individuals' decisions to enter the formal system might co-vary with one another. In Chapter 1, I presented evidence that this is already happening in Dar es Salaam, where proximate groups of households seem to be adopting land titles together. In this chapter, I investigate whether the adoption of formal property rights is *contagious*, where the action of one agent adopting a new regime increases the chance that another does the same. In the peer effects literature these are known as *endogenous peer effects* (Manski 1993).

The discovery of endogenous peer effects in property rights adoption is useful for several reasons. First, the existence of adoption spillovers is informative as to whether or not property rights should be considered solely as a private good, or as one with substantial externalities. Second, if the channel through which adoption peer effects operate can be identified, we might learn something more about the expected benefits of titling.[1] Finally, even if the exact mechanisms remain hidden, the presence of positive endogenous peer effects is interesting from a policy perspective, as policies aimed at encouraging take-up would have a subsequent knock-on effects on other households, otherwise known as a social multiplier effect (Glaeser, Scheinkman, and Sacerdote 2003).

Endogenous peer effects are notoriously difficult to identify, as they are subject to both 'reflection' bias (where the direction of causality cannot be determined) and correlated effects (where shared unobservable characteristics drives similar decisions). In this chapter, I overcome these standard identification challenges by using exogenous variation in land title purchases resulting from a unique land titling experiment in the unplanned settlements of Dar es Salaam. The experiment randomly allocated a subset of informal landowners to treatment groups which received massive subsidies to obtain a land title, leaving others excluded. I then combine this variation in the incentive to title with spatial information on the location and treatment status of each household's set of nearest-neighbours, allowing me to identify the impact of each neighbour's adoption decision on the probability that a given household will purchase a land title. This approach is similar to a number of studies which use randomised selection into a programme to identify peer effects (Duflo and Saez 2003; Lalive and Cattaneo 2009; Bobonis and Finan 2009; Oster and Thornton 2009).

My results suggest that there are strong, positive endogenous peer effects in land

[1] For example, if peer effects only operate through spatially-proximate households which are concerned about government expropriation, we might infer that land titling has spatial spillovers in the reduction of expropriation risk.

title adoption. In my main specification, the probability that a household chooses to purchase a land title increases by 8-15% for every neighbour that also chooses to purchase one, an effect equivalent in size to a 25-50% discount on the price of the land title. I also show that these results not only diminish with distance, but they appear to be operating primarily through physical proximity, rather than social proximity, and are not necessarily due to the exchange of information. Furthermore, I show that there is some evidence that households with a higher ex-ante perception of expropriation risk are more responsive to the behaviour of their neighbours, suggesting that there are strategic complementarities in adoption to those most fearful of expropriation, although it is not clear if these complementarities lead to aggregate reductions in expropriation risk. For robustness, I show that these results hold for some basic changes to the structure of the peer group. I then go on to show that these results remain roughly consistent for an identical experiment rolled out in a neighbouring community a year later. Finally, I turn to a database covering roughly 45,000 land parcels in Dar es Salaam, using popular non-experimental methods of identifying peer effects to show that positive effects also exist in this larger setting, albeit with a slightly different type of land title.

In the next section, I discuss the setting of urban Tanzania in more detail, as well as the types of land titles this chapter will be covering. Section 2.3 covers some reasons why peer effects in land titling take-up are likely to exist. Section 2.4 outlines the randomised controlled trial which I will exploit to identify peer effects. Section 2.5 discusses identification and the empirical set up. Section 2.6 covers the main results of the chapter, Section 2.7 covers the results from the second experiment and administrative data, and I conclude with Section 2.8.

2.2 Land tenure in urban Tanzania

In Tanzania, formal access to urban land is controlled exclusively by the government, as all land in the country is owned by the Office of the President (Kironde 1995). Given the rates of growth that Tanzania's cities experienced, the post-independence management and distribution of urban land has generally been haphazard and insufficient (Kombe

2005). Following the 1999 Land Act, the Tanzanian government introduced two new forms of land tenure in urban areas in an attempt to pave the way for more rapid formalisation of existing settlements. The first form of tenure was a temporary, two-year leasehold known as a residential license (RL), which had the benefit of being cheap and easy to implement, but lacked many of the features desired in full titles, such as perpetual security, transferability and collateralisability.

The second form of tenure has been considered to be much closer to a full land title: a certificate of right of occupancy (CRO) lasts 99 years, is transferable and is seen by many as reasonable proof of land ownership by credit providers. Despite the obvious appeal of the CRO, the Tanzanian government has largely failed to encourage urban land owners to purchase them.[2] The lack of progress has been principally due to the large practical and monetary hurdles that urban landowners face, including expensive prerequisites such as cadastral surveying and application fees (Collin, Dercon, Nielson, Sandefur, and Zeitlin 2012).

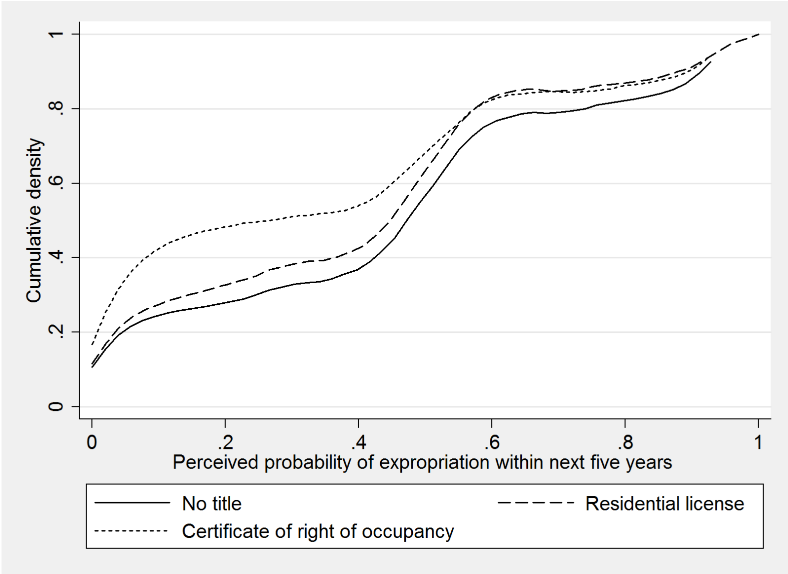
The benefits of CRO ownership

While the Land Act includes relatively straightforward provisions on the legality of using CROs to obtain credit or sell land, the interaction between CRO ownership, expropriation and compensation is less clear. The Land Acquisition Act of 1967 gives the Tanzanian Government broad powers to expropriate land for "any public purpose", even if the owner is in possession of a CRO. This includes government schemes, general public use, sanitary improvements, upgrading or planning, developing airfields or ports and uses related to mining or minerals. Indeed, recent history suggests such expropriation seems most likely to occur from government-driven development initiatives (Hooper and Ortolano 2012). While exact figures on government expropriation are not known, the practice seems frequent enough to elicit alarm in the media: Kironde (2009) found six expropriation-related stories in local newspapers in just one week.

While the Tanzanian government is legally obligated to relocate displaced residents

[2]Records from the Kinondoni Municipality in Dar es Salaam indicate that a little over 2,000 applications from CRO have been made, out of a total population of 60,000 land parcels.

Figure 2.1: Perceived impact of formal land tenure on expropriation risk



Note: Graph shows cumulative density functions of self-reported perceived expropriation probability, conditional on (hypothetical) ownership of different forms of land titles. Data taken from baseline census of landowners in Kigogo Kati and Mburahati Barafu wards in Kinondoni Municipality, Dar es Salaam.

and provide adequate compensation when it acquires land, case studies of recent land conflicts reveal that these efforts are at best mismanaged and at worst completely neglected (Kombe 2010). While a CRO does not legally protect a household from expropriation, it might very well *indirectly* protect a land parcel from government expropriation by raising the value of said land and making the compensation transfer more straightforward. The Land Acquisition Act only provides for compensation in the case where the owner can be identified (Ndezi 2009). Incidents of government expropriation of urban land reported in newspapers and in case studies suggest that informal settlements face the highest risk, so there is reason to believe that, when faced with a choice, governments will usually go for the low-hanging fruit of untitled land.

Even if CRO ownership had no discernable impact on the probability of expropriation, many residents still believe that it does. Figure 2.1 illustrates these beliefs: each line represents an approximated cumulative density function of households' perceived probability of expropriation in the next five years. Residents of two unplanned settlements were asked to condition their predictions on hypothetically having no title at all, having a residential license, or having a CRO.[3] It is clear, at least for a substantial portion of residents, that CROs are *perceived* to be effective at mitigating expropriation risk. As mentioned before, the Land Act also establishes the legal basis for CROs to be used as collateral for loans. Anecdotal discussions with formal lenders in Dar es Salaam suggests that, while CROs are readily accepted as collateral, they do not necessarily offer a substantial benefit over than of a residential license. Still, evidence from the baseline survey used for the experiment described in this chapter suggests that households, on average, also expect that CROs will lead to an increase in both credit supply and land values. While it is clear that households recognise a private benefit to titling, what this fails to reveal is whether or not landowners perceive any *externalities* in adoption which would give rise to peer effects, a possibility I will explore further in Section 2.3.

Before describing the experimental setting where people have been induced to adopt this new form of land title, I will first consider the reasons why we might expect peer effects in CRO adoption to exist in this context.

2.3 Externalities and peer effects in land rights adoption

Most work on formal property rights bundles the benefits and expected impacts of titling into three broad categories, initially summarised by Besley (1995) and later expanded upon in Besley and Ghatak (2010). The first of these is through an (expected) reduction in expropriation risk: formalisation should, in theory, reduce the chance a landowner loses his or her land to to either the state or other individuals. While reducing expropriation should be welfare-enhancing whenever owners derive direct utility from their land, in a context where land is also used as a productive asset, reducing expropriation might have knock-on benefits by encouraging effort or reducing the need for guard labour (Field 2005). In most theoretical contexts, the benefits of reducing expropriation risk are strictly private and positive.

Tenure formalisation is also expected to make it easier for landowners to leverage the value of their land to access credit. Many banks or other lending institutions use collateral

[3] The order of the conditional questions were randomised to avoid priming the respondents.

requirements to relieve adverse selection and moral hazard problems in the credit market. As it should be easier to foreclose on collateralised land when it is covered by a formal document, titling should, in theory, lead to an expansion of credit. Finally, formalisation is expected to increase the transferability of land, allowing landowners to take advantage of rising land prices and for ownership to shift to those who can use it most productively.

With the exception of general equilibrium credit market and land price impacts, which are often ambiguous (Besley and Ghatak 2010), many of these benefits are often modeled as private returns, with the act of one landowner having obtained formal land tenure having no impact on other landowners. There are a number of reasons why there might be immediate, direct spillovers from the decision to buy a land title, both of which have implications for the existence of peer effects. In this section I will consider the most plausible ones, given the context, and then discuss how, in this chapter, I will attempt to discern between them.

Complementarity or substitutability in the returns to land title adoption

One particular area which remains understudied is whether or not there are spillovers in the *returns* to adopting formal property rights. Individual formalisation efforts, such as land titling, might not only result in a private benefit, but might also impact the returns to titling for other individuals. Consider a very simple model of decision-making for household i who must choose between spending an exogenous source of income y_i on consumption c_i or on a (continuous, for simplicity) amount of land tenure T_i.

Imagine the utility function looks like this:

$$U_i = c_i + F_x(T_i, \overline{T_j})H \tag{2.1}$$

Where H is the household's utility from owning land and $F_x(.)$ is a parameter which scales that utility, itself a function of the household's tenure status T_i and (possibly) the average tenure status of the household's neighbours $\overline{T_j}$. Parameter $F_x(.)$ could represent, for instance, the probability the land is protected from future expropriation, the potential sale price or the collateral value of the land. Assuming a price for each unit of tenure p_T, first order conditions leave us with the following equality:

$$\frac{\partial F(.)}{\partial T_i} H = p_T \tag{2.2}$$

The household buys formal land tenure up until the point when the expected gain in utility from land is equal to the price. Now consider the cross-partial derivative: $\frac{\partial^2 F_x(.)}{\partial T_i \partial T_j}$. In the case where the cross-partial derivative is zero, land titling has no externalities across tenure effectiveness; each household makes its choice independently, as the choice of its neighbours has no effect on the marginal returns to titling.

However, we might expect that the returns to titling would be increasing in the number of neighbours taking the same action. This is the classic case of *strategic complementarity*, when the private returns to an action are greater when other agents also take it (Schelling 1978; Bulow, Geanakoplos, and Klemperer 1985). In the above example, this would be the case when the cross-partial derivative is greater than zero, with household i's returns to titling *increasing* as more neighbours adopt.

Where might we see strategic complementarities in practice? For one, there might be a snowballing effect in the reduction in expropriation risk, with the government taking formal tenure more seriously as more people adopt it, possibly due to the rising implicit costs of paying out compensation. However, even in the presence of strategic complementarities, expropriation spillovers might not be entirely positive. If, for example, a government decides to expropriate land which has the lowest level of formal tenure, the act of land titling might just shift expropriation risk from one set of households to another. In this instance, households will be induced to title when their neighbours do the same, not because the decision leads to a net gain in welfare, but because they must do so to prevent a rise in their risk of expropriation. This implies that titling creates a 'race to the bottom,' where all households title in order to improve their security of tenure, but are no better off at the end of the titling scheme. This result is akin to De Meza and Gould's (1992) burglar alarm example: while there is a private benefit for a given household installing a burglar alarm, it increases the probability of neighbouring houses being burgled and hence a no-alarm equilibrium is preferable to an all-alarm one.

Complementarities might also exist in the other standard benefits of land titling. For example, banks may be more likely to accept land titles as form of collateral if they are widely used and accepted in a community (Fort, Ruben, and Escobal 2006) or the impacts of titling on land prices might increase as more neighbours adopt.[4]

Of course, titling decisions could also be substitutes. If the cross-partial $\frac{\partial^2 F_x(.)}{\partial T_i \partial T_j}$ is negative, then the marginal utility from titling *decreases* as more neighbours take up, making it more likely that household i will opt out.[5] If the main benefits of titling are through reducing expropriation risk, this would reflect a context where low levels of titling are enough to deter a government from clearing an area, and so subsequent titling is less effective. Similarly, some have argued that the credit-supply effects of large scale titling will be smaller than individual titling: if lenders consider titling to be a signal of borrower quality, rather than as a collateralisable asset, then large-scale titling would imply a lower signal-to-noise ratio (Dower and Potamites 2012).

Strategic complementarity (substitutability) in the returns to titling would imply a positive (negative) endogenous peer effect, as the effect of neighbour take-up increases (decreases) the marginal benefit to titling for a given household. For most of the impacts discussed here, we would also expect these spillovers to be inherently spatial: both expropriation risk and land values are typically highly correlated across space (as might be collateral values, as lenders might be more confident in extending loans to areas they are already familiar with). Later on on this chapter, I will take advantage of the spatial nature of my data to discern whether or not the observed endogenous peer effect varies with distance.

Learning and rule-of-thumb behaviour

Peer effects might also arise from learning behaviour: based on their peers' experience, individuals update their beliefs on the efficacy of a product. This 'social learning'

[4]Note that both of these channels might also be subject to net negative impacts. If banks switch to a regime where formal titles are the only legitimate form of collateral, non-adopting households might be rationed out of the market (Van Tassel (2004) shows a similar result might happen even if all households are given title) Similarly, if titles become the *de facto* means of transferring property, households relying on informal channels may feel the need to adopt if they are to sell in the future.

[5]This opens the door for standard public goods/collective action problems, as everyone has a private incentive to disinvest if they know their neighbour is investing.

behaviour has already been revealed in the decision of farmers to adopt new farming techniques or new types of crops (Bandiera and Rasul 2006; Conley and Udry 2010; Zeitlin 2012). This could equally apply to landowners in urban areas who observe their neighbours obtaining land titles and possibly being secure from expropriation, gaining access to credit or selling at a high price. Yet, in the context of this study, the benefits of holding a land title would be impossible to measure: as I will discuss in the next section, land titles have yet to be issued for landowners involved in the field experiment. This prevents the sort of wait-and-see learning observed in previous studies.

However, if landowners believe that the adoption decisions of their peers reveal their *knowledge* about the benefits of land titling, high rates of peer adoption may act as a signal for high returns. Recent evidence suggest that peers' adoption decisions transmit important information, irrespective of actual adoption outcomes (Bursztyn et al. 2012). In this circumstance, any observed endogenous peer effects would be unambiguously positive, as take-up conveys a signal of high-returns to titling.

It is normally difficult to disentangle peer effects created by strategic complementarities from signaling/learning behaviour. However, we might expect peer effects determined by the latter to transmit through traditional social networks, as households observe the behaviour of not only their neighbours, but also their friends and acquaintances. Later in this chapter, after establishing that that endogenous peer effects in land title take-up exist between spatially-proximate households, I will then take advantage of some basic social network data to investigate whether or not endogenous peer effects also exist across this alternate network structure, which would suggest that effects other than complementarity/substitutability spillovers are at play.

Other channels

Another concern might be strategic expropriation on the part of those obtaining a land title, with early-movers grabbing a portion of their neighbour's land by making an early claim. While this might be a concern in other settings, it is unlikely to be a factor here, as contiguous neighbours must sign off on the CRO application forms affirming the boundaries of the plot. Furthermore, these sorts of actions would still fall under the 'complementarity' channel: if adopting a CRO protects me from my neighbour's attempt to grab land, my neighbour's action increases the marginal gain from adopting that title.

Finally, there might be information-transfer peer effects, where households learn about the benefits of CRO adoption and share this information, then make entirely independent decisions to title. I will discuss this channel and my attempt to rule it out more in the next section.

2.4 Experimental design and data collection

As I described in the Section 2.2, most households in Dar es Salaam face formidable barriers to the adoption of formal land titles. In this section, I will describe an experimentally-provided land titling programme designed to overcome these barriers. The random variation in CRO adoption induced by the experiment will then be used to identify the impact of a neighbour's adoption on a given household's propensity to adopt.

2.4.1 An experimental land titling programme

The setting is Kinondoni, which is the largest of the three municipalities which make up Dar es Salaam and houses approximately 50% of the city's population. The land titling programme was introduced in two adjacent neighbourhoods (known as sub-wards or *mitaa*), first in Mburahati Barafu then a year later in Kigogo Kati. Barafu will be the main focus of this chapter, due to the completeness and robustness of its data, although I will be using the subsequent replication in Kati as a robustness check in Section 2.7.1.

Both neighbourhoods are located approximately five kilometers from the city centre. While there are a number of pre-planned parcels at the core of each settlement, each mtaa is primarily composed of unplanned, informal settlements. Table 2.1 displays some basic administrative data from both neighbourhoods alongside Kinondoni as a whole. Typical of most informal settlements, Barafu is a high density area with relatively low reported land values and a lack of access to public services and infrastructure. Informality is the norm here, with very few households holding formal tenure: estimates from a baseline census of Barafu put the total number of CRO owners at around 10 households, less than

Table 2.1: Summary Statistics on Parcel Characteristics

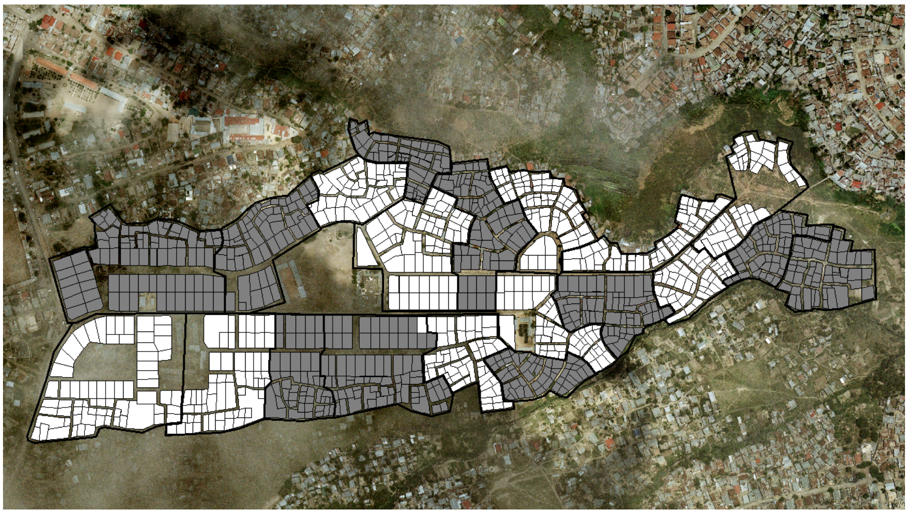
	Kinondoni Municipality	Kigogo Kati	Mburahati Barafu
Formal employment	49.9%	44.6%	44.3%
Size and Value of Property			
Area in square meters	439	264	247
Property value in '000 TSh.	12,562	9,939	8,910
Land rent in TSh.	3,679	2,125	1,907
Accessibility to the Property			
No access	1.3%	1.1%	1.1%
Foot path	55.2%	71.3%	82.0%
Feeder road	36.4%	19.8%	16.2%
Main road	5.5%	6.6%	0.6%
Highway	1.6%	1.1%	0.0%
Access to Public Utilities			
Piped water (incl. public)	22.7%	22.0%	5.6%
Electricity connection	46.1%	38.6%	35.1%
Waste removal services			
Burn/buried on plot	35.4%	25.4%	55.7%
Gutter/river/street	20.0%	49.6%	35.4%
Collected by priv. company	40.8%	24.4%	8.4%
Collected by municipality	3.8%	0.7%	0.5%
Number of properties	65,535	1,474	990

Source: Author's calculations based on the land registry maintained by Kinondoni Municipality.

1% of the community, and administrative data suggests less than 40% of households have ever purchased a residential license.

In October, 2010, the University of Oxford and the World Bank began implementing a land titling programme aimed at increasing the rate of CRO adoption in Barafu. This was done in partnership with the Woman's Advancement Trust (WAT), a Tanzanian NGO which specialises in large-scale titling programmes. The intervention proceeded as follows: prior to launching the land titling programme, all land parcels in Barafu were identified via a household listing of the community and a recent map drawn up by a town-planning firm. This map was used to divide the community up into twenty 'blocks' of roughly 40-50 parcels each. Using a set of basic characteristics taken from the listing to establish balance, ten blocks were randomly allocated to a treatment group and ten to a control group. Balance results for the overall treatment are available in Table 2.10 in Appendix 2.A.1. Figure 2.2 shows the map of Barafu with treatment and control blocks outlined. Parcels in treatment blocks (and their owners) were subject to several interventions:

Figure 2.2: Treatment and control blocks in Mburahati Barafu



Note: Treatment blocks are shaded grey.

1. *All* parcels in treatment blocks were subject to a cadastral survey (demarcation of boundaries using cement beacons), one of the prerequisites for applying for a CRO.

2. Parcels in treatment blocks were invited to meetings to discuss involvement in the land titling programme and the benefits of CRO ownership, run by WAT.

3. During these meetings and subsequent follow-up visits, treatment parcels were invited to pay 100,000 TSh to WAT (approximately $62, the average cost of the cadastral surveying plus application fees) over a period of about five months in return for a CRO. In exchange for this, WAT would manage the application process and any related fees.

4. Within treatment blocks, parcels were randomly allocated voucher discounts through a public lottery. Two types of vouchers were allocated: general vouchers, which were redeemable without condition, and conditional vouchers, which required that a female member of the household be included as an owner on the final documents. A parcel could receive both voucher types, just one type, or none at all, and vouchers could take on values of 20, 40, 60 or 80 thousand shillings.[6]

[6]Complete details of the voucher allocation process are discussed in Chapter 4.

Following this, households in treatment blocks were free to sign up up and begin repayment. Through an agreement with the municipal government, treated households could not obtain a CRO through conventional means, only through the NGO. Households in control blocks were free to obtain CROs through the municipality, at the regular cost, although a subsequent review of municipal records revealed that none have done so to date. At the time of writing, the project is still underway, with no land titles having yet been issued, but with household decisions and payments having been completed.

2.4.2 Data sources

In this chapter I use three primary sources of data from Barafu. The first was collected prior to the randomised intervention: in the summer of 2010, roughly six months before the start of the land titling programme, the University of Oxford conducted a complete census of all known parcels in Barafu, using records obtained from the Kinondoni Municipality. For each parcel, an owning household was identified and interviewed, resulting in a rich data set of owner and parcel characteristics. However, as this data was collected earlier and used a different sample frame than the administrative project data, there are a number of missing observations, mainly due to parcels which were missed during the baseline census or those that were sold to a new owner in the interim. Baseline data is available for roughly 92% of unplanned parcels in treatment blocks, but only 72% of control blocks have linked baseline data, due to a lack of project information for these households. I will use this data both for testing balance and for controls in my main specification.

The second is detailed parcel-level data taken from project records, including meeting attendance, sign-up and repayment information. As households in control blocks were excluded from participating in the project, this data only contains information on treated parcels. However, data obtained from the Kinondoni municipality reveals that no parcels in the control blacks purchased a CRO during the time frame of the project. Finally, the third source of data I use is detailed geographic information service (GIS) encoded data on the location, shape and size of each parcel in both treatment and control areas.[7] This

[7]This data was taken from a ‘town plan’ of Barafu, the final planning document drawn up for a

will allow me to calculate 'nearest-neighbour' peer groups for every parcel and compare the probability of a household choosing to purchase a CRO with the average adoption rate of that household's nearest-neighbour set.

Throughout the remainder of this chapter, I will primarily be using *unplanned* parcels in the analysis, as this group was the original target of the research. While there is complete project data and limited baseline data on planned parcels (those who were in a pre-planned area or had already obtained a cadastral survey prior to the intervention), I will only be using them as a robustness check for the main result.

2.5 Identification of endogenous peer effects

Consider a basic linear probability model for household i's decision to adopt a land title:

$$T_i = \alpha + \rho \overline{T}_{g(-i)} + x_i \beta + \overline{x}_{g(-i)} \delta + u_i + \epsilon_g + \varepsilon_i \tag{2.3}$$

Where T_i is the household's choice to adopt a land title, $\overline{T}_{g(-i)}$ is the average choice of the households group of neighbours g (excluding i), x_i is a vector of household characteristics, $\overline{x}_{g(-i)}$ is the same set averaged over the group, u_i is a household-specific effect and ϵ_g is a vector of group-specific characteristics. Using Manski's (1993) terminology, ρ is known as the *endogenous* effect, the impact of i's neighbours' choices on i's choice. The parameter δ represents a vector of effects stemming from i's neighbours' characteristics, known as *exogenous* or *contextual* effects. Finally, ϵ_g contains unobserved within-group *correlated* effects.

There are two primary challenges to the identification of ρ, the parameter of interest. The first is a result of Manski's 'reflection problem', where the direction of causality is difficult to discern. At first glance, we are unable to identify whether ρ captures aggregate effects of i's neighbours' adoption on i or vice versa. In the extreme case where peer groups are perfectly transitive,[8] it is difficult to separately identify endogenous peer effects ρ and the set of contextual effects δ_g.[9] However, when peer or neighbour groups are partially

community before CROs can be provided

[8]Transitivity implies that if i and j are peers and j and k are peers, then i and k must also be peers.

[9]Brock and Durlauf (2001) exploit nonlinearities in discrete choice models to identify linear-in-means

overlapping (i.e. when the neighbours of i's neighbours can reasonably be excluded from i's neighbour set) identification is made possible by exploiting variation in characteristics of these excluded neighbours (Bramoullé et al. 2009; De Giorgi et al. 2010), a popular method I will apply to a larger, non-experimental data set in Section 2.7.2.

The second concern is over conflating endogenous peer effects with correlated effects. The latter can arise when peer groups or neighbours are affected by common background characteristics or shocks which also predict adoption. For example, if land title adoption depends on unobserved (to the researcher) land quality, then adoption rates will be correlated across neighbours even in the absence of endogenous effects. Similarly, if the endogenous sorting of households into peer groups or neighbour sets is marked by *homophily*, then correlated adoption decisions might solely be the result of correlated individual characteristics, such as wealth or risk aversion.

In this chapter, I use the random variation in the price and accessability of land title purchase to identify exogenous changes in $T_{g(-i)}$, allowing me to estimate (2.3) using two-stage least squares (2SLS) with reduced concerns for correlated effects and reflection. I do this using the percentage of household i's neighbours who were included in treatment blocks as well as their average voucher values[10] as instruments for the average adoption of the neighbour set. Since households in control blocks were effectively excluded from purchasing CROs,[11] my sample will only cover households in treated blocks (although I will consider neighbours from both treatment and control blocks). I will discuss the suitability of these instruments and possible reasons why identification might still fail in the next subsection.

While many studies have used random variation in *group* assignment to estimate peer effects (Sacerdote 2001; Guryan, Kroft, and Notowidigdo 2009), my approach in this chapter is more similar to those which use random variation in *programme* assignment

models, yet identifying assumptions are heavily dependent on functional form, and do not allow for correlated effects.

[10] Averaged over included-neighbours. For precision I use both regular and conditional vouchers separately.

[11] While it would be tempting to include control households in the main sample and code their take-up as zero (as Municipal records revealed that no control households purchased CROs), it is worth noting that control households are effectively excluded (almost no households purchased CROs prior to the intervention, due to the extremely high costs and barriers to entry. Hence, the endogenous peer effects estimated in this chapter should be thought of as being conditional on switching from a regime where all households are excluded to one in which land titles become affordable.

as an instrument for peer-level adoption. For example, both Lalive and Cattaneo (2009) and Bobonis and Finan (2009) use the random assignment of a conditional cash subsidy in PROGRESA villages to instrument for the school enrolment of a child's peer group. Similarly, Oster and Thornton (2009) use the random assignment of menstrual pads to Nepalese school girls to study the impact of group-level treatment on individual utilization of the pads. Both Godlonton and Thornton (2012) and Ngatia (2011) use randomized price incentives to get tested for HIV/AIDS in Malawi to instrument for peer group testing. In each of these studies, social interactions are treated as a specific type of treatment spillover: an individual's peer group is randomly shocked and the resulting change in behaviour affects the individual's adoption choice. This method was first laid out by Robert Moffitt as the *partial-population* approach (Moffitt et al. 2001).

2.5.1 Interpreting estimates

Econometric issues

There are a couple of caveats to the interpretation of ρ using the partial-population approach. First, while properly instrumenting $T_{g(-i)}$ solves the reflection problem and bypasses any group or individual-level unobservables, the resulting estimate of ρ is the endogenous peer effect, conditional on groups already having formed endogenously. These 'true' peer effects might be stronger or weaker for households which have chosen to live together as opposed to those randomly sorted into the same neighbourhood. For instance, households from the same religious background might be more likely to associate and share information about adoption decisions. In this instance, we might expect the estimate of ρ, post-endogenous sorting, to be higher than the estimate under random sorting.

Which estimate do we care about? While the 'randomly-assigned' endogenous peer effect might be more appealing to those concerned with pure social interactions, in reality the policymaker has little control over the formation of these peer groups, in which case the 'post-sorting' endogenous peer effect is clearly the preferred parameter. In the context of urban formalisation, most policymakers are burdened with the significant task of getting large, informal settlements to take up formal property rights. As these settlements have not formed randomly, the post-sorting peer effect gives us an idea as to whether significant policy multipliers are present for property rights interventions.

Another issue follows directly from using 2SLS with an exogenous treatment instrument to identify peer effects. Under the assumption of heterogenous effects, instrumental variables regressions only allow the researcher to identify the local average treatment effect (LATE) (Imbens and Angrist 1994). The implications of this for the estimation of peer effects are nonnegligible. For example, when using the block-level treatment as an instrument for neighbour adoption, the effect identified ρ is only defined for compliers, households whose *neighbours* were induced to adopt from the treatment, but otherwise would have not done so. As mentioned in the previous section, there are no always-takers, so estimates of ρ using project treatment of an instrument will only be leaving out never-takers, those that do not respond to the treatment. If we have reason to believe that peer effects are heterogenous, then LATE estimates of ρ might deviate substantially from the average treatment effect estimate. The peer effects literature has largely been silent on this issue, with some exceptions.[12]

Finally, it should be emphasised that while the randomised control trial described above has generated geographic variation in the take-up of CROs, the RCT itself was not designed for the purpose of of studying peer effects. Thus most[13] of the identifying variation in take-up will be generated by the large-scale block-level variation in treatment. While this is not as precise as a parcel-level treatment, identification will be possible as long as treated neighbours are not systematically different from untreated neighbours and households *with* treated neighbours are not systematically different than those without. To allay any concerns, I will show in Section 2.5.3 that when compared using baseline data, treated and untreated parcels are, on the whole, very similar.

Conceptual issues

While the data from the experiment only comprises actual adoption decisions, the

[12]To date, only Dahl, Løken, and Mogstad (2012) and Ngatia (2011) have explicitly acknowledged that peer effects estimated using 2SLS are subject to a LATE interpretation. Ngatia (2011) explicitly models these heterogenous effects and estimates their effects by exploiting multiple instruments for adoption.

[13]Some of the variation will still be driven by variation in the voucher allocation received by treated neighbours.

empirical adoption equation (2.3) can have the usual latent variable interpretation: each household makes a decision to purchase if the expected gain from obtaining a title exceeds the cost. A positive (negative) estimate of ρ indicates that the household's expected gain is increasing (decreasing) in the number of neighbours who take up. However, while identifying ρ would be sufficient to understand the sign of the effect of neighbour decisions on household expectations, it is not enough to discern whether or not these expectations are affected by strategic complementarity/substitutability,[14] or by signaling of the quality of the take-up decisions, as described in Section 2.3.

One option would be to consider different constructions of the relevant neighbour group g_{-i}. When peer effects are being determined by direct spatial spillovers in the return to titling, we would not only expect estimates of ρ to be positive or negative, but to be stronger for spatially-proximate households. Furthmore, if endogenous peer effects were being generated by a positive signal of the private returns to titling, endogenous peer effects should exist even when considering non-proximate neighbours who are still part of the household's social network. In sub-section 2.6.1 I will consider different constructions of g_{-i} in an attempt to discern between direct spillovers and signalling.

2.5.2 Challenges to identification

Even though the instruments I use in this chapter are randomly drawn, there are still a number of ways the above identification strategy might be undermined. For instance, despite the randomisation, a bad draw in assignment of treatment status or voucher values might have resulted in spurious correlation with relevant unobservable characteristics. Later, I will show that not only both treatment and voucher assignment are well-balanced across a range of observable characteristics obtained from the baseline census, but that the main results presented in Section 2.6, are unaffected by the inclusion of these characteristics. While balance and conditioning on observables does not guarantee identification (Bruhn and McKenzie 2009), randomisation is as close as we're ever likely to get, as in expectation the instruments should be uncorrelated with the error term in the main

[14] Although there is little scope for negative peer effects under the signaling hypothesis, as a neighbour taking up a land title should not reduce the probability it is a good decision for the household to the same.

equation.

A more pertinent problem is the exclusion restriction. In order for the estimate of ρ to be interpreted solely as an endogenous peer effect, the instruments (being in a treatment block and the random voucher draw) must only affect a household's adoption of a land title through the adoption of its neighbours. There are a few reasons why this might not be the case:

One valid concern is that direct-adoption peer effects might be confused with information exchange. Prior to the intervention, most residents knew very little about CROs. Since households in treated blocks are invited to meetings in which they are given extensive information on the benefits of these titles, it is possible that attending households passed this information on to their non-attending neighbours. Thus the observed peer effect ρ might include the impact of this information transfer. To account for information in my main specification, I will use data on household and neighbour meeting attendance to proxy for knowledge of CROs.

Another potential problem is related to a second change in neighbour characteristics driven by the treatment. Recall that all parcels in treatment blocks are subject to a cadastral survey, even if the owners do not go on to purchase a land title. The act of surveying a neighbour's plot could have an independent effect on a household's decision to purchase a CRO if, for instance, being in a heavily-surveyed area affects the perceived value of a title. Recent evidence suggests that land demarcation has important implications for the function and growth of land markets (Libecap and Lueck 2011), so it is possible that a shift from the previous regime[15] to tightly-regulated cadastral surveying could have substantial impacts independent of land title adoption.

To deal with this, I first turn to data from the baseline census, which suggests that a household's perceived expropriation risk is unaffected by proximity to previously-surveyed parcels (these results are discussed in detail in Appendix 2.A.3). Secondly, I also present results which include previous-surveyed parcels as neighbours and show that endogenous peer effects are of a similar magnitude for this group. Finally, the timing of the intervention suggests that adoption decisions might have been independent of surveying: while

[15]Prior to the introduction of the town plan, parcels were delineated with hand-drawn maps produced using aerial photography.

treatment and control blocks were decided at the beginning, actual cadastral surveying did not begin until several months following the initial sign-up period, and took over a year to complete, so the final surveying status of treated-neighbours would have been unconfirmed for most households.

Another assumption behind the exclusion restriction is that proximate neighbours have independent budget constraints. This would be undermined if two neighbours act as a single household or take part in risk-sharing groups.[16] However, while spontaneous risk-sharing groups have been observed in randomised controlled trials in the past,[17] the chances of such an arrangement existing in this context are slim, given that the households were presented with non-transferable vouchers which were tied to individual parcels.

The possibility of other types of social interactions highlights another issue with the structure of the intervention: the public nature of the voucher draws. If households felt under pressure to use the vouchers due to their relatively high-profile distribution or conversely were less likely to use them due to fairness concerns, there might be direct voucher spillovers on take-up, apart from the indirect effect implied by the 2SLS setup. However, auxiliary work (not shown here) suggests that while household demand affected by the levels of their own voucher values, it appears to be independent of its relative position in the peer group draw. Furthermore, households which attended meetings appear to be no more or less responsive to voucher values, indicating that peer pressure to either use or ignore value values might not be at play. However, as meeting attendance is endogenous, it is still possible that these effects are at play.

The public nature of the programme might have other unintended effects on take-up: as a part of the application process, households which opt to purchase a CRO must have all contiguous neighbours sign to certify that the applicant is the rightful owner of the parcel. If boundaries are in dispute, the decision to purchase a CRO might trigger a conflict with nearby neighbours, indicating that some households might avoid adopting for fear of their neighbours. However, as all households in treatment blocks are subject to

[16] Lalive and Cattaneo (2009) discuss this as a potential threat to identification, where sharing of PROGRESA transfers might lead to a spurious social interaction result.

[17] Blattman (2011) discusses difficulties with lottery recipients exchanging winnings. Similarly Angelucci and De Giorgi (2009) shows that ineligible households are affected by cash transfers to treatment households.

a cadastral survey, whether or not they go on to purchase a CRO, all parcel boundaries must be worked out prior to the titling process. This should, in effect, decouple the decision to title from concerns over starting a boundary conflict.[18] This should also apply to treated households with neighbours outside of treated blocks, as continuous, untreated neighbours will still share a surveyed boundary. However, households in untreated blocks with no treated neighbours do risk starting boundary conflicts if their make the private decision to have their plot surveyed, which may be one explanation for why the untreated take-up rate is so low.

Finally, the exclusion restriction might be undermined if households decide not to participate in the programme because of concerns for fairness (for their neighbours not being included) or if high/low voucher allocations to neighbours elicit feelings of envy or unfairness which stop them from adopting. However, anecdotally there is not much evidence that these sort of feelings are at play on the ground.

2.5.3 Empirical setup

Reconsider the empirical model presented in equation (2.3), which is presented as a linear probability model (LPM):

$$T_i = \alpha + \rho \overline{T}_{g(-i)} + x_i\beta + \overline{x}_{g(-i)}\delta + u_i + \epsilon_g + \varepsilon_i$$

While it is possible to estimate this using a nonlinear specification, such as a probit or logit, the LPM makes interpretation of the results relatively straightforward. The chief concern over the use of a LPM is over out-of-sample predictions and the potential bias which results from its use. In Table 2.15 in Appendix 2.A.1, I show that the percentage of out-of-sample predictions is extremely low, which suggests that there is not much scope for bias in the LPM (Horrace and Oaxaca 2006).

A dummy variable equal to one if a household has fully paid for its CRO will be used as my main measure of title adoption T_i.[19] In my main specification, for household/parcel characteristics x_i, I will include the general and conditional voucher values that the house-

[18] There appears to be no correlation between experience with border conflicts, as measured in the baseline survey, and subsequent purchase of CROs by households.

[19] Results are also robust to using household sign-up as a measure of adoption instead of full purchase.

hold received and a control for whether or not that household attended the block-level meeting. In addition, I will include a series of baseline controls, including the natural log of the parcel's area, the year the parcel was obtained, the household's monthly income, total value of all assets (TSh), household size, average schooling and dummy variables for whether the parcel is rented out, the owner is resident on the parcel, and there has been recent investment in the parcel. Each of these controls is also averaged across the household's neighbour set and included in $\overline{x}_{g(-i)}$, with the exception of the neighbour's voucher values, which are used as instruments. I have also included a control for whether or not the household has neighbours outside of the treatment block, so as not to conflate differences in neighbour treatment with the household's relative location within the block.

Using GIS data to calculate distances between parcel borders, I construct peer groups using the n closest neighbours to household i. This approach allows for results which are intuitive and easy to understand, as each house has equal-sized peer groups. For robustness, I will also present results using fixed-distance neighbour sets (which include all neighbours within a certain distance d), but the differences are minor. As it offers a reasonable trade-off between proximity and the power of the instruments,[20] my main results will use the five nearest-neighbours, but extensions on the size of the neighbour sets are presented in Section 2.6.1.

Summary statistics for the main controls, as well as their balance across voucher allocations and the percentage of five nearest-neighbours treated, are shown in Table 2.2. Parcels which faced a high price were slightly less likely to be electrified and had slightly higher levels of schooling, but neither of these differences are substantial. Households with a high percentage of treated neighbours were more likely to attend meetings and were less likely to have purchased a residential license. Apart from these differences, the sample appears to be fairly well-balanced.

Finally, I will be using both average voucher values across the neighbour set and the percentage of treated neighbours as instruments for $\overline{T}_{g(-i)}$. While the results are robust to including these as separate instruments, the estimates are most precise when they are aggregated into a single instrument. This instrument is defined as the 'total' price of a

[20] The larger the neighbour set the greater the number of households which fall outside i's block and therefore have the potential to be treated.

Table 2.2: Summary statistics and balance (voucher distribution and treated neighbours)

	Mean/SD	Own price	% neighbours treated	Mean neighbour price
	(1)	(2)	(3)	(4)
Attended meeting	0.61	0.002	0.515	-.0008
	(0.488)	(0.0009)**	(0.165)**	(0.0004)**
Year parcel acquired	1992.126	0.014	6.454	-.016
	(12.307)	(0.023)	(4.228)	(0.009)*
Parcel rented out	0.4	0.001	-.040	0.00007
	(0.512)	(0.001)	(0.176)	(0.0004)
Owner resides on parcel	0.827	-.00007	0.05	-.00008
	(0.395)	(0.0007)	(0.136)	(0.0003)
Applied for CRO in past	0.014	-.00007	-.061	0.0001
	(0.124)	(0.0002)	(0.042)	(0.00009)
Applied for RL in past	0.386	0.0001	-.288	0.0006
	(0.508)	(0.001)	(0.175)*	(0.0004)
Parcel was inherited	0.107	0.0008	0.113	-.0002
	(0.322)	(0.0006)	(0.111)	(0.0002)
Parcel has electricity	0.408	-.002	0.033	-.0003
	(0.513)	(0.001)**	(0.177)	(0.0004)
# buildings on parcel	1.332	0.0007	-.065	0.0002
	(0.56)	(0.001)	(0.193)	(0.0004)
Invested in parcel	0.175	-.0006	0.06	-.00003
	(0.397)	(0.0007)	(0.137)	(0.0003)
Monthly income	356.346	0.477	-175.492	0.328
	(464.245)	(0.876)	(159.710)	(0.349)
Total assets (tsh 000')	4140.882	15.221	-1836.007	4.724
	(6848.912)	(12.908)	(2357.852)	(5.151)
Average schooling	12.263	0.009	1.479	-.002
	(2.783)	(0.005)*	(0.956)	(0.002)
Household size	4.716	-.007	-.669	0.0003
	(2.508)	(0.005)	(0.863)	(0.002)
Ln(area m^2)	5.096	0.001	0.146	-.0002
	(0.529)	(0.001)	(0.182)	(0.0004)
Obs	459	459	459	459

Column (1) displays the mean and standard deviation for each variable. Columns (2)-(4) display the mean and standard error of β_2 from the linear regression of each variable $var = \beta_1 + \beta_2 * Z$, where Z is overall price the household faced (2), the percentage of five-nearest neighbours who were in treatment blocks (3) and the average price faced by the household's five-nearest neighbours (setting p = 500,000 TSh for neighbours in control blocks)(4). Price measured in ('000 TSh).
$^{*}(p < 0.10),^{**}(p < 0.05),^{***}(p < 0.01)$

CRO per household, which is set to TSh 500,000 for untreated neighbours (which is in line with previous estimates)[21] and set to the actual project price, net of vouchers, for treated neighbours.

To account for spatial dependence of observations, all standard errors are calculated using Conley's (1999) method, where the estimated covariance matrix is adjusted to allow for arbitrary spatial correlation between observations. The degree of correlation is allowed to decrease linearly with distance and is set at zero beyond a specified cutoff. For all nearest-neighbour specifications, cutoff values are set at the average distance of the fifth neighbour across observations. For distance-band specifications, cutoff values are set equal to the distance-band. In general, the results are not qualitatively different from standard heteroskedastic-robust estimates.

2.6 Main results

Table 2.3 shows the results from the estimation of equation (2.3) using the five nearest-neighbours as the relevant peer group. The first three columns display results from an OLS estimation of the probability that household i adopts a land title on the number of neighbours in the neighbour set also adopting.[22] In column (1), the controls included are household i's allocated vouchers, whether or not someone from the household attended the information/voucher distribution meeting held for the treatment block, the percentage of i's neighbours who attended a meeting and the percentage of neighbours who are in a different treatment/control block. Column (2) restricts the sample to households with baseline data and the nearest-neighbour set to neighbours with baseline data, but does not include baseline controls. These are introduced in column (3), so as not to conflate sample-selection differences with the changes induced by including controls. Also included are average values for these controls for household i's neighbour set. Columns (4), (5), and (6) repeat the same pattern, but using 2SLS to estimate equation (2.3), using the average 'total' price households in the neighbour set faced as an instrument.

[21] Average estimates put this at about $500-1000 per parcel.

[22] This estimation is equivalent to using the average adoption rate for the neighbour set, multiplied by the size of the neighbour set, which is a constant. For results using distance bands instead of nearest-neighbour sets, I multiply by the average neighbour set size.

Table 2.3: Barafu - impact of neighbour's CRO take up on own take up - 5 closest neighbours

	OLS			2SLS		
	(1) Basic	(2) Restricted	(3) Restricted + Controls	(4) Basic	(5) Restricted	(6) Restricted + Controls
# of neighbours adopting	0.0773***	0.0836***	0.0835***	0.147***	0.137***	0.148***
	(0.0170)	(0.0173)	(0.0182)	(0.0409)	(0.0425)	(0.0396)
Voucher (tsh '000)	0.00386***	0.00318***	0.00389***	0.00290**	0.00244*	0.00301**
	(0.00113)	(0.00120)	(0.00120)	(0.00128)	(0.00136)	(0.00137)
Gender voucher ('000)	0.00388***	0.00394***	0.00424***	0.00294***	0.00316***	0.00342***
	(0.000946)	(0.000974)	(0.000994)	(0.00111)	(0.00116)	(0.00116)
Attended meeting	0.192***	0.126**	0.124**	0.203***	0.133**	0.129**
	(0.0514)	(0.0558)	(0.0550)	(0.0531)	(0.0567)	(0.0561)
% neighbours attended	-0.164**	-0.118	-0.119	-0.180**	-0.132	-0.128
	(0.0832)	(0.0854)	(0.0878)	(0.0850)	(0.0866)	(0.0895)
% neighbours out of block	0.0202	0.00630	0.00511	0.0499	0.0290	0.0359
	(0.0493)	(0.0503)	(0.0222)	(0.0518)	(0.0531)	(0.0516)
Constant	0.148*	0.176**	5.865	-0.00645	0.0559	0.00114
	(0.0798)	(0.0820)	(4.209)	(0.111)	(0.115)	(0.112)
Baseline controls	No	No	Yes	No	No	Yes
Adj. R-Square	0.110	0.106	0.121	0.0784	0.0865	0.0946
Obs	456	421	421	456	421	421
C-D Wald F-stat				84.52	67.75	75.94

Dependent variable is a dummy variable = 1 if the household purchases a CRO

Basic columns include only controls shown + # of neighbours attending meeting and a control for whether household has neighbours outside treatment block

Restricted columns are the same as basic, except sample and neighbour sets are restricted to households with non-missing baseline data

Restricted + Controls columns include household and average neighbour set controls for Log(parcel area), year of purchase, rental status, owner residence RL ownership, electricity access, number of buildings, recent parcel investment, monthly income, assets, average schooling and hh size

Instruments in 2SLS specification: average priced faced by neighbours (setting untreated parcels at price = tsh 500,000

Conley-adjusted standard errors in parentheses, $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

OLS estimates of the endogenous peer effect ρ are positive and of similar size, even when including baseline controls, with the predicted probability that household i purchases a land title increasing by 7-8 percentage points with each neighbour that takes up. When instrumented, these estimates nearly double, with the probability that the household purchases a CRO increasing by 14-15 percentage points with each neighbour that takes up. In previous literature, IV estimates of peer effects are nearly always higher than the OLS estimates. In a pure Manski world, this is perplexing, as simultaneity bias and correlated effects should, on average, lead to bias away from zero, rather than towards it.

One possibility relies on the local average treatment effect interpretation of the estimated coefficient: as ρ is estimated using 2SLS, it is defined only over households whose neighbours were affected by the treatment, thus leaving out all households with neighbours who decided, despite facing large subsidies, not to purchase a land title. This decision might convey unobserved information which also interacts with the mechanisms driving peer effects: for example, the choice of a neighbour not to purchase a title might reveal that expropriation complementarities are not expected to be particularly strong in a given location. Also, if some neighbours never intend to adopt CROs (even if they were to face a price of zero) their non-adoption might convey little-to-no information to other households, resulting in lower average peer effects when they are included.

The other possible reason why 2SLS results are higher than OLS is due to a mechanical downward bias in OLS estimates inherent in most endogenous peer effects models. Guryan, Kroft, and Notowidigdo (2009) show that when peer groups are constructed which *exclude* the household itself and peers are considered as observations as well, OLS estimates will be biased downward.[23] Guryan et al. (2009) also show that controlling for the average take-up of the pool from which a household's peers are selected corrects for this bias. However, in the current context, this 'pool' comprises all observations from Barafu except for the household of interest: as all variation in the pool average is being driven by variation in T_i, it is impossible to include it as a control. Caeyers (2013) shows that this bias is removed when using 2SLS, as valid instruments for $\overline{T}_{g(-i)}$ also side-step

[23]The intuition is as follows: as households are being excluded from their own peer group, if the household had a high value of the outcome of interest Y_i then the resulting peer group will have, in expectation, a lower average outcome $\overline{Y}_i$. When, in turn a household from the same group with a low value of Y_i is considered, the constructed peer group will have a higher average value.

the mechanical bias, hence resulting in higher estimates under 2SLS than OLS.

Both types of vouchers have strong, significant effects on take up. Meeting attendance is correlated with higher take-up, although it is unclear if this due to the effect of the meeting or driven by unobserved demand for CROs. Interestingly, neighbour attendance of meetings is negatively correlated with CRO adoption, indicating that the direction of information channels is not straightforward. As meeting attendance is endogenous, Table 2.13 in the appendix shows the main results still hold when meeting attendance is excluded from the specification. The dummy indicating that the household has neighbours living outside the treatment block does not appear to be a significant correlate of adoption.

The voucher results give us a novel way to interpret the size of the peer effect results. In the 2SLS specification with baseline controls a 1,000 TSh voucher is associated with approximately a .03% increase in the predicted probability that a household purchases a CRO, the decision of a nearest-neighbour to purchase a CRO leads to approximately a 15% increase. Thus, the peer effect generated by a single neighbour adopting is roughly equivalent to a 50,000 TSh voucher transfer.

That peer effects are large and strictly positive suggests positive strategic complementarities in the purchase of CROs. I will investigate this further using a variety of robustness checks throughout this section. More substantial robustness checks are performed in Appendix 2.A.2, where I show these results a robust to the inclusion of block fixed effects and controls for the take-up decisions of household's outside of the nearest-neighbour set.

2.6.1 Distance and social connections

To confirm that these results aren't isolated to a single specification, Table 2.4 shows estimates of ρ across different nearest-neighbour sets. In both the OLS and 2SLS specifications, peer effects are strong, positive and significant. Table 2.11, located in the appendix, shows these results to be similar when using distance-bands. From these results, it is clear that peer effects are decreasing with distance. The average effect per-neighbour in the three-neighbour 2SLS specification is roughly seven times greater than the twenty-neighbour neighbour one (although this gradient is less steep for the OLS and distance-band specifications). Figure 2.3 shows the decrease in the effect for both

Table 2.4: Barafu - impact of neighbour's adoption for nth nearest-neighbour sets

	(1)	(2)	(3)	(4)	(5)
OLS					
Basic	0.0933**	0.0773**	0.0513**	0.039**	0.0299**
	(0.0232)	(0.017)	(0.0104)	(0.0086)	(0.0074)
Restricted	0.0944**	0.0836**	0.048**	0.0365**	0.0302**
	(0.0236)	(0.0173)	(0.0112)	(0.0088)	(0.0072)
Covariates	0.0914**	0.0835**	0.0448**	0.0301**	0.028**
	(0.0242)	(0.0182)	(0.0118)	(0.0101)	(0.0078)
2SLS					
Basic	0.2339**	0.147**	0.0483**	0.0349**	0.0239**
	(0.0593)	(0.0409)	(0.0194)	(0.0123)	(0.0094)
Restricted	0.1896**	0.1368**	0.0474**	0.0302**	0.0217**
	(0.0607)	(0.0425)	(0.0203)	(0.0129)	(0.01)
Covariates	0.2031**	0.1478**	0.0611**	0.0404**	0.0292**
	(0.0629)	(0.0396)	(0.0199)	(0.0127)	(0.0088)
# nearest neighbours =	3	5	10	15	20

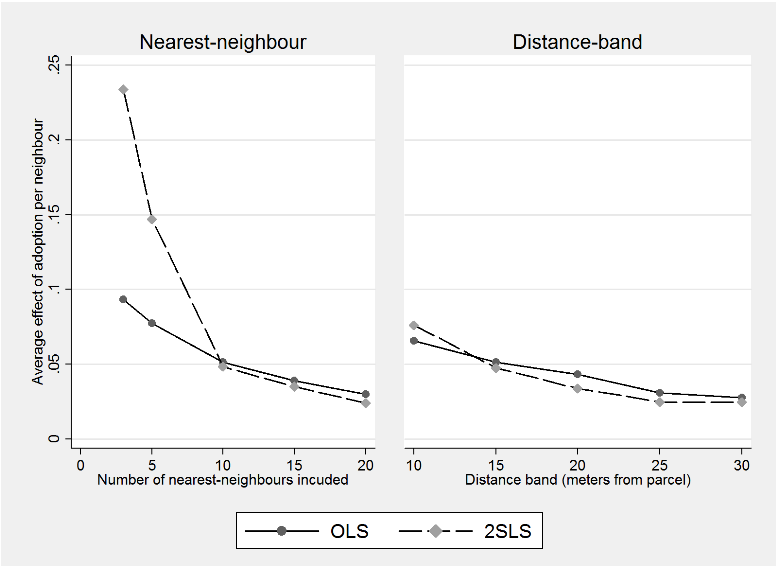
Dependent variable is a dummy variable = 1 if the household purchases a CRO. **"Basic"** rows include only controls shown & # of neighbours attending meeting and a control for whether household has neighbours outside treatment block. **"Restricted"** rows are the same as basic, except sample and neighbour sets are restricted to households with non-missing baseline data. **"Covariates"** columns include household and average neighbour set controls for Log(parcel area), year of purchase, rental status, owner residence. Each column represents a different nearest-neighbour set (i.e. 3 = 3 closest neighbours). Conley standard errors in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

nearest-neighbour and distance-band approaches as the number of neighbours included is increased.[24] While this is consistent with the story of spatially-proximate strategic complementarities, it is also possible that physical distance might just be a convenient proxy for social distance, as those who live close to one another are more likely to interact on a day-to-day basis. To understand whether or not signaling might be driving positive estimates of ρ, information on social networks with some degree of independence from geographic proximity would be necessary.

Data taken during the baseline survey might prove helpful in solving this conundrum. Prior to the baseline data collection, for each of fifteen administrative blocks of households (note that these blocks do not correspond to the blocks used for the experiment) a random

[24]While it is possible that these effects could be generated by bias in estimates of ρ when using inappropriately-small neighbour sets, I show in Section 2.A.2 in the Appendix that the distance effects hold even when controlling for take-up outside the neighbour set of interest.

Figure 2.3: Average neighbour peer effect as neighbour set increases in distance



sample of ten households were chosen to form a network questionnaire. During the baseline survey, each household was asked if they knew the head of each household from the network roster. For all households with baseline data, I have matched up those listed on the network roster with programme take-up data. Matching these responses in the network questionnaire has allowed me to construct a limited dyadic sample of 402 parcels, each with 9.24 links on average, for a total of 3,718 observations. The i dimension of the dyad includes all treated households with responses to the network questionnaire. The j dimension includes all of those listed on the roster with take up data. This will allow me to investigate whether adoption peer effects are higher for households closer together or those that know each other.

Table 2.5 shows the results from a regression of i's probability of take up on j's take up, including an interaction term if household i knows household j and a second interaction for the geographic distance between i and j in meters. Standard errors are clustered at both the i and j level using Cameron et al.'s (2011) method, which provides a good approximation of the dyad-specific approach proposed by Fafchamps and Gubert (2007). The first column of Table 2.5 shows the results using OLS, which show that j's purchase of a CRO is associated with a 10% increase in the probability that i purchases a CRO. This effect increases by roughly one percentage point if i knows j, but the effect is insignificant at the 10% level. However, the peer effect decreases with distance: the effect is 1% lower for every 15 meters of distance between the two households. Column (2) shows a 2SLS specification, again using aggregate price of a CRO as an instrument for j's take-up.[25] The coefficients in the 2SLS specification are very similar to those of OLS, with the negative coefficient on the distance interaction being nearly identical and still significant at the 10% level.

While the results here are based on a limited sample (those who answered the network questionnaire and those who were randomly selected to be on the network questionnaire), they do suggest that peer effects are primarily running through physical proximity, rather than ex-ante familiarity between households. While not conclusive, this indicates that complementarities in the marginal gain from CRO adoption, rather than signaling or information flows, are likely to be driving the large positive endogenous peer effects observed in the data.

2.6.2 Planned areas

So far, the results that I have presented are solely for unplanned areas of Barafu, with neighbour sets constructed only out of neighbours who have previously not had a cadastral survey. If the knowledge that a treated neighbour will be surveyed increases the perceived value of a CRO, then the instrument might have an effect on take up, outside of the peer effect. To determine whether or not this might be a problem, I first use the baseline data to check and see if there is any correlation between perceived expropriation risk and proximity to surveyed parcels, and I find none (these results are presented in Appendix 2.A.3). Next, I re-run the main specification, this time including surveyed parcels as neighbours. If part of the observed peer effect is actually proxying for a surveyed-effect, then peer effects from already-surveyed neighbours should be lower. However, including already-surveyed neighbours does not seem to change the results in any meaningful way. Figure 2.4 shows a comparison between estimates of ρ from both before and after planned neighbours are included, using the basic specification without covariates. Table 2.12 in

[25]To instrument the interaction terms, I use interactions between the main instrument (average neighbour price) and the two dummies of interest, i knowing j and the distance between i and j.

Table 2.5: Impact of neighbour's CRO take up on own take up - matched network list

	(1) OLS	(2) 2SLS
Household j is adopting	0.103**	0.137**
	(0.0425)	(0.0615)
(j adopting) * (i knows j)	-0.00437	0.00769
	(0.0716)	(0.104)
(j adopting) * (i-j distance)	-0.000656**	-0.000709*
	(0.000274)	(0.000379)
Household i knows household j	0.0509	0.0399
	(0.0500)	(0.0647)
Distance between parcels i and j	0.000397***	0.000438***
	(0.000135)	(0.000156)
Unconditional voucher	0.00264**	0.00265**
	(0.00122)	(0.00122)
Conditional voucher	0.00469***	0.00466***
	(0.000974)	(0.000977)
Constant	0.317***	0.302***
	(0.0558)	(0.0574)
Adj. R-Square	0.0515	0.0506
Obs	3718	3718
C-D Wald F-stat		15.32

Dependent variable is a dummy variable = 1 if household i purchases a CRO
Instruments in 2SLS specification: j household in treatment block, (i knows j)*(j treated) and (i-j distance)*(j in treatment block). Robust standard errors in parentheses, two-level clustering at both i and j parcel level. $^{*}p < 0.10$,$^{**}p < 0.05$,$^{***}p < 0.01$

Appendix A shows the full set of results, revealing no substantial difference. Given that peer effects for already-surveyed parcels are of a similar magnitude to unplanned parcels, it does not appear that surveying of neighbours is a key factor in take-up decisions.

2.6.3 Peer effects and baseline perceptions of expropriation risk

In order to investigate further the role of expropriation risk in this context, I turn to baseline data on the parcel owner's perceived risk of expropriation (presented earlier in Figure 2.1). While respondents could choose probabilities anywhere from zero to one, predictions were clumped around zero, 0.5 and 1, as Figure 2.5 illustrates. To see if those who believe themselves to be at a higher risk of expropriation are more responsive to their

Figure 2.4: Average nearest-neighbour peer effects for both unplanned and all neighbours

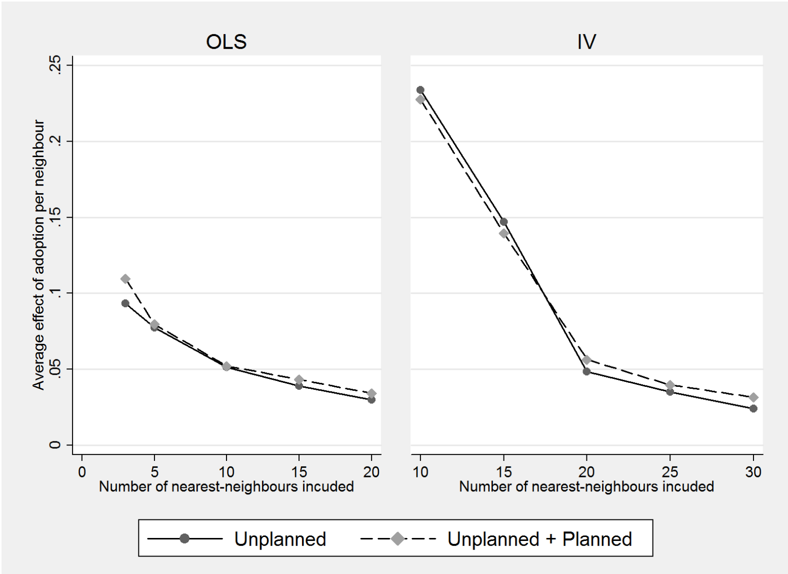
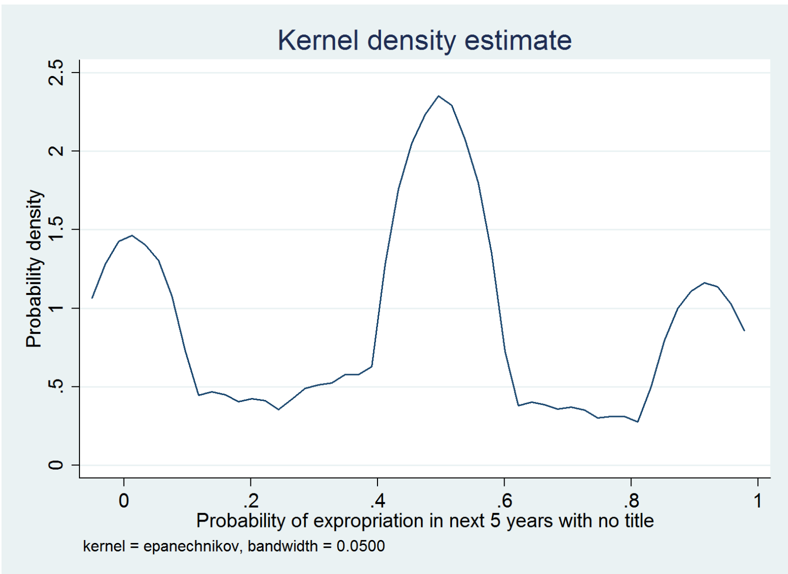


Figure 2.5: Barafu - density of perceived expropriation risk



neighbours' adoption, I create a dummy variable ($exprop_i$) which is equal to one if the house reported their perceived expropriation risk (conditioned on not having any form of title) to be greater than or equal to 50%. I then interact this dummy with $\overline{T}_{g(-i)}$, the adoption rate of the neighbours, and proceed with the same specification as before. For the 2SLS estimates, I retrieve predicted values of $\widehat{\overline{T}_{g(-i)}}$ from the first stage regression, then use these predicted values and their interactions $\widehat{\overline{T}_{g(-i)}} \times exprop_i$ as instruments for $\overline{T}_{g(-i)}$ and $\overline{T}_{g(-i)} \times exprop_i$.[26]

Table 2.6 shows the results from both the OLS and 2SLS estimation for three different neighbour sets, all with baseline controls included. In all OLS specifications, the interaction effect is significant and positive, where the 2SLS results show a significant effect at the 10% level in the two largest-neighbour sets (columns (4) and (6)). These results suggest that the peer effect is stronger for those that had a higher ex-ante perceived probability of expropriation. The coefficient of the level effect of $exprop_i$ is consistently large, negative and significant in most specifications. It appears that while households with a higher ex-ante expropriation risk are more responsive to peer effects, they have a lower absolute level of take-up. This is consistent with a model in which households with a high perceived risk only bother to purchase if they observe others around them doing them same, suggesting that there are complementarities in the reduction of expropriation risk. It does not, however, rule out other possible channels for complementarities, such as credit or house prices, which might imply even larger externalities.

[26] Wooldridge (2010) suggests that there are efficiency gains when using predicted values as interaction terms, rather than the original instruments.

Table 2.6: Barafu - interaction between perceived expropriation risk and impact of neighbour's CRO take up

	5 nearest		10 nearest		15 nearest	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
# of neighbours adopting	0.0284	0.129**	0.000145	0.0310	-0.000481	0.0204
	(0.0350)	(0.0571)	(0.0197)	(0.0277)	(0.0161)	(0.0190)
High exprop risk × # adopting	0.0771*	0.0318	0.0682***	0.0527*	0.0464***	0.0337*
	(0.0393)	(0.0581)	(0.0228)	(0.0305)	(0.0176)	(0.0195)
High exprop risk	-0.260**	-0.111	-0.444***	-0.351*	-0.443***	-0.331*
	(0.131)	(0.186)	(0.141)	(0.184)	(0.158)	(0.174)
Voucher (tsh '000)	0.00404***	0.00301**	0.00451***	0.00423***	0.00464***	0.00442***
	(0.00125)	(0.00136)	(0.00124)	(0.00122)	(0.00122)	(0.00119)
Pink voucher (tsh '000)	0.00417***	0.00333***	0.00460***	0.00427***	0.00495***	0.00469***
	(0.00105)	(0.00115)	(0.00102)	(0.00101)	(0.00100)	(0.000971)
Attended meeting	0.122**	0.129**	0.133**	0.144***	0.122**	0.132**
	(0.0571)	(0.0560)	(0.0541)	(0.0531)	(0.0535)	(0.0522)
Constant	7.156	4.853	9.963	9.981	6.037	7.317
	(10.90)	(10.43)	(16.30)	(15.50)	(19.73)	(18.80)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-Square	0.127	0.0927	0.151	0.143	0.145	0.140
Obs	421	421	421	421	421	421
C-D Wald F-stat		32.48		62.17		144.3

Dependent variable is a dummy variable = 1 if the household purchases a CRO. **High expropriation risk** a dummy = 1 if household's perceived probability of expropriation >= 50%. Specification includes main and baseline controls discussed in previous tables. **Instruments** in 2SLS specification: predicted values from first stage regression of 2SLS regression (using average priced faced by neighbours (setting untreated parcels at price = 500,000 TSh) as an instrument, interacted with high expropriation risk. Robust standard errors in parentheses. $^{*}p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$

2.7 Additional results

In this section I will present results from two other sources of data, first from a replication of the above experiment in Kigogo Kati, an adjacent community, and then from administrative data from the Kinondoni Municipality.

2.7.1 A second experiment

Nearly a year following the allocation of treatment and control blocks and the voucher distribution in Mburahati Barafu, the same intervention was introduced in Kigogo Kati, which borders Barafu to the south. Due to the length of time between the two interventions, this provides us with an interesting replication of the Barafu experiment. Other than location, Kati differs from Barafu in two key ways which might affect estimates of peer effects. Firstly, shortly after the intervention began, Dar es Salaam was subject to some of the worst flooding in 60 years, with Kati being one of the areas which was affected the most. This subsequently depressed CRO adoption, as households were subject to a shortfall in income. While this should not necessarily dampen peer effects, the low levels of take-up (roughly 15% versus approximately 60% in Barafu), indicate that the instruments used to identify peer effects will be significantly weaker.

Secondly, Kati has been the recipient of a community infrastructure upgrading project (CIUP) for several years, which has led to a number of parcels being demolished to make way for road expansion and electrification. This increased probability of expropriation and the changes in the gains for land titling which might come from being in a heavily invested area are both likely to interact with peer effects. Furthermore, while the take up data from Barafu is considered complete, Kati is still in the process of collecting repayment and soliciting more participants, so these results should be considered preliminary.

Table 2.7 replicates the same specification seen in Table 2.3 for the five nearest-neighbours, first showing the results for OLS with and without baseline covariates and then using 2SLS. In order to maximize the explanatory power of the instrument, I use average voucher values and average assignment-to-treatment as individual instruments, rather than the composite price measure I used in the previous section. For the OLS

Table 2.7: Kati - impact of neighbour's CRO take up on own take up - 5 closest neighbours

	OLS			2SLS		
	(1) Basic	(2) Restricted	(3) Restricted + Controls	(4) Basic	(5) Restricted	(6) Restricted + Controls
# of neighbours adopting	0.0944***	0.0980***	0.0788***	0.206***	0.225***	0.231**
	(0.0186)	(0.0192)	(0.0211)	(0.0737)	(0.0749)	(0.0915)
Voucher (tsh '000)	0.00114*	0.00117*	0.00123*	0.000819	0.000777	0.000953
	(0.000604)	(0.000653)	(0.000639)	(0.000649)	(0.000724)	(0.000688)
Gender voucher ('000)	0.00147**	0.00153**	0.00153**	0.00114	0.00116	0.00129*
	(0.000641)	(0.000672)	(0.000643)	(0.000695)	(0.000736)	(0.000700)
Attended meeting	0.133***	0.136***	0.141***	0.138***	0.142***	0.148***
	(0.0260)	(0.0274)	(0.0292)	(0.0274)	(0.0296)	(0.0321)
Constant	-0.0202	-0.00417	0.171	-0.0450	-0.0331	-0.0365
	(0.0461)	(0.0518)	(2.177)	(0.0458)	(0.0498)	(0.0542)
Baseline controls	No	No	Yes	No	No	Yes
Adj. R-Square	0.0892	0.0922	0.107	0.0135	-0.00721	-0.0217
Obs	684	615	615	684	615	615
C-D Wald F-stat				12.38	13.38	9.129

Dependent variable is a dummy variable = 1 if the household purchases a CRO

Basic columns include only controls shown + # of neighbours attending meeting and a control for whether household has neighbours outside treatment block

Restricted columns are the same as basic, except sample and neighbour sets are restricted to households with non-missing baseline data

Restricted + Controls columns include household and average neighbour set controls for Log(parcel area), year of purchase, rental status, owner residence RL ownership, electricity access, number of buildings, recent parcel investment, monthly income, assets, average schooling and hh size

Instruments in 2SLS specification: average programme treatment status and average priced faced by neighbours

Conley-adjusted standard errors in parentheses, $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

specification, peer effects are of similar magnitude to the results from the Barafu experiment, with each neighbour adopting associated with a 8-9.5% increase in the probability the household will also adopt. As before, the 2SLS result is significantly higher, with each neighbour adopting associated with a 20% increase in the predicted probability (note that this is the maximum peer effect size allowed in this specification). Note that the Cragg-Donald Wald F statistic, reported at the bottom of the table, is quite low, so these results should be taken with some caution.[27]

In Appendix 2.A.1, Table 2.14 shows the results from replication of the specification across different-sized neighbour sets. The results seem reasonably robust to variations of the peer group, again showing a decreasing effect as the neighbour set grows to include parcels which are further away. Not all results are significant at the 10% level, but the coefficients are on the same order of magnitude of the results from Barafu.

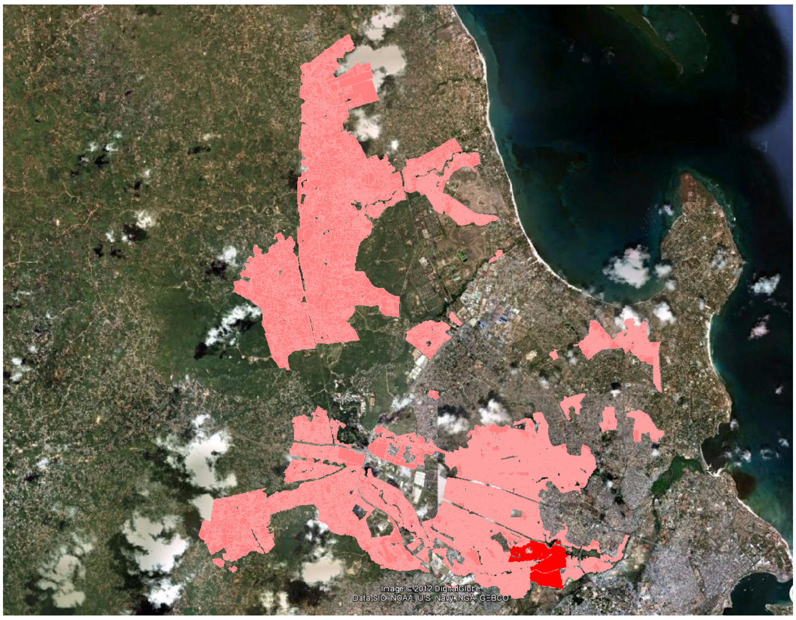
2.7.2 Peer effects in residential license take-up

While endogenous peer effects appear to be a determinant of take-up in both Mburahati Barafu and Kigogo Kati, it is not immediately clear that the results are generalisable to other settings or necessarily scalable. This is a common criticism of micro-empirical work, including most randomised controlled trials (Ravallion 2008; Deaton 2010) and one that is rarely dealt with.

Ideally, a replication of the experiments in Barafu and Kati at a larger level would show that such results are scalable. For lack of such an experiment, I turn to administrative data: the Kinondoni Municipal land office keeps records of each and every unplanned parcel in the municipality, including basic characteristics for both the parcel and the owning household. Also available are records for every purchase of a residential license, the short-term land title mentioned in Section 2.2, from the time they first became available in mid-2005 until late 2008. Finally, I also have access to GIS shapefile data on the size and location of each of the 63,000 unplanned land parcels in Kinondoni, the same data I used to show that adoption decisions were correlated in 1. Figure 2.6 displays the location of these land parcels, including Barafu and Kati. By matching these three data sources

[27] Also, this CDW F-stat has not been corrected for spatial correlation.

Figure 2.6: Location of Kinondoni Municipality GIS data



Dark red shapes indicate Mburahati Barafu and Kigogo Kati

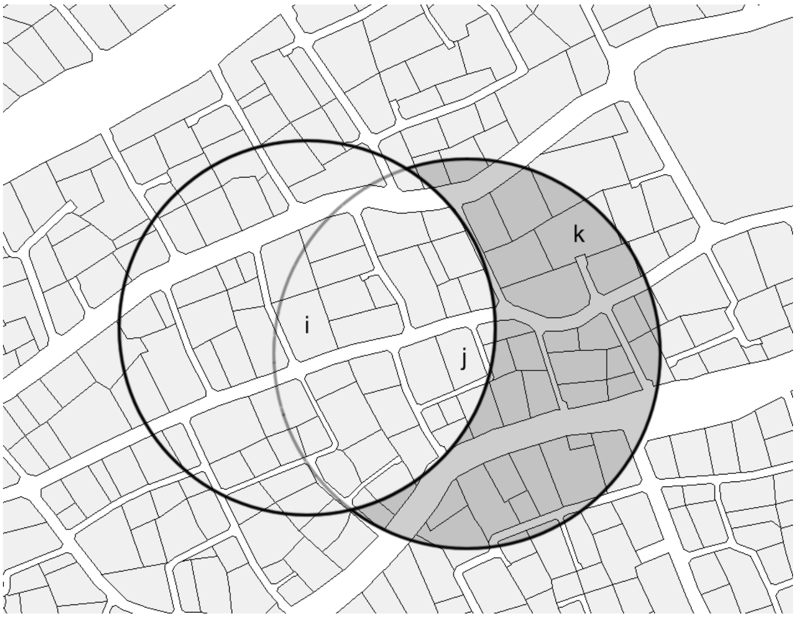
together, it is possible to investigate whether or not peer effects in residential license adoption exist at a larger scale.

While there is no experimental variation in residential license take up, the peer effects literature has developed several methods of identifying peer effects, given some limiting assumptions on how neighbours interact. A common method of overcoming the reflection problem is to take advantage of the structure of partially-overlapping peer groups. This is this case when peer group structures are not transitive; for example, when j being part of of i's peer group and k being part of j's peer group does not guarantee that k will be in i's peer group. When this is the case, there were will be characteristics in k's equation which can be used as instruments for j's adoption.

The intuition is this: k's exogenous characteristics affect k's adoption decision directly, and thus j's adoption decision indirectly (through the endogenous peer effect).[28] Since i doesn't directly interact with k, the latter's exogenous characteristics *only* affect i's adoption through j's adoption. Thus, k's exogenous characteristics satisfy the exclusion restriction and are potential instruments for j's take up. This method was developed

[28]In the presence of exogenous (contextual) effects, k's characteristics will also affect j's adoption decision directly.

Figure 2.7: Example of excluded neighbours



Note: Black circles indicate the boundaries of i and j's neighbour set. Shaded area indicates *excluded* neighbours.

simultaneously by Bramoullé, Djebbari, and Fortin (2009) and De Giorgi, Pellizzari, and Redaelli (2010) and has since become a popular method of overcoming the simultaneity bias inherent in peer effects models.

Figure 2.7 shows a hypothetical case using a map of Mburahati Barafu. In this example, the shaded area indicates all parcels which are in j's peer group, but not in i's peer group, and therefore can be used as "excluded" neighbours. When peer groups are constructed spatially, the partially-overlapping requirement for intransitivity is usually met.

Letting $g(-i)$ indicates household i's neighbour set and $g^2(-i)$ indicate set of neighbours of i's neighbours (both discluding i), reconsider the empirical adoption equation:

$$T_i = \alpha + \rho \overline{T}_{g(-i)} + x_i \beta + \overline{x}_{g(-i)} \delta + u_i + \epsilon_g + \varepsilon_i$$

To identify ρ, we need to instrument $\overline{T}_{g(-i)}$ with $\overline{x}_{g^2(-i)}$, the average exogenous characteristics of the neighbours of i's neighbours.[29] The exclusion restriction, that $\overline{x}_{g^2(-i)}$ only

[29]Recall that the identifying information is coming only from *excluded* neighbours. Those which are

affects T_i through $\overline{T}_{g(-i)}$, is heavily dependent on the assumption that household i doesn't interact with households outside of its designated peer group. While this assumption is more easily defended when peer groups are defined by a rigid structure (such as friends in a network roster), in a dense slum it is a little more precarious. Furthermore, while the neighbours-of-neighbours approach theoretically deals with the reflection problem, it does not eliminate correlated effects. For example, if wealth is positively correlated with unobserved land quality, but also affects residential license adoption, then the wealth of excluded neighbours might be correlated with unobserved land quality in i's equation as well.

My approach is as follows: first I restrict the sample and neighbour-sets to all non-empty parcels with non-missing observations for a set of characteristics, which reduces the total sample size to approximately 44,000 land parcels. This includes the area of the parcel, a dummy for the parcel being used for residential purposes, a dummy for the parcel being used for both residential and commercial purposes, whether the parcel is on hazard land or not (land deemed unlivable by the government), the number of rooms, the number of households living there, the number of people living there, a dummy equal to one if the parcel has a positive property value, and an interaction between the positive property value dummy and the natural log of the household's value in Tanzanian shillings.

This set of characteristics will be included both as a set of controls for the household/parcel in question i, averaged across the household's neighbour set. For instruments, I have taken a subset of these characteristics for the neighbour's excluded neighbours which are the most informative about $\overline{T}_{g(-i)}$ (the number of rooms, people living on the parcel, and the property value variables). To account for correlated effects, I have included first mtaa/ward fixed effects, then administrative block fixed effects.

Table 2.8 displays the results from the estimation of the five nearest-neighbours peer effects specification, using residential license take-up as the measure of property rights adoption. The first two columns show the OLS results, while controlling for neighbour characteristics, first with mtaa fixed effects then with block fixed effects. The second two columns show results from the 2SLS specification, when the neighbour-set's average

part of j's neighbour set but not part of i's neighbour set.

take up of CROs is instrumented with the characteristics the neighbours of neighbours. Both OLS and 2SLS estimates are of similar magnitude to what was seen in both Kigogo Kati and Mburahati Barafu. However, the 2SLS estimates which incorporate controls for correlated effects, when administrative block fixed effects are included, are approximately 20% smaller than the estimates for CRO adoption (11.8% in Table 2.8 versus 14.8% in Table 2.3).

Given that residential licenses only have a limited tenure value, as they must be renewed every five years, it is possible that there is less room for complementarities in reducing expropriation risk. Furthermore, given the differences in the two forms of tenure, and the fact that these results are based on a four-year span of adoption, it is likely that the exogenous peer effects revealed here are operating through entirely different channels than in the field experiments.

Finally, Table 2.9 shows the results for other neighbour set sizes. Again, the results are of a very similar magnitude to what was seen in the experimental data, especially when the neighbour set is extended to the twenty nearest-neighbours. Peer effect estimates also seem to decline as the size of the neighbour set grows. It should be noted that most of these specifications seem to suffer from a weak instrument problem, as Kleibergen-Paap F statistic (standard errors are clustered at the block level, violating the homoscedasticity assumptions necessary for using standard Cragg-Donald test)r is very low. Also, some specifications also fail their overidentification tests, suggesting that the instruments here are not entirely valid.[30] Despite these problems, there is still evidence here that residential license uptake decisions are correlated, possibly as a result of strategic complementarities in their adoption.

[30] Although in a world of heterogenous peer effects, the Hansen J test may just be highlighting the local average treatment effect interpretation.

Table 2.8: Kinondoni - impact of neighbour's RL adoption on own adoption - 5 nearest neighbours

	OLS		2SLS	
	(1)	(2)	(3)	(4)
# neighbours adopting RL	0.0562***	0.0388***	0.130***	0.118**
	(0.00536)	(0.00334)	(0.0471)	(0.0574)
Elevation (m)	-0.0143***	-0.0148***	-0.0138***	-0.0141***
	(0.00391)	(0.00412)	(0.00380)	(0.00398)
Parcel area (100 sqm)	0.0109**	0.0116**	0.0118**	0.0120***
	(0.00474)	(0.00457)	(0.00464)	(0.00461)
Mixed use plot	-0.0629***	-0.0607***	-0.0621***	-0.0611***
	(0.0190)	(0.0203)	(0.0186)	(0.0206)
Residential plot	-0.0392***	-0.0357***	-0.0352***	-0.0334***
	(0.0112)	(0.0115)	(0.0108)	(0.0118)
# rooms in house	0.0119***	0.0118***	0.0111***	0.0112***
	(0.00120)	(0.00107)	(0.00147)	(0.00120)
# households living on plot	-0.0136***	-0.0134***	-0.0137***	-0.0136***
	(0.00145)	(0.00149)	(0.00144)	(0.00156)
# people living on plot	0.000635	0.000584	0.000799*	0.000772
	(0.000437)	(0.000515)	(0.000428)	(0.000542)
Plot is on hazard land	-0.482***	-0.482***	-0.482***	-0.482***
	(0.0245)	(0.0149)	(0.0251)	(0.0158)
Property value > 0	-0.124***	-0.115***	-0.132***	-0.126***
	(0.0165)	(0.0166)	(0.0175)	(0.0193)
Ln(Prop value) * (value > 0)	0.0101***	0.00975***	0.0106***	0.0104***
	(0.000969)	(0.000966)	(0.00102)	(0.00111)
Constant	0.174*	-2.18e-09		-2.59e-09*
	(0.0958)	(1.50e-09)		(1.57e-09)
Neighbour controls	Yes	Yes	Yes	Yes
Mtaa fixed effects	Yes	No	Yes	No
Admin block f.e.	No	Yes	No	Yes
Adj. R-Square	0.147	0.0739	0.0549	0.0418
Obs	44162	44162	44162	44162
K-P Wald F-stat			3.846	4.005
Hansen J p-value			0.100	0.0399

Dependent variable is a dummy variable = 1 if the household purchases a RL in 2005-2008
Neighbour controls are average values of household/parcel characteristics for neighbour set
Instruments in 2SLS specification are average values of household/parcel characteristics for excluded neighbours of neighbours. Standard errors clustered at mtaa level in columns (1) and (3) and at block level in (2) and (4). $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table 2.9: Kinondoni - 2SLS results - different neighbour sets

	Nearest-neighbour groups			
	5	10	15	20
# neighbours adopting RL	0.118**	0.0818***	0.0709***	0.0329***
	(0.0574)	(0.0235)	(0.0174)	(0.0121)
Neighbour controls	Yes	Yes	Yes	Yes
Admin block f.e.	Yes	Yes	Yes	Yes
Adj. R-Square	0.0418	0.0399	0.0259	0.0638
Obs	44162	45524	46163	46445
K-P Wald F-stat	4.005	3.846	3.500	4.820
Hansen J p-value	0.0399	0.0637	0.0756	0.610

Dependent variable is a dummy variable = 1 if the household purchases a RL in 2005-2008
Neighbour controls are average values of household/parcel characteristics for neighbour set
Instruments in 2SLS specification are average values of household/parcel characteristics for excluded neighbours of neighbours. Standard errors clustered at block level.
$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

2.8 Conclusion

The options for the many developing countries grappling with high levels of urban growth could be boiled down to formality-by-force or formality-by-nudge. The former is characterized by high levels of urban planning and slum clearance, bringing new arrivals immediately into the formal system and dragging in older ones kicking and screaming. For the latter, the incentives to switch are introduced ex-post, through the introduction of simple, robust formal tenure systems and slum-upgrading. After advocating the former camp for decades following independence, the Tanzanian government has finally found itself pushing the latter. However, its efforts to entice informal settlements to shift to a new tenure system have broadly failed, partly due to the government's lack of knowledge of how to spur demand for land titles.

In this chapter, I set out to determine whether or not endogenous peer effects in land titling adoption exist. Using the results from two randomised controlled trials in Dar es Salaam, I exploited random variation in the incentive to title in order to identify the impact of a neighbour's adoption on a household's propensity to adopt. The results suggest there are strong, positive endogenous peer effects, and these results are robust to different neighbour set specifications, as well as a replication of the main experiment in a second location. There is also evidence that positive exogenous peer effects are present at a much larger scale, as results from municipal records suggest that residential license take up show similar signs of being contagious. While the exact mechanism for these results is elusive, evidence strongly points towards geographic proximity as a determinant of the size of the peer effects. This, combined with evidence that households with a higher ex-ante perception of expropriation risk are more responsive to peer effects, suggests that perceived complementarities in risk-reduction are at least one of the drivers of the result. However, as data limitations make it difficult to test the other benefits of CRO ownership, such as a rise in house prices, it is also possible that strategic complementarities exist in these dimensions as well.

This chapter has established that not only is encouraging take-up possible, as evidenced by the effectiveness of the land titling programme, but it also has spillovers which can encourage larger levels of adoption. This is encouraging from a narrow policy perspective, but some caution must be taken in interpreting the results with respect to the expected benefits of land titling. While these complementarities appear to be encouraging take-up, the adoption of titles by some households could, for instance, increase perceived expropriation probabilities for other households, driving a rush towards titling which does not actually result aggregate welfare gains. If households are being driven by a fear of being left behind in an untenable informal system, peer effects could drive communities to make costly, but entirely unnecessary shifts between regimes. However, this chapter does provide robust evidence that the switch between tenure regimes exhibits contagion, with households encouraging each other to make the switch, although it cannot verify whether or not the switch is strictly beneficial.

If large positive externalities to land titling adoption do exist, then why haven't more communities embraced large scale formalisation, even without further government intervention? While the cost of an individual cadastral survey is prohibitively expensive, en-mass surveying can be considerably cheaper. Given that the demand for title has been shown to be substantial once these hurdles have been overcome (in this instance, by our intervention), the fact that households had not already coordinated to take advantage of these returns to scale suggests these communities already face significant barriers to collective action. These are not universally insurmountable, as there are a few examples of communities in Dar es Salaam coordinating to get the entire neighbourhood titled.[31] What remains to be seen is what policies best take advantage of this social multiplier effect and whether or not it is enough to ensure a full shift to a formal system.

[31] Magigi and Majani (2006) presents a case study of an informal community in Dar es Salaam with atypically high social capital organising a full cadastral survey of the entire unplanned settlement.

2.A Chapter 2 Appendix

2.A.1 Additional tables

Table 2.10: Barafu - main treatment and control balance

	Treatment	Control	T-test	OLS
	(1)	(2)	(3)	(4)
# Rooms	5.118	5.033	-0.560	0.085
	(2.604)	(2.175)	[0.575]	(0.149)
Electrical connection	0.739	0.733	-0.264	0.007
	(0.439)	(0.443)	[0.792]	(0.026)
Owner occupied	0.847	0.848	0.023	-0.000
	(0.360)	(0.360)	[0.982]	(0.022)
Tenants on parcel	0.636	0.602	-1.170	0.034
	(0.482)	(0.490)	[0.242]	(0.029)
Access to road	0.243	0.231	-0.471	0.012
	(0.429)	(0.422)	[0.638]	(0.025)
Log(parcel area m^2)	5.298	5.336	1.089	-0.038
	(0.605)	(0.594)	[0.276]	(0.035)

Notes: Columns (1)-(2) show means for treatment and control groups, standard deviations in ( ). Column (3) shows test statistic for t-test, p-values in [ ]. Column (4) shows coefficient and standard error for OLS regression of outcome variable on treatment.

Table 2.11: Barafu - impact of neighbour's adoption for neighbours within distance d

	(1)	(2)	(3)	(4)	(5)
OLS					
Basic	0.0657**	0.0512**	0.0432**	0.0307**	0.0277**
	(0.0124)	(0.0101)	(0.0086)	(0.008)	(0.0068)
Restricted	0.0652**	0.0458**	0.0418**	0.0294**	0.0267**
	(0.0136)	(0.0113)	(0.0094)	(0.0086)	(0.0075)
Covariates	0.0666**	0.0445**	0.039**	0.0245**	0.024**
	(0.0139)	(0.0119)	(0.0103)	(0.0098)	(0.0088)
2SLS					
Basic	0.0762**	0.0475**	0.0337**	0.0248**	0.0247**
	(0.0273)	(0.0178)	(0.0126)	(0.0099)	(0.0083)
Restricted	0.0806**	0.0383*	0.0338**	0.025**	0.0242**
	(0.033)	(0.0211)	(0.0145)	(0.0113)	(0.0097)
Covariates	0.1003**	0.0511**	0.0429**	0.0301**	0.0302**
	(0.0323)	(0.0202)	(0.0141)	(0.0116)	(0.0101)
Distance band length (m) =	10	15	20	25	30

Dependent variable is a dummy variable = 1 if the household purchases a CRO. **"Basic"** rows include only controls shown & # of neighbours attending meeting and a control for whether household has neighbours outside treatment block. **"Restricted"** rows are the same as basic, except sample and neighbour sets are restricted to households with non-missing baseline data. **"Covariates"** columns include household and average neighbour set controls for Log(parcel area), year of purchase, rental status, owner residence. Each column represents a different distance band (i.e. 10b = all neighbours within 10 meters) Conley-adjusted standard errors in parentheses. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

Table 2.12: Barafu - impact of neighbour's adoption for nth nearest-neighbour sets, including previously-surveyed neighbours

	(1)	(2)	(3)	(4)	(5)
OLS					
Basic	0.1094^{**}	0.0795^{**}	0.0521^{**}	0.0433^{**}	0.0341^{**}
	(0.0235)	(0.0167)	(0.0101)	(0.008)	(0.0071)
Restricted	0.1045^{**}	0.086^{**}	0.0501^{**}	0.0384^{**}	0.0323^{**}
	(0.0236)	(0.0169)	(0.0108)	(0.0085)	(0.0069)
Covariates	0.0995^{**}	0.0801^{**}	0.0436^{**}	0.0301^{**}	0.0296^{**}
	(0.0242)	(0.0178)	(0.0117)	(0.0099)	(0.0081)
2SLS					
Basic	0.2276^{**}	0.1395^{**}	0.0563^{**}	0.0396^{**}	0.0316^{**}
	(0.0665)	(0.0457)	(0.0222)	(0.0136)	(0.0105)
Restricted	0.2353^{**}	0.1537^{**}	0.0496^{**}	0.0349^{**}	0.0263^{**}
	(0.06)	(0.0434)	(0.0212)	(0.013)	(0.0101)
Covariates	0.2268^{**}	0.1513^{**}	0.0616^{**}	0.0362^{**}	0.03^{**}
	(0.0618)	(0.0414)	(0.0205)	(0.0133)	(0.0097)
# nearest neighbours =	3	5	10	15	20

Dependent variable is a dummy variable = 1 if the household purchases a CRO. **"Basic"** rows include only controls shown & # of neighbours attending meeting and a control for whether household has neighbours outside treatment block. **"Restricted"** rows are the same as basic, except sample and neighbour sets are restricted to households with non-missing baseline data. **"Covariates"** columns include household and average neighbour set controls for Log(parcel area), year of purchase, rental status, owner residence. Each column represents a different nearest-neighbour set (i.e. 3n = 3 closest neighbours). Conley-adjusted standard errors in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 2.13: Barafu - impact of neighbour's adoption for nth nearest-neighbour sets, without meeting controls

	(1)	(2)	(3)	(4)	(5)
OLS					
Basic	0.0917**	0.0735**	0.0476**	0.0351**	0.0246**
	(0.0232)	(0.0173)	(0.0107)	(0.0087)	(0.0075)
Restricted	0.0941**	0.0814**	0.0447**	0.0328**	0.0256**
	(0.0235)	(0.0173)	(0.0113)	(0.0091)	(0.0074)
Covariates	0.0919**	0.0822**	0.0417**	0.0267**	0.0248**
	(0.0242)	(0.0181)	(0.0121)	(0.0102)	(0.0078)
2SLS					
Basic	0.2478**	0.1421**	0.0392*	0.0262**	0.0147
	(0.0604)	(0.0435)	(0.0202)	(0.0132)	(0.0098)
Restricted	0.1911**	0.1309**	0.0352*	0.0211	0.0145
	(0.0611)	(0.0435)	(0.0213)	(0.0134)	(0.0097)
Covariates	0.2024**	0.1426**	0.0511**	0.0338**	0.0261**
	(0.0628)	(0.041)	(0.0213)	(0.0133)	(0.0088)
# nearest neighbours =	3	5	10	15	20

Dependent variable is a dummy variable = 1 if the household purchases a CRO. **"Basic"** rows include only controls shown & # of neighbours attending meeting and a control for whether household has neighbours outside treatment block. **"Restricted"** rows are the same as basic, except sample and neighbour sets are restricted to households with non-missing baseline data. **"Covariates"** columns include household and average neighbour set controls for Log(parcel area), year of purchase, rental status, owner residence. Each column represents a different nearest-neighbour set (i.e. 3n = 3 closest neighbours). Conley-adjusted standard errors in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table 2.14: Kati - impact of neighbour's adoption for nth nearest-neighbour sets

	(1)	(2)	(3)	(4)	(5)
OLS					
Basic	0.1141**	0.0944**	0.063**	0.0456**	0.0379**
	(0.0247)	(0.0186)	(0.0116)	(0.0094)	(0.0071)
Restricted	0.1201**	0.098**	0.0662**	0.0431**	0.0361**
	(0.0261)	(0.0192)	(0.0121)	(0.0095)	(0.0073)
Covariates	0.1122**	0.0788**	0.0503**	0.0316**	0.0247**
	(0.0269)	(0.0211)	(0.0149)	(0.0107)	(0.0086)
2SLS					
Basic	0.1869	0.2061**	0.0957**	0.0605**	0.034
	(0.1293)	(0.0737)	(0.0468)	(0.0277)	(0.0242)
Restricted	0.2249*	0.2248**	0.1079**	0.0606**	0.0324
	(0.1239)	(0.0749)	(0.0423)	(0.0252)	(0.0218)
Covariates	0.1437	0.2311**	0.1217*	0.0454	0.0256
	(0.1301)	(0.0915)	(0.0727)	(0.0511)	(0.0484)
# nearest neighbours =	3	5	10	15	20

Dependent variable is a dummy variable = 1 if the household purchases a CRO. **"Basic"** rows include only controls shown & # of neighbours attending meeting and a control for whether household has neighbours outside treatment block. **"Restricted"** rows are the same as basic, except sample and neighbour sets are restricted to households with non-missing baseline data. **"Covariates"** columns include household and average neighbour set controls for Log(parcel area), year of purchase, rental status, owner residence. Each column represents a different nearest-neighbour set (i.e. 3n = 3 closest neighbours). Conley-adjusted standard errors in parentheses. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

Table 2.15: Percentage of predictions outside of [0,1] in LPM model (Barafu)

# Neighbours	3	5	10	15	20
OLS					
Basic	0	0	.0022	0	0
Restricted	0	0	0	.0024	0
Restricted + Controls	.021	.038	.043	.036	.033
IV					
Basic	.036	.062	.04	.04	.036
Restricted	.02	.0022	.0022	0	0
Restricted + Controls	.019	0	0	0	0

Each cell shows the % of observations with predicted values which fall outside of the [0,1] interval for a given specification.
Results are from Barafu nearest-neighbour regressions.

2.A.2 Extra robustness: block fixed effects and outside-neighbour set adoption

To ensure that the main results in this chapter are not being driven by block-level unobservables, I have re-run the main specification with treatment block fixed effects. The results are presented in Table 2.16. The results are broadly similar to what has been seen before, although they are less-precisely estimated.

Another concern is the identification of the correct neighbour set. Under the assumption that a nearest-neighbour set of size n is the "correct" neighbour set (i.e. it captures all neighbours relevant for the CRO adoption decision), then estimates of ρ using smaller nearest-neighbour sets may be biased upward. For instance, if the main specification (2.3) is estimated using the index household's five nearest-neighbours, the titling decisions of neighbours outside of this group will influence *both* the five nearest-neighbors and the household in question. Therefore, nearest-neighbour sets which are "too small" will also be proxying for the larger, "correct" neighbour sets, and so per-neighbour estimates will be biased upwards. This does not prevent identification of ρ, but it does complicated its interpretation. Instrumenting might take care of this problem, but the instruments used here (the percentage of neighbour who are treated) will also be correlated across different-sized neighbour sets.

To account for this, Table 2.17 re-runs the main specification whilst controlling for the percentage of neighbours outside of the chosen neighbour set (up to twenty-nearest neighbours) who have purchased a CRO. So, for example, when the specified neighbour set is the five-nearest neighbours, a control is included for the % of excluded neighbours (those between the sixth and twentieth-nearest neighbours) who have also adopted a CRO. The results indicate that specifications using smaller neighbour sets might be upward-biased (the 2SLS estimates for three-nearest neighbours are 0.159 for the full specification, versus 0.203 in Table 2.4). However, estimates of larger neighbour sets are very close to what was seen in previous results.

Table 2.16: Barafu - impact of neighbour's adoption - nearest neighbour - block fixed effects

	(1)	(2)	(3)	(4)	(5)
OLS					
Basic	0.077**	0.0668**	0.0489**	0.0379**	0.0328**
	(0.0246)	(0.0182)	(0.0117)	(0.01)	(0.0098)
Restricted	0.0702**	0.0687**	0.0421**	0.0341**	0.0302**
	(0.0251)	(0.0191)	(0.0127)	(0.011)	(0.0112)
Covariates	0.0642**	0.068**	0.0352**	0.0189	0.0175
	(0.0253)	(0.0199)	(0.0139)	(0.013)	(0.0121)
2SLS					
Basic	0.2259**	0.1501**	0.0489*	0.0363**	0.0265*
	(0.0715)	(0.0507)	(0.0261)	(0.0168)	(0.0144)
Restricted	0.1664**	0.1398**	0.0511*	0.0314*	0.0177
	(0.0714)	(0.0532)	(0.0267)	(0.0187)	(0.0179)
Covariates	0.1825**	0.1446**	0.0644**	0.0479**	0.0312**
	(0.0734)	(0.0505)	(0.0276)	(0.0176)	(0.0156)
# nearest neighbours =	3	5	10	15	20

Dependent variable is a dummy variable = 1 if the household purchases a CRO. **"Basic"** rows include only controls shown & # of neighbours attending meeting and a control for whether household has neighbours outside treatment block. **"Restricted"** rows are the same as basic, except sample and neighbour sets are restricted to households with non-missing baseline data. **"Covariates"** columns include household and average neighbour set controls for Log(parcel area), year of purchase, rental status, owner residence. Each column represents a different nearest-neighbour set (i.e. 3 = 3 closest neighbours). Conley standard errors in parentheses. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

Table 2.17: Barafu - nearest neighbour - controlling for adoption outside of neighbour set

	(1)	(2)	(3)	(4)	(5)
OLS					
Basic	0.0943**	0.0781**	0.0515**	0.0389**	0.0299**
	(0.0226)	(0.0167)	(0.0104)	(0.0086)	(0.0074)
Restricted	0.0911**	0.0815**	0.0472**	0.0366**	0.0302**
	(0.023)	(0.0171)	(0.0112)	(0.0089)	(0.0072)
Covariates	0.0879**	0.0821**	0.0447**	0.0301**	0.028**
	(0.0233)	(0.0178)	(0.0118)	(0.0101)	(0.0078)
2SLS					
Basic	0.1992**	0.1266**	0.0458**	0.0352**	0.0239**
	(0.0581)	(0.0397)	(0.0182)	(0.0118)	(0.0094)
Restricted	0.1524**	0.1171**	0.041**	0.0303**	0.0217**
	(0.062)	(0.0444)	(0.0202)	(0.0128)	(0.01)
Covariates	0.1597**	0.1299**	0.0555**	0.0402**	0.0292**
	(0.0615)	(0.0406)	(0.0194)	(0.0125)	(0.0088)
# nearest neighbours =	3	5	10	15	20

Dependent variable is a dummy variable = 1 if the household purchases a CRO. **"Basic"** rows include only controls shown & # of neighbours attending meeting and a control for whether household has neighbours outside treatment block. **"Restricted"** rows are the same as basic, except sample and neighbour sets are restricted to households with non-missing baseline data. **"Covariates"** columns include household and average neighbour set controls for Log(parcel area), year of purchase, rental status, owner residence. Each column represents a different nearest-neighbour set (i.e. 3 = 3 closest neighbours). For all nearest-neighbour sets < 20, a control is included for the take up of households outside of the neighbour set, but within the 20-nearest neighbours cut-off. Conley standard errors in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

2.A.3 Cadastral survey proximity and perceived expropriation risk

To investigate this, I turn to the baseline data collected prior to the intervention. One of the questions asked in the survey requires the land owner to guess the probability that the parcel will be expropriated in the next five years, providing an excellent measure of self-perceived expropriation risk.[32] Using these responses by residents of unplanned areas, I have regressed the perceived probability of expropriation on several measures of proximity to already-surveyed parcels, distance to various geographic features in Barafu, and a set of household and parcel covariates. The results are displayed in Table 2.18.

The first two columns use two measures of proximity: distance of the household to the nearest cadastral-surveyed parcel and a dummy variable for whether or not the parcel is adjacent to a surveyed parcel. Neither is statistically significant at the 10% level, and while the coefficient on the adjacency dummy is of the "correct" sign, distance from a surveyed parcel seems to counter-intuitively reduce perceived expropriation risk. Columns (3) and (4) use the percentage of the nearest 20 neighbours who are surveyed as a measure, and while the coefficient is negative (indicating that parcels with more surveyed neighbours have lower perceived risk), it is not significant.

While one might be worried that measurement error in perceived expropriation risk might make it difficult to pick up *any* correlation, it is worth noting that many of these results do make sense. Proximity to "hazard land", areas which are deemed by the local government to be unsafe to build and are often subject to mass expropriation, is positive correlated with perceived expropriation risk. Similarly, proximity to the nearest primary road, where many parcels had already been marked for demolition to make way for expansion of existing infrastructure, is also correlated with higher perceived expropriation risk.

[32]Note that these data are only currently available for roughly 65% of the Barafu sample, so the following analysis might be subject to selection bias.

Table 2.18: Perceived expropriation risk at baseline and proximity to surveyed parcels

	Distance measures		Nearest-neighbor	
	(1)	(2)	(3)	(4)
Dist to nearest surveyed parcel (m)	-0.000103	0.000147		
	(0.000307)	(0.000430)		
HH is adjacent to surveyed parcel?	-0.0288	-0.0108		
	(0.0328)	(0.0343)		
% nearest 20 neighbours surveyed			-0.0967	-0.124
			(0.0898)	(0.110)
Distance to nearest church		0.000119		0.0000791
		(0.000654)		(0.000651)
Distance to nearest open field		0.000226		0.000209
		(0.000260)		(0.000256)
Distance to hazard land		-0.00191***		-0.00185***
		(0.000488)		(0.000464)
Distance to nearest mosque		0.00107*		0.00108*
		(0.000640)		(0.000638)
Distance to local govt office		0.000483		0.000537
		(0.000645)		(0.000642)
Distance to nearest walking path		0.000774		0.000699
		(0.000515)		(0.000498)
Distance to nearest primary road		-0.00207***		-0.00200***
		(0.000780)		(0.000762)
Distance to nearest river		0.000828		0.000784
		(0.000777)		(0.000773)
Distance to nearest secondary road		-0.000529		-0.000703
		(0.000444)		(0.000475)
Distance to nearest school		-0.00103*		-0.000953*
		(0.000530)		(0.000530)
Controls	No	Yes	No	Yes
R-Squared	0.000975	0.0979	0.00162	0.0991
Obs	755	755	755	755

Results are from an OLS regression of the household's self-reported perceived probability of expropriation within next five years on measures of proximity to parcels which have been cadastral surveyed. Sample is limited to households in un-surveyed areas. Robust standard errors in parentheses
$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Chapter 3

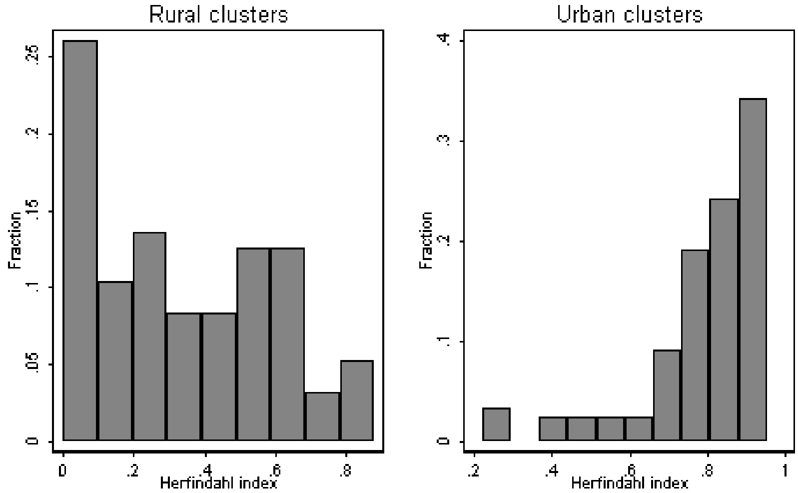
Ethnic enclaves and the demand for formal land tenure

3.1 Introduction

In the last fifty years the proportion of the African population living in urban areas has more than doubled (UN-HABITAT 2010). A common feature of this rapid urbanisation is the growth of unplanned settlements, characterized by high population density, insecure tenure and low levels of infrastructure and public services. Over 61% of sub-Saharan Africa's urban population lives in slum conditions (UN-HABITAT 2010). Most governments have been keen either to curb this informal growth or tackle existing settlements through slum upgrading, which usually entails infrastructure investment, land registration and the introduction of formal tenure instruments (Deininger, Augustinus, Enemark, and Munro-Faure 2010). However, such endeavors can be expensive (Woodruff 2001) and many governments rely on innate demand for formal land tenure to drive their titling programs. This approach is not always successful: as discussed in previous chapters, Tanzania's attempts to fully establish a system of formal tenure in its cities have largely failed, in part due to very low levels of demand for land titles on the part of urban landowners.

Another feature of rapid urbanisation has been the integration of some very diverse, segregated societies. This is especially true of Tanzania, which has one of the highest levels of ethnolinguistic fractionalisation in the world. Table 3.1 shows the average tribal

Figure 3.1: Ethnic fractionalisation in Tanzania



Notes: Author's calculations from 1993 Human Resource Development Survey
Figure shows distribution of cluster-wide Herfindahl Index across Rural and Urban clusters.

Herfindahl Index[1] defined over rural and urban clusters from the World Bank's 1993 Human Resource Development Survey, one of the few representative surveys which recorded data on tribal affiliation in Tanzania.

While the modal rural cluster is perfectly homogenous, with a Herfindahl Index of zero, the modal urban cluster is marked by an extremely high degree of heterogeneity, the end result of high levels of rural-urban migration. While there is already some evidence that such large scale desegregation is highly beneficial for outcomes like trust and quality of governance (Alesina and Zhuravskaya 2011), one might question how the movement of people from homogenous societies with similar customs and traditions (such as customary land tenure) to very mixed, heterogenous environments might affect their demand for state-led formalisation.

The focus of this paper is on the interaction between these two facets of urbanisation: does ethnic sorting hurt or hinder efforts to formalise unplanned settlements? To do this, I use data from a unique census of a large slum in Dar es Salaam to investigate whether or not households living near neighbours of the same ethnic background are less likely to

[1] The Herfindahl Index is a common measure of heterogeneity. For any setting with N groups, each with a different population share $s_1, s_2, \ldots s_N$, the Herfindahl Index is calculated as $1 - \sum_{n=1}^{N} s_n$, and can be interpreted as the probability that any two individuals selected at random come from *different* groups.

accept a limited form of land title recently offered by the Tanzanian government. The main challenge to identification comes from endogenous sorting: households can choose which parcels of land they wish to acquire, so unobserved household characteristics might be driving both the decision to locate with coethnics and the demand for formal tenure. I attempt to bypass these concerns by controlling for how ethnically-similar the household's neighbours were at the time the household moved into the slum, using variation in the ethnicity of neighbours who arrived at a later date to drive the result.

The results suggest that the impact of having neighbours from the same ethnic background on the demand for formal tenure is consistently negative, even when using different measures of 'coethnicity', different distance cutoffs for who qualifies as a neighbour, controls for the religious similarity as well as proxies for nearby relatives. The results are also robust to the inclusion of spatial fixed effects, introduced to control for any spatially-correlated unobservable characteristics. While the exact channels are difficult to identify, I argue that the negative coefficient is due to a variety of factors which reduce a household's expropriation risk when they are surrounded by coethnics, reducing the need for a formal title. This is supported by the data, which suggest that households living near coethnics are less fearful of losing their land.

This chapter makes several contributions to the literature. First, while many studies have considered the aggregate impact of ethnic heterogeneity on community or national-level outcomes, few have focused on decisions made at the household level with respect to the adoption of state-provided goods. Second, to my knowledge this is the first empirical study to show that differences in how households are ethnically sorted *within* a community can have a non-negligible impact on their acceptance of a formal system, a result which might be shrouded when only aggregate levels of ethnic heterogeneity are used. Finally, while many studies have suggested that rural-urban migration leads to a breakdown of social ties commonly seen in tribal networks (Lilleør and Lassen 2008), this work shows that these ties can still play a role in an urban setting.

In Section 3.2 I will discuss what we know from the literature about how coethnics interact with each other, the context of ethnicity in Tanzania and a framework for the hypothesis that households living with coethnics will be less likely to accept an offer of formal tenure from the government. In Section 3.3 I will discuss the data gathered from Dar es Salaam, the empirical model and construction of the relevant variables, as well as my main identification strategy. Section 3.4 will present and discuss the main results, Section 3.5 will cover robustness checks and I will conclude with Section 3.6.

3.2 Context and framework

3.2.1 Land rights in Tanzania

While rural land ownership and distribution in Tanzania have historically been dictated either by customary law or by the large scale 'villagisation' and collectivisation experienced under Julius Nyere's *Ujamaa* policies, urban land has officially been owned and allocated by the central government, itself initially hostile to the growth of informal settlements (Kironde 2006). Despite the government's position and *de jure* ownership of all urban land, informal acquisition by migrants continued nearly unabated in the years following independence, with *de facto* tenure arrangements being supported either through customary law or informal arrangements of ownership (De Soto and Cheneval 2006).

Much of this changed with the passing of the 1999 Land Act and Village Land Act, which restricted the mandate of customary law to rural villages, implicitly removing the government's recognition of customary land holdings in urban areas. While informal ownership in urban areas was acknowledged in this new legislation, formal recognition of tenure would only be made available to those who purchased a land title authorized by the Ministry of Lands.

While the previous chapter focused on CROs, which are effectively full land titles, this chapter will return the focus to residential licenses (known as a *leseni ya makazi* in Swahili). Residential licenses are short-term leases of urban land from the central government,[2] initially with a renewable two year duration, although this was later extended to five years. Seeing them as a stepping-stone to full title, the Tanzanian Ministry of Lands has focused most of its attention since the passing of the Land Act on getting urban

[2]Despite the fact that an RL is a leasehold, the central government generally does not allocate this land, nor does it claim it back after an RL expires. The expiration of a RL can be seen as an expiration of the government's *recognition* of ownership.

landowners registered with residential licenses.

Aside from the inconvenience of requiring a periodic renewal, RLs are limited in several other ways: they are non-transferable, so new owners must purchase a new RL, even if the previous owner already owned one. They were initially also seen as insufficient for obtaining credit: banks in Tanzania showed little interest in making loans on the basis of such a short term leasehold. However, recent anecdotal evidence has suggested that after RLs were extended to five years, many microfinance lenders began accepting them as collateral. Residential licenses are also expected to provide a degree of private tenure security. Not only do they provide households with a government document certifying them as the landowner, but they are also registered as residential license holders in the municipal government's database. Additionally, while the government's official policy on post-expropriation compensation is murky, by some accounts the residential license also guarantees the owner compensation after holding the title for three years or longer (Kironde 2006).

As opposed to more comprehensive forms of title, it is less likely that the purchase of an RL would result in direct externalities on neighbouring parcels. Due to the fact that a residential license has very short lifespan, it is less likely to factor into large-scale government decisions on slum-clearance or expropriation, where large-scale full-titling might have a more substantial impact. This means that the theoretical titling externalities discussed in the previous chapter are less likely to be at play here.

The base price for a RL is roughly \$6, yet one of the prerequisites to applying for one is the payment of two forms of taxation: land rent and property tax. A further implication of these requirements is that households who opt in to the RL system are easier to track and tax. This not only raises the price each household faces, but might also change the very nature of the purchase decision. Rather than weighing the direct benefit a residential license has in reducing expropriation risk versus the price, the decision to purchase an RL might be part of a grander bargain: allowing the government to extract tax in exchange for more public services (such as slum upgrading and tenure security). While I will largely present the trade-off between tribe and title as a simple story of expropriation risk, it is also likely that there are other, larger trade-offs at hand.

3.2.2 How do coethnics interact?

Much of what we know about how coethnics interact is derived from literature on the correlates of ethnic heterogeneity, which has suggested that coethnics have greater levels of trust (Alesina and La Ferrara 2002; Zerfu, Zikhali, and Kabenga 2008), which can have implications as far reaching as tax compliance (Lassen 2007). There is also some evidence, thanks primarily to the work of Habyarimana, Humphreys, Posner, and Weinstein (2007), that coethnics have a distinct advantage in reaching cooperative outcomes, either because they have a *strategy-selection* advantage (i.e. coethnics fall back on norms of cooperative strategies) or because they have a *technological* advantage in enforcing cooperative outcomes through mechanisms such as social sanctions.

This evidence is primarily confined to experimental settings, but the literature has broadly concluded that these mechanisms produce observable differences in real life collective action outcomes such as public goods provision, where coethnics are again seen to have a significant advantage in many settings (Alesina and Ferrara 2000; Miguel 2004; Miguel and Gugerty 2005; Algan, Hémet, and Laitin 2011), although this relationship is not quite ubiquitous (Glennerster, Miguel, and Rothenberg 2010).

Also, perhaps more closely related to the theme of this chapter, there is a growing body of work on the effect of migrating into an 'enclave', or ethnically-similar environment. Most of these studies are confined to Scandinavian countries preoccupied with immigration, many of them showing positive effects of coethnic sorting on labour market outcomes (Edin, Fredriksson, and Åslund 2003; Damm 2009). There has been little work on the effect of coethnic sorting on the demand for government-provided public goods, save for one study which finds no correlation between living in an enclave and support for the size of government (Gerdes 2011).

3.2.3 Ethnicity in Tanzania

There has been some debate over the extent to which ethnicity remains a salient issue in Tanzania, which saw universal adoption of Swahili as well as the firm establishment of a national identity in the years following independence (Polome 1980; Court 1984). The Tanzanian government has not gathered any census data on ethnic affiliations since 1967,

Table 3.1: Trust within and across tribes

Year	**Mean**		
	Trust tribe members?	Trust other tribe?	Diff (1-2)
2001	0.77	0.68	0.08***
2005	0.32	0.22	0.10***
Total	0.60	0.51	0.09***

Notes: Data taken from Afrobarometer Survey, Rounds 1 and 2.
Third column shows results of t-test of difference of means between first two columns.

considering the matter to be taboo. In the latest round of the Afrobarometer survey, 67% of respondents in Tanzania responded that they "feel only Tanzanian", when asked about their national and ethnic identities (compared to 20-24% of Nigerians, Ghanians, Malawians and Kenyans).[3] Miguel (2004) compared differences in ethnic heterogeneity and public goods provision between villages on either side of the Kenyan-Tanzanian border and found that ethnic heterogeneity only negatively impacted provision on the Kenyan side.

Despite the obvious strength of Tanzanian national identity, there are still signs that many urban Tanzanians readily recognise and identify with their ethnic origins. Respondents in our survey had no difficulty in identifying their tribe of origin, and over 50% of landowners considered themselves fully fluent in their tribal language. The local daladalas (private buses) often have indigenous phrases written on the back. There is also evidence that tribalism still affects trust: Table 3.1 displays some data from the first two rounds of the Afrobarometer survey in Tanzania, where respondents were asked if they trusted people from within their own tribe and from other tribes. While the differences between the two levels of trust are not massive, they are still significant. Finally, some recent work by Lilleør and Lassen (2008) has revealed a connection between tribal heterogeneity and social capital in Tanzania, the latter measured using remittances from family members who have migrated out (under the assumption that communities with higher levels of social capital are better able to enforce remittance compliance).

Tanzania has a high degree of ethnolinguistic fractionalisation, yet much of this diversity is confined to urban areas, as indicated before in Figure 3.1. The stark difference in

[3] Own calculations.

tribal heterogeneity between rural and urban areas is perhaps one of the main drivers of the dichotomy between rural and urban land policy: while the Village Land Act allowed homogeneous rural communities with similar practices to use customary law as a precedent, the Land Act ensured that urban areas were subject to formal, individualistic land rights.

3.2.4 The demand for formal land tenure

Given what we know about the way coethnics interact with one another, how might the presence of coethnics depress or reinforce the demand for formal land tenure?

For one, greater levels of trust with one's neighbours may lead to lower levels of perceived expropriation risk. Macours (2007) shows that Guatemalan landowners who lack the security of formal tenure are more likely to lease land out to coethnic tenants than non-coethnics. A household might also be better defended from challenges outside its immediate neighbourhood if it is surrounded by coethnics. In the absence of formal tenure arrangements, land disputes are typically settled locally (Kombe and Kreibich 2002), where the side with the larger cohort would have an advantage. Kombe and Kreibich (2002) describes how informal security of tenure in Dar es Salaam is highly dependent on local acceptance:

> "Like in Kihonda and the other informal settlements studied, social-recognition of an individuals rights on land by other settlers, especially the adjoining landowners, by local leaders and relatives or friends is the key factor guaranteeing security of tenure."

Durand-Lasserve (2003) refers to this mix of old-style customary tenure and informal systems as the "neo-customary" tenure system. Anecdotally, this social recognition appears to be crucial to informal tenure security: during field-work in the area of study, the research team interviewed a landowner who, returning from a ten-day trip, found that an undeveloped portion of his land had been sold off and that a house was already being built in its place. The landowner complained that his neighbours, with whom he had a poor relationship, had supported the sale in his absence.

Occasionally disputes make their way to the courts, which have sometimes deferred to customary law (Kironde 2006), where again numbers may play a role. One could also think of tenure as a 'club' good over which coethnics have distinct advantage of self-provision. Non-coethnics lack this advantage, and so if neighbour-enforced informal tenure and government-provided formal tenure are substitutes, non-coethnics should be more likely to embrace the latter.[4] Finally, while it may not have a direct impact on expropriation risk, coethnics are known to have an advantage in creating risk-sharing networks (Grimard 1997) and a household which loses its land might have a softer place to land if surrounded by coethnics.

The interaction between ethnicity and the demand for formal titling might not only be dependent on concerns over expropriation. As several studies have suggested that credit is often provided along ethnic lines (Fafchamps 2000; Biggs, Raturi, and Srivastava 2002; Fisman 2003), households which are ethnically-isolated may look to other solutions for borrowing, such as formalising their property to either use it as collateral (Besley 1995) or to signal their credit-worthiness to lenders (Dower and Potamites 2012). Similarly, if households obtain formal tenure as a means of paying for local public goods (both the residential licences discussed in this chapter and the CROs discussed in the previous one require annual payment of property tax), those from ethnically-fractionalised neighbourhoods may face a natural disadvantage in self-provision, so will be more likely to turn to the state to provide these goods.

There might also be 'negative' reasons why households living near coethnics would have lower demand for formal tenure. One common feature of kin networks is a cultural imperative to share wealth (Platteau 2000) and there is growing evidence that households will often hide a portion of their income for fear of ill-treatment from their kin (Jakiela and Ozier 2012). Investment in land, either directly or indirectly through tenure formalisation, might be looked down upon by one's neighbours if they happen to be kin. Similarly, some models of kin networks predict that ethnic groups are worried that members will exit the 'club' and migrate to the formal sector, thus undermining the whole network (Hoff and Sen 2006). To prevent this, they raise costly barriers to exiting, preventing households from

[4] Of course, informal and formal tenure arrangements can also be complementary, or even orthogonal, which would suggest either a positive or zero relationship with coethnic sorting.

leaving, or, in this context, embracing formal tenure. It is also possible that coethnics might actually be a larger threat to tenure security than outsiders: Midheme (2007) cites instances in the Dar es Salaam slum of Midizini where landowners shared land with kin, only to have their relatives attempt to register the land formally in their own name.

Note that while it is very difficult to empirically distinguish between these different mechanisms, they all result in the same basic prediction: that the presence of nearby coethnics should reduce a household's demand for formal land tenure. However, it is also possible that coethnics might push households towards formalisation, as land registration could act as a way for households to decouple from customary inheritance obligations, preventing kin from inheriting land they wish to pass on to someone else. Also, if coethnic neighbours are more likely to communicate and share information, information about land titles might spread quicker amongst these groups, promoting adoption through social interactions.

Finally, as discussed in the previous chapter, many of these theories assume that titling, on the whole, results in positive outcomes for adopting households. However, in certain contexts large-scale titling could make communities worse of if, for instance, it made it easier for governments to tax them without providing anything in return. For instance, there is anecdotal evidence that Dar es Salaam's residential license scheme failed because the government failed to fulfill its promise to reciprocate the increase in property tax contributions with systematic slum upgrading. This could leave communities unambiguously worse off. If the individual returns to titling are high enough to induce households to move to this "high-tax" state, then only substantial coordination amongst households might keep neighbourhoods from ending up in this equilibrium. Coethnics would likely have a distinct advantage in coordinating to avoid titling if large-scale adoption would make everyone worse off.

The rest of this paper will be concerned with testing whether or not having coethnic neighbours is associated with different levels of demand for formal land tenure, in the context of a slum in Dar es Salaam where residential license were recently introduced.

3.3 Data and empirical model

3.3.1 The Tanzanian Land Rights Survey

In the summer of 2010 the University of Oxford conducted a census of land parcels in two adjacent mitaa[5] (subwards) in the district of Kinondoni, Dar es Salaam. The two communities, Kigogo Kati and Mburahati Barafu, are unplanned, low-income communities located less than five kilometers from the center of the city.

The aim of the Tanzanian Land Rights Survey (TLRS) was to cover every parcel of land listed in the Kinondoni property register, a database of all presumed property owners constructed in 2004. The result of the survey is a sample of 2,384 individual parcels of land across the two mitaa. The survey data includes detailed information on the characteristics of the owning household and the parcel itself. Property owners were also asked about the tribal affiliation of every member of the household, which will provide a basis for measuring coethnicity in Section 3.3.4. Finally, the precise location of each land parcel can be identified using a series of land registry maps produced by the Kinondoni Municipality (the same maps used to produce the analysis in Section 2.7.2), which will allow me to identify each household's relevant neighbour set. First I will consider the empirical model I will use to examine the relationship between coethnic location and the demand for formal tenure.

3.3.2 Empirical model and framework

Following from the discussion in Section 3.2.4, there are a number of theoretical reasons why a household's demand for formal tenure might be reduced by the presence of coethnics in the neighbourhood. Consider a household's latent demand for a residential license, defined as a function of household characteristics $\mathbf{X}_i$ and the proportion of neighbours who are coethnic c_{i_N}. Given some simple assumptions over the structure of the error term, it is straightforward to re-write the household's demand equation as a linear probability model in which both c_{i_N} and $\mathbf{X}_i$ affect the probability of purchasing a residential license:

[5] Mtaa is the Swahili singular for subward, where mitaa is the plural.

$$RL_i = \gamma + \theta c_{i_N} + \beta \mathbf{X}_i + \delta \overline{\mathbf{X}}_{i_N} + L_i + u_i + \varepsilon_i \tag{3.1}$$

Where RL_i is an indicator variable equal to one if index household i has ever acquired a residential license for its parcel.[6]. Again, c_{i_N} is a measure of ethnic similarity between i and its relevant neighbour set N, both of which I will describe further in the next section.

The key parameter of interest is θ, the effect of coethnic neighbours on the propensity to adopt a residential license. If, as described in Section 3.2.4, the presence of coethnics reduces the demand for formal tenure, then estimates of θ should be negative. If titling decisions are independent of any spillovers from nearby coethnics, then any tests of θ should fail to reject a null hypothesis of zero. Finally, if coethnics generate a demand for tenure, either by pushing households away or through information peer effects, then θ might be greater than zero. While I will argue in this section that it will be possible to identify θ, neither a positive nor negative result would be enough to directly identify the channel through which coethnicity affects tenure demand. Thus, even if estimates of θ are strictly less than zero, we cannot discern whether or not this is due to complementarity in expropriation risk, for instance, or coethnics directly preventing their neighbours from signing up. However, in Section 3.4.2 I will investigate whether or not coethnics also have an effect on a few potential channels, such as expropriation risk and public goods expenditure, in an attempt to better identify the mechanisms through which coethnic affects tenure demand.

Returning to equation (3.1), $\mathbf{X}_i$ is a vector of parcel and owner household characteristics, the contents of which I will describe shortly. $\overline{\mathbf{X}}_{i_N}$ is the average of $\mathbf{X}_i$ across i's neighbour set. In the context of the peer effects literature, δ measures *contextual/exogenous* peer effects (Manski 1993). L_i is a vector of geographic characteristics for parcel/household i, which is intended to capture any *correlated effects*, another concern of the peer effects literature. For example, if coethnics are more likely to cluster in areas where tenurial investment incentives are low, such as in hazardous areas like flood plains, a negative estimate of θ may be the result of these common characteristics, rather

[6]Some households may have obtained a RL but then let it expire by the time of the survey. The questionnaire did not specify whether or not the household currently has a residential license

than coethnicity itself. The parameter u_i is a vector of unobserved household or land characteristics thought to be correlated with both RL_i and c_{i_N}. This is my chief concern for identification, which I will discuss further in Section 2.5.

Finally, despite my hints that the parameters in this specification have a distinct peer effects interpretation, such as the setup in Chapter 2, equation (3.1) explicitly excludes an *endogenous* peer effect. That is, the characteristics of household i's neighbours are allowed to affect i's residential license choice, but not their take-up, $\overline{RL}_{i_N}$. This is essentially a reduced-form version of equation (2.3), where any estimate of θ might comprise the net effect of coethnicity, operating through any observed endogenous peer effects, rather than just as a purely exogenous peer effect. While the previous chapter hinted that there might be positive endogenous peer effects in residential license take-up, the identification issues created by including them in this specification would distract from the main objective of identifying coethnic effects.

Next, I will discuss how the neighbour set and measure of coethnicity, the two chief components of c_{i_N}, are constructed.

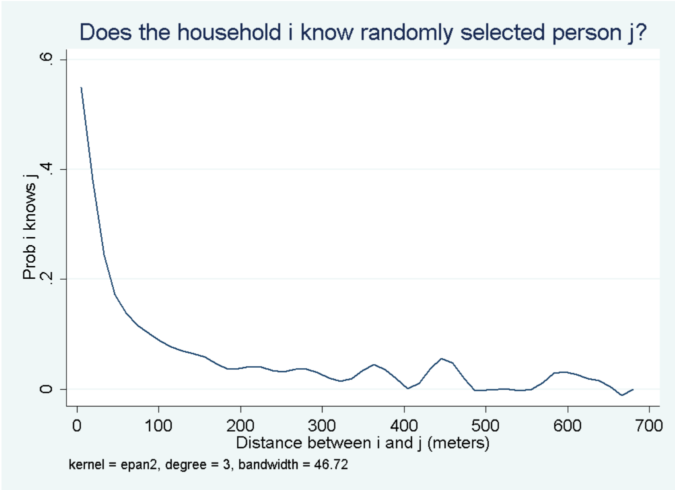
3.3.3 Choosing an appropriate neighbour set

There is already good reason, a priori, to believe that the relevant social network for a household is characterized by geographic proximity. Prior to the start of survey work, researchers randomly selected 10 households from each 'block' of parcels (these blocks were arbitrarily defined by the municipality) to frame a small network questionnaire.[7] During the survey, each respondent was asked if they knew any of the households from the list. Figure 3.2 shows the results of a kernel regression of the probability that household i knows household j on the distance in meters between i and j's parcels. The chance of a connection between i and j drops off sharply with distance, with the greatest drop occurring within the first 50 meters.

Not only are close neighbours more likely to know and interact with one another, but they may also have special status with respect to land tenure. The results from Chapter 2 already suggest that, when it comes to land title adoption decisions, households look to

[7] This is the same questionnaire used to discuss network effects of CRO adoption in Section 2.6.1

Figure 3.2: Distance and social connections



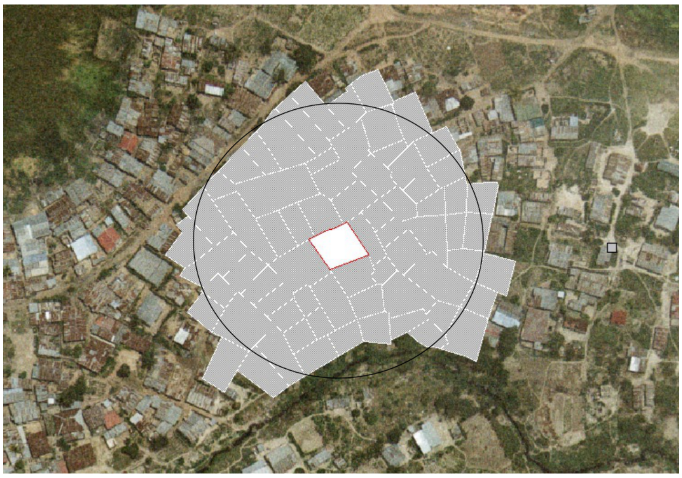
their spatially-proximate peers. While it is not a requirement for obtaining a residential license, other forms of land registration, such as the CRO, often require the signature of the four closest neighbours to verify ownership. Proximate neighbours also pose the biggest threat to tenure security, as they might be more likely to make ownership challenges than households further away.

I define the relevant neighbour set as every household/parcel with a border within a given distance d of the index parcel.[8] However, the choice of d is a challenge in itself. Intuition suggests that contiguous neighbours may be the most important. However, land acquisition in this area is characterized by subdivision, either through sales (I sell my front yard to a new arrival) or through inheritance (I cede my front yard to my newly-wed son or daughter). This suggests that existing owners have significant control over who moves directly next to them. Thus, while it is tempting to set $d = 0$ and use only contiguous neighbours in the analysis, this results in a neighbour set the household has very likely chosen, and thus concerns over correlation with u_i grow.

Setting d too high also runs the dual risk of allowing in too many neighbours who

[8]In Chapter 2 I used a nearest-neighbour approach instead of a distance band specification. As I explain later in the section, I will not be relying on variation in the ethnicity of contiguous neighbours due to concerns over endogeneity. To ensure that there is enough variation to identify an effect, I have used larger neighbour sets than I do in the previous chapter. Using very large nearest-neighbour sets can be troublesome, as the nearest nth neighbour can be very distant for some observations (such as a case where a cluster of $n-1$ households is isolated from the rest of the community). In this case, distance-band measures might be more reasonable proxies for the surrounding environment.

Figure 3.3: Example of neighbour set using 50m threshold



have little relevance to tenure security and trading off too much geographic variation in c_{i_N} for tribe-specific variation.[9] As a compromise, I set $d = 50m$, thus including all neighbouring parcels within 50 meters of the index household's borders. Figure 3.3 shows an example of the neighbour set for a randomly-chosen parcel in the sample. Fifty meters may appear to be an arbitrary choice, but it has the benefit of being a reasonable cutoff for two households knowing one another, as evidenced by Figure 3.2. Later on, in Section 3.5.3, I will consider how alternate values of d affect the main results.

3.3.4 Choosing coethnicity measures

The household roster questionnaire for the TLRS contains a single question which will form the basis for this measure: "what is your tribe?" Respondents were allowed to choose from an extensive list of tribes derived from *Ethnologue*, a language encyclopedia (Lewis 2009). If unhappy with the choices offered, their subjective response was recorded. 97% of responses were directly mappable to *Ethnologue*'s classification, with over 70 different tribes represented. Those that could not be matched with a known tribe or language group have been dropped from the analysis.

I feel that, within the Tanzanian context, tribal affiliation is a reasonably objective

[9] As d increases, c_{i_N} will eventually converge to a neighbourhood-level estimate. For example, if d is large enough to consider all other parcels in the neighbourhood, all households of the same tribe will have the same estimate of c_{i_N}.

measure of ethnic affiliation, circumventing many of the problems associated with self-identification measures. Despite the government's aversion towards recording ethnic data, "what is your tribe?" seems to be as straightforward and uncontroversial a question in Dar es Salaam as "where did you go to school?". The results reveal that the unplanned settlements are incredibly diverse: the Herfindahl Index for the sample is 0.91. Figure 3.4 illustrates this high degree of heterogeneity, with each tribe coded with a separate color.[10]

While a simple measure of coethnicity would be a tribal dummy equal to one if two households are from the same tribe and zero otherwise, this might be too restrictive a measure. Studies which use such a highly disaggregated definition of ethnicity have been open to criticism that they throw away crucial information on cultural similarity *between* ethnic groups (Desmet, Ortuño-Ortín, and Wacziarg 2012). For example, consider the language tree in Figure 3.5: using a simple tribal measure of coethnicity would assign a value of zero to both a Sagala-Asu and a Sagala-Ngulu pairing, even though the Sagala and Ngulu tribes share a common language branch and are presumably closer in many ways than the Sagala and the Asu.

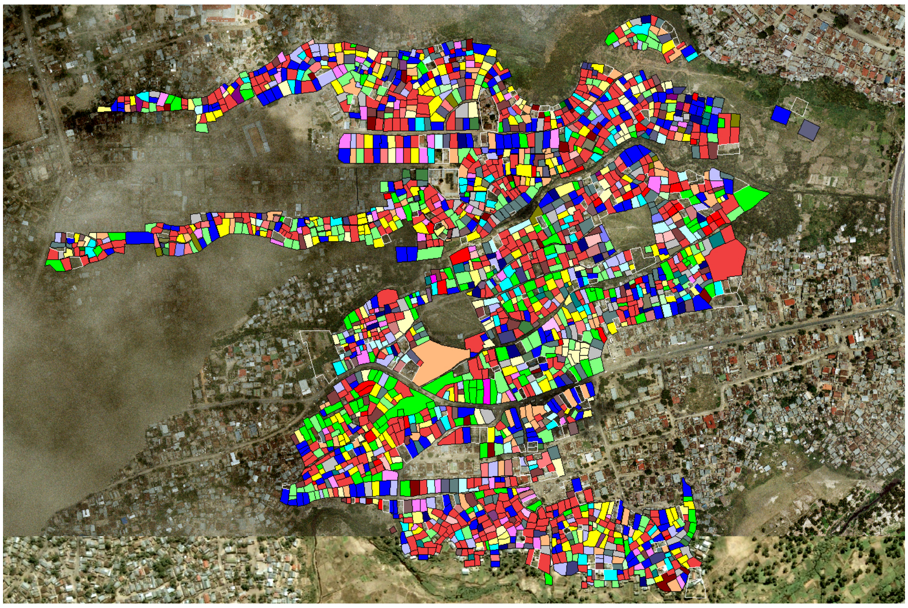
Using a more aggregate measure of tribe trades off variation in coethnicity with more accurate information on the linguistic differences between tribes. For example, aggregating to the highest level possible results in over 99% of the sample being relegated to a single language group (Niger-Congo). A reasonable trade-off (and the one I will suggest) is to aggregate up to the second row in Figure 3.5, to what I will refer to from now on as the *language stem*. This reduces the measured ethnic heterogeneity in the sample from a Herfindahl of 91% to 83% and the number of ethnic groups to thirty-two.

The coethnicity measure is derived as follows: each household i is paired with every other household in its neighbour set N. For every pairing, an indicator variable is constructed which is equal to 1 if both households are from the same language stem and 0 otherwise. This indicator is then averaged across the neighbour set, so c_{i_N} can be interpreted as the percentage of households in i's neighbour set from the same language stem.

Of course, the language stem might not be the most appropriate or relevant measure

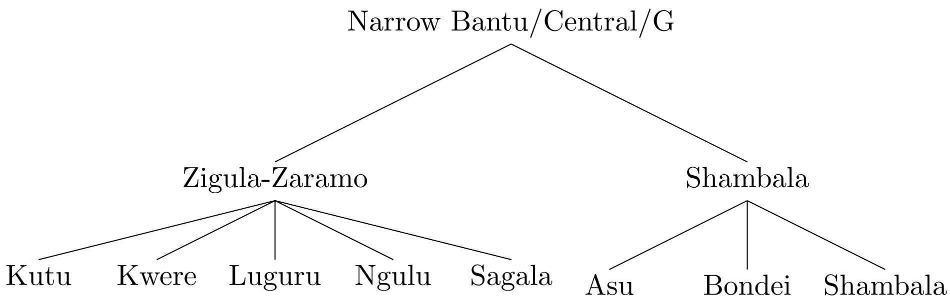
[10] Actual tribe names are not revealed here for purpose of anonymity.

Figure 3.4: Tribal heterogeneity of sample



Note: Each colour represents a different tribe

Figure 3.5: Language tree example



Note: Bottom level names indicate tribes, mid-level stems indicate language stems, and top-level indicates language group.

of ethnic similarity. While I will use it as my main measure for this paper, I also consider two other measures: one based on Fearon's (2003) index of 'cultural similarity', which uses *all* information from the language tree and the other using the geographic distance between tribal homelands as a proxy for ethnic similarity. Both of these will be discussed in more detail in Section 3.5 and in Appendix 3.A.2.

3.3.5 Identification strategy

The main threat to identification of θ, the impact of an ethnically-similar neighbour set on residential license uptake, is u_i: unobserved household level characteristics which drive both coethnic sorting *and* the demand for tenure. For example, more risk averse households might be more concerned about overall expropriation and thus will both seek extra protection through locating near other coethnic households and also have a higher innate demand for formal tenure, which would bias estimates of θ upward.

Similarly, there might be characteristics which drive coethnic sorting yet depress the demand for tenure. Imagine some households have an innate demand for village-based governance. These households would both desire to locate with coethnics and would have less enthusiasm for a state-backed land tenure instrument. If such preferences are at play, they might result in a spurious negative correlation between c_{i_N} and RL_i.

An ideal situation would be one where residential location had been randomly assigned. Unfortunately, while there are numerous examples of credible natural experiments in coethnic location from developed countries (Edin, Fredriksson, and Åslund 2003; Damm 2009; Algan, Hémet, and Laitin 2011), there are few studies in the East African context where urban location as been subject to a credible level of exogenous variation. Panel data would allow for household fixed effects, but this would also require enough variation in c_{i_N} across time, which might be difficult given the current density of the slum.

In lieu of the ideal, my attempt to solve this identification problem relies on data gathered on the timing of the household's arrival in the slum. Each household was asked about the year they acquired the parcel, allowing me to establish which households were already established when a given household i arrived in the neighbourhood. Define c_{i_B} as the ethnic similarity of all households in the neighbour set which were *already established* when the index household arrived.

Assuming that households only choose locations based on existing rather than future levels of neighbourhood coethnicity, then overall levels of neighbourhood coethnicity should be uncorrelated with unobserved determinants of coethnic location, conditional on c_{i_B}. To be precise, I assume:

$$cov(c_{i_N}, u_i | c_{i_B}, \ldots) = 0 \tag{3.2}$$

Intuitively, by including c_{i_B} as a control in equation (3.1), I will be comparing households which chose to locate next to similar numbers of coethnics, but experienced different numbers of coethnics moving in after they arrived.

Several more assumptions need to hold before (3.2) becomes plausible. Firstly, since we do not observe the actual makeup of the neighbourhood at the time of the households arrival, c_{i_B} needs to be be a good proxy for the surrounding environment at that time: there cannot be other, unobserved parcels occupied at the time which drove the household's location decision. The only model of slum formation which would guarantee this is one in which parcels are acquired either through squatting or subdivision and, once households arrive, they can never leave or move within the mtaa. While the latter requirement is harder to justify (some households own more than one parcel), movement is still infrequent. The former requirement, that households can never leave the mtaa, seems to be backed up by the data: roughly 94% of all parcels were undeveloped (lacking any sort of structures) at the time they were acquired. This suggests that parcels are empty until they are recorded in the survey as being acquired, and so c_{i_B} should be a good proxy for the ethnic makeup at the time household i arrived.

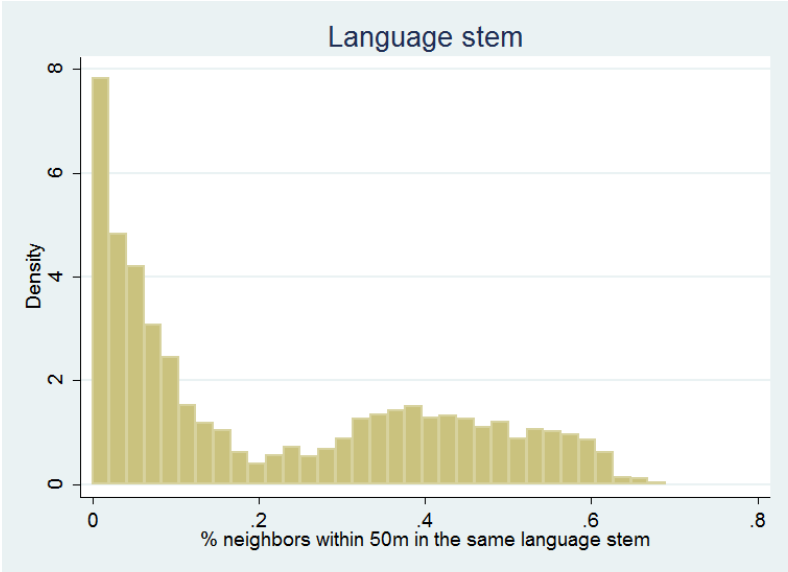
However, if expropriation risk is one of the main mechanisms at play, there might be concerns that selection out of the sample is still a problem if households are losing entire plots of land to families moving in. Anecdotally, it appears that most expropriation by other households is only partial, with households that have only developed part of their parcel losing the undeveloped portion to households migrating in, as described in Section 3.2.4. Thus, while expropriation remains a concern for households, it should not result in a sample selection problem, as households as households who have experienced expropriation should not actually drop out of the sample.

The second prerequisite for (3.2) to hold is that when choosing their location, households cannot anticipate future inflows of coethnics into their neighbour set. To guard against this, as well as recall error in the year of arrival, I extend c_{i_B} to include not only households which arrived before household i, but also five years after. In essence, all the remaining variation in c_{i_N} will be driven by households which arrived more than five years after i. To impose this restriction I must drop all households from the sample who lack neighbours that arrived more than five years later. This automatically includes all households arriving after 2004. This might have already been a desirable restriction: given that residential licenses were first made available in 2005, the location decision of households becomes more complex once coethnic location and residential license adoption becomes a simultaneous decision.

Finally, conditional on their original choice, households cannot encourage or discourage new coethnics into their neighbour set. While households do have some control over their contiguous neighbours through subdivision, this should be accounted for by the inclusion of all contiguous neighbours from the same tribe as a separate control. However this assumption might still be violated if households have some veto power over land sales in their immediate neighbourhood.

Aside from concerns over u_i, there may be tribe-specific characteristics which can drive both sorting and tenure. Tribes which originate near Dar es Salaam, such as the native Zaramo, may feel they have more secure informal tenure due to more proximate customary institutions and may be more likely to locate near each other due to lower moving costs. It is also important to disentangle sorting from overall size effects: households from tribes with a large presence in the mtaa are more likely to have coethnics in their neighbour set due to pure chance, and may feel more protected. Thus, it is important to control for as many tribe-specific characteristics as possible. I do this using a variety of characteristics (including the population size of each language stem in the neighbourhood) and eventually language-stem fixed effects.

Figure 3.6: Ethnic fractionalisation in Tanzania



3.3.6 Controls and summary statistics

Following from the empirical specification (3.1), my measure of coethnicity derived in sections 3.3.4 and 3.3.3 is defined as the percentage of neighbours in household i's neighbour set from the same language stem. Figure 3.6 shows the distribution of this measure across the sample. Note that no household has more than 70% of its neighbours from the same language stem.

For household characteristics I use household size, monthly income (TSh), the natural log of the household's asset holdings,[11] average age and schooling in the household and a dummy for Muslim households. I also include a proxy for how many households from the randomly-chosen network set the household knew and a measure of how familiar the household head is with his or her tribal language. For parcel characteristics, I have included the year the parcel was acquired, the log of the size (m^2) of the parcel and dummy variables for whether or not the parcel was inherited, the presence of a pit latrine, business or electricity connection, and whether any part of the parcel is being rented out. All of these characteristics are included in $\mathbf{X}_i$ as well as average values for the whole neighbour set $\overline{\mathbf{X}}_{i_N}$. Summary statistics for these characteristics are shown in Table 3.2.

[11]Households were only asked about the size, not the value of their assets. Median values for each asset type from the 2000/01 Tanzanian Household Budget Survey were used to calculate the total wealth stock.

Table 3.2: Summary statistics

Variable	Mean	(Std. Dev.)	Min.	Max.	N
Applied for RL	0.534	(0.499)	0	1	2247
% same language stem	0.218	(0.2)	0	0.690	2247
Household:					
HH size	5.181	(2.604)	1	20	2247
Log(assets)	14.426	(1.222)	8.313	17.957	2247
HH monthly income (tsh)	470.807	(1834.711)	0	58008.332	2247
HH avg years of schooling	12.293	(2.708)	2	24	2247
HH average age	29.797	(10.066)	9.333	88	2247
Muslim	0.573	(0.495)	0	1	2247
Language fluency	1.389	(1.323)	0	3	2247
# known in network roster	1.082	(1.345)	0	8	2247
Parcel:					
Year parcel aquired	1991.975	(11.908)	1960	2010	2247
Log(parcel size m^2)	5.335	(0.553)	3.462	9.095	2247
Parcel was inherited	0.141	(0.348)	0	1	2247
Toilet = pit latrine	0.77	(0.421)	0	1	2247
Electricity connection?	0.603	(0.489)	0	1	2247
Business on parcel?	0.077	(0.266)	0	1	2247
Parcel rented out?	0.394	(0.489)	0	1	2247
Ethnic:					
Dist of tribe homeland to Dar	341.048	(286.8)	29.521	2222.4	2246
Area (km) of tribe homeland	9723.972	(11571.419)	321.348	59972.592	2230
% of tribe in mtaa	0.083	(0.082)	0	0.217	2247
Patrilineal ranking of tribe	1.698	(0.87)	1	3	1982
Geographic:					
Kigogo Kati	0.585	(0.493)	0	1	2247
D to nearest church (m)	631.104	(272.392)	2.684	1239.563	2247
D to nearest field (m)	149.174	(99.430)	0	464.396	2247
D to nearest hazard (m)	428.279	(274.907)	0	961.464	2247
D to nearest mosque (m)	280.558	(148.568)	0.889	737.599	2247
D to nearest path (m)	129.572	(146.885)	0	756.814	2247
D to nearest primary road (m)	290.149	(177.591)	0	641.401	2247
D to nearest river (m)	156.348	(114.211)	0	510.516	2247
D to nearest road (m)	200.559	(144.589)	0	583.463	2247
D to nearest school (m)	228.682	(125.637)	0	578.339	2247
D to nearest mtaa office (m)	347.871	(161.188)	28.652	792.550	2247

For ethnicity-specific characteristics, I have included the total size of each language stem in the community, the geographic size of of each tribe's district of origin and its distance from Dar es Salaam[12] and measures for how patrilineal each tribe's descent system is.[13] These measures will be wiped out when I subsequently include language-stem fixed effects.

For the vector of geographic characteristics L_i, I include measures of the parcel's distance from a variety of geographic features, including the nearest church/mosque, fields, flood plains (hazard areas), walking paths, primary and secondary roads, schools, rivers and local government offices. I also include a dummy for whether or not the parcel is in Kigogo Kati or Mburahati Barafu. As mentioned before, I also control for the percentage of contiguous parcels who are from the same tribe (to account for the household's control over these parcels and within-family subdivision). Finally, I also control for the total size of the neighbour set N, to avoid conflating size with ethnic effects.

3.3.7 Inference

In estimation of (3.1), standard methods of statistical inference would suffice under the assumption of independence across observations. However, given the inherently spatial structure of this specification, this assumption is certain to be violated.[14]

A standard approach would be to make limiting assumptions about the correlation of error terms: for example, assign all parcels to geographic distinct districts and assume no correlation across these districts. However, the density of these unplanned settlements implies that such assumptions would be meaningless as many of these arbitrarily-defined districts would be adjacent, and observations along the boundaries of adjacent districts would be highly correlated.

To deal with this spatial dependence, I use a method derived by Conley (1999), where the estimated covariance matrix is adjusted to allow for an arbitrary spatial correlation between observations which declines as the distance between observations grows and is

[12]These measures are estimated using maps provided in *Ethnologue*.

[13]This measure is derived from the *Atlas of Precolonial Societies* (Müller, Marti, Schiedt, and Arpagaus 2000)

[14]It is assured by the design of the specification: as c_{i_N} increases, the probability that other households in i's neighbour set are from the same language stem, and thus have a similar value of c_{i_N} also increases.

zero beyond a defined threshold. I set this threshold equal to the neighbour set distance cutoff, which for the main specification is fifty meters. I report my main results using the Conley standard errors. Across all esimates θ, the Conley standard errors are uniformly larger and so these can be though of a reasonable upper bound for inference.

3.4 Results

In this section, I will report the results from the estimation of (3.1), first as a 'naive' regression, without concern over endogenous sorting, then with controls for the household's choice of neighbours at the time of arrival. Due to the large number of controls I will present only the coefficients of interest here, but extended versions of these tables are available in the appendix (Tables 3.11 and 3.12).

3.4.1 Main results

Table 3.3 reports the result from the naive OLS regression. The first column shows the results from a simple bivariate regression of RL_i on c_{i_N} and each subsequent column introduces additional sets of the controls described in Section 3.3.6. Columns (2) and (3) introduce household/parcel characteristics, averages for neighbour sets and location characteristics. Column (4) includes the tribal controls[15] and column (5) uses language-stem fixed-effects.

The first thing to note is the significant, negative correlation between the share of coethnic neighbours and residential license uptake. While the point coefficient tends towards zero as household, contextual and location characteristics are included, introducing tribal controls or tribe fixed effects results in a stronger negative relationship. This suggests that some tribes or language groups are more likely to sort together *and* obtain a residential license.[16]

The point coefficients in these naive regressions are not only significant, but econom-

[15] Data for some tribes in unavailable, hence the drop in sample size.

[16] Through repeated randomization of observed location, it is possible to calculate the expected value of c_{i_N} under the assumption of random assignment. Language stems with a large positive deviation from this value ($c_{i_N} - E[c_{i_N}]$) are observably more likely to purchase a residential license (the correlation coefficient between these two is 0.42).

Table 3.3: Ethnic co-location and residential license takeup

	(1)	(2)	(3)	(4)	(5)
% same stem (50m)	-0.253***	-0.242***	-0.210***	-0.348***	-0.380***
	(0.0535)	(0.0611)	(0.0611)	(0.107)	(0.146)
HH char		Yes	Yes	Yes	Yes
Parcel char		Yes	Yes	Yes	Yes
Neighbor char			Yes	Yes	Yes
Location char			Yes	Yes	Yes
Tribe char				Yes	
Tribe f.e.					Yes
R^2	0.010	0.079	0.127	0.126	0.139
Obs	2247	2247	2247	1978	2247

Conley standard errors in parentheses: $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Variable of interest is the percentage of other hhs within 50m from the same language stem

Dependent variable is a dummy for RL takeup

ically meaningful. A 10% increase in the percentage of neighbours from the same stem results in a 3.8% reduction in the predicted probability a household has obtained a residential license. Moving from being completely surrounded by neighbours of a different language stem ($c_{i_N} = 0$) to the maximum observed percentage ($c_{i_N} = .7$) results in roughly a 27% reduction.

Table 3.4 reports results once c_{i_B}, the percentage of neighbours who arrived five years after *or* earlier than household i who are coethnics is included as a control.[17] The first column replicates column (6) from Table 3.3. The second column restricts the sample to all households which arrived prior to the year 2005. The third column further restricts the sample to all households which have any neighbours arriving more than five years later. Note that the estimate of θ is lower for this restricted sample.

Keeping the sample size constant, column (4) reveals the estimate of θ once c_{i_B} is included as a control. The result is a further downward movement in θ^*. Despite a decrease in precision, the point estimate is still highly significant. A 10% increase in the percentage of coethnics in a household's neighbour set is now associated with nearly a 6% decrease in the predicted probability the household has obtained a residential license.

[17] This includes households which arrived in the same year as the index household.

Table 3.4: Results controlling for previous coethnic choice

	(1) OLS	(2) t < 2005	(3) Restricted	(4) Final
% same stem (50m)	-0.380***	-0.434***	-0.468***	-0.576***
	(0.146)	(0.165)	(0.171)	(0.219)
% same stem at arrival				0.0972
				(0.123)
HH char	Yes	Yes	Yes	Yes
Parcel char	Yes	Yes	Yes	Yes
Neighbor char	Yes	Yes	Yes	Yes
Location char	Yes	Yes	Yes	Yes
Tribe f.e.	Yes	Yes	Yes	Yes
R^2	0.139	0.141	0.142	0.142
Obs	2247	1886	1784	1784

Conley standard errors in parentheses: $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$
Column 1 is the full OLS model from the previosu table
Column 2 restricts the sample to parcels acquired prior to 2005. **Column 3** further restricts the sample to those who who had neighbors move in more than 5 years later.
Column 4 introduces a control for the percentage of % of neighbors at arrival (and within 5 years later) from the same language stem.
Dependent variable is a dummy for RL takeup

Moving from complete isolation to the maximum saturation seen in the data ($c_{i_N} = 0.70$) results in a 40% decrease in the probability of having obtained an RL. While it is not significant at standard levels of inference, the positive coefficient (0.0972) on c_{i_B} indicates that households which moved into ethnically similar areas are actually more likely to purchase a CRO.

3.4.2 Channels

While the results in the previous sections suggest a persistently negative partial correlation between how ethnically similar a household's surrounding environment is and its demand for a residential license, the underlying mechanisms are not yet obvious. As discussed in section 3.2.4, the presence of coethnic neighbours may reduce a household's perceived expropriation risk, thus reducing the need to buy in to the formal tenure system. However, there might be other factors at play here: there is some evidence that landowners in Dar es Salaam believed that obtaining an RL would give them access to credit (Midheme 2007), and anecdotal evidence suggests that, after RLs were extended to a five year period, many formal lending organisations began accepting them as a form of collateral. As there is already some evidence that informal credit is often provided along ethnic ties (Fafchamps 2000; Biggs, Raturi, and Srivastava 2002; Fisman 2003), households with preexisting connections might put less value on residential licenses. Furthermore, informal land tenure might be part of a range of other club goods provided by coethnics. Given the tax burden implied by RL ownership, households isolated from their ethnic peers may be more likely to adopt because they desire access to state-provided public goods which they cannot obtain in their immediate neighbourhood.

As households were interviewed well after their decision to purchase a residential license, it is impossible to directly identify the channel through which ethnic sorting is depressing demand for tenure. However, the reduced-form 'impact' of having coethnic neighbours on household beliefs and behaviours might still suggest which channels might be active, even following the residential license choice. To test this, I consider three different measures of the expropriation, credit and public goods channels.

The first is perceived expropriation risk: during the baseline survey, households were asked to estimate the probability that they would lose their land within the next five years. My second outcome measure is a dummy for whether or not the household has borrowed, from any source in the past year.[18] Finally, while we do not observe any direct evidence of a household's consumption of public or club goods, we do observe a household's contribution levels to local (mtaa-level) public goods,[19] so the dependent variable in this specification is a dummy equal to one if a household has ever contributed to any public good in the past year (only 30% of the sample have done so). If households can choose between goods self-provided with their immediate neighbours and higher levels of public goods provision, we should see a negative relationship between having coethnic neighbours and mtaa-level public goods provisions.

[18]The results are similar using measures of perceived ability to borrow rather than actual borrowing behaviour.

[19]This includes public toilets sewerage, garbage collection, recycling, road building or repair, street lights, neighbourhood security and a local infrastructure upgrading project.

Table 3.5: Coethnic location and possible channels of tenure demand

	Expropriation risk		HH has borrowed		Public goods?	
	(1)	(2)	(3)	(4)	(5)	(6)
% same stem (50m)	-0.230**	-0.260*	0.0189	0.0662	-0.00320	0.0197
	(0.0967)	(0.136)	(0.0758)	(0.0947)	(0.149)	(0.200)
% same stem at arrival		0.0274		-0.0431		-0.0209
		(0.0830)		(0.0541)		(0.116)
Standard controls	Yes	Yes	Yes	Yes	Yes	Yes
Tribe f.e.	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.117	0.117	0.134	0.134	0.085	0.085
Obs	1715	1715	1715	1715	1715	1715

Conley standard errors in parentheses: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Dependent variables: (i) the perceived probability of expropriation in the next five years, (ii) indicator variable = 1 if the household has borrowed in the past five years, and (iii) indicator variable = 1 if the household has contributed anything to neighborhood public goods in past year

Table 3.5 shows the results of regressing these three measures on c_{i_N} and the other controls listed in Section 3.3.6. The results suggest that perceived expropriation risk is somewhat mitigated by the presence of coethnic neighbours: a 10% increase in the share of neighbours from the same language stem is associated with roughly a 2.5% decrease in perceived expropriation risk. Both the propensity to borrow and contribute to local public goods appear to be effectively orthogonal to c_{i_N}. While this does not guarantee that expropriation risk is the main channel through which coethnic neighbours suppress the demand for RLs, it appears to be at least one plausible connection. While a household's residential license status is not included in these specifications, the results are robust to its inclusion.

3.5 Robustness checks

3.5.1 Ethnic versus religious or family enclaves

While this paper has focused on ethnicity as being the most pertinent social dimension to investigate, it is possible that religion might also play a role. These two are somewhat correlated, as many of the 'indigenous' tribes from Tanzania's east coast have strong roots in Islamic practice, while migrants from the west tend to be from tribes with a predisposition towards Christianity. Historically, religion has been a more significant social cleavage in Dar es Salaam's local politics, although much of this has been fairly conspicuous racial politics (i.e. the indigenous population at odds with migrants from South Asia) (Brennan 2007).

To reaffirm that the results in the previous sections are not conflating the two types of identity, I have re-run the main specification, including the percentage of households in the neighbour set with the same religion (Christian or Muslim) as the index household. The results are presented in Table 3.6 in columns (1) and (2). θ^*, the coefficient on c_{i_N} is of a slightly larger magnitude once religious similarity is controlled for. The coefficient on religious similarity is insignificant, but positive. This confirms not only that religious heterogeneity is not driving the main results, but that it is, at best, a predictor of uptake.

Finally, it might still be the case that ethnic similarity is just acting as a proxy for kinship, with the measured effect running through direct family ties rather than indirect ethnic channels. To account for this, I use roster data from the baseline survey to match households with similar name structures. I isolate the middle and last names for every member of the household, convert them to a common character set using the Stata 11 command `Soundex`. This allows me to isolate similar-sounding middle and last names without relying on identical spelling. I then calculate the percentage of neighbours who have at least one household member with a similar middle or last name to at least one member of the target household. Columns (3) and (4) in Table 3.6 show that the inclusion of this variable has no substantial effect on the results here, suggesting that they are not being driven by nearby relatives.

What about channels? While in Section 3.4.2 I showed that coethnic location was correlated with perceived expropriation risk and not credit access or contributions to public goods, it is also worth investigating whether these alternate measures of neighbour similarity upset that result. Table 3.13 in the appendix replicates the main specification from Table 3.5, this time including both the percentage of same-religion neighbours and same-name neighbours as controls. Again, the percentage of coethnic neighbours appear to be correlated with a reduction in perceived expropriation risk, while neither the name or religion measures are significantly correlated with any of the three outcomes.

Table 3.6: Religious similarity, kinship and RL uptake

	(1)	(2)	(3)	(4)
% same stem (50m)	-0.488***	-0.594***	-0.474***	-0.585***
	(0.172)	(0.219)	(0.171)	(0.218)
% same religion	0.128	0.128		
	(0.0931)	(0.0932)		
% same stem at arrival		0.0956		0.100
		(0.122)		(0.122)
% same name			0.163	0.167
			(0.170)	(0.170)
Standard controls	Yes	Yes	Yes	Yes
Tribe f.e.	Yes	Yes	Yes	Yes
R^2	0.143	0.143	0.142	0.143
Obs	1784	1784	1784	1784

Conley standard errors in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$
Dependent variable is a dummy for RL takeup

3.5.2 Measuring coethnicity

Given that my main measure of coethnicity is constructed, it is possible the results presented in previous sections are unique to this definition. In this section I will present results from the estimation of (3.1) with two alternate measures of coethnicity.

The first is Fearon's (2003) measure of cultural similarity. A detailed description of the construction of this measure can be found in Appendix 3.A.2. Fearon's measure is useful as it utilizes information on how many branches of the language tree two language groups share, rather than just evaluating whether or not two language groups share a particular branch. Using this approach, c_{i_N} now becomes a measure of the average cultural similarity between household i and its neighbour set N.[20] Similarly, c_{i_B} is also calculated using this new measure.

Table 3.7 displays the results from this exercise. Columns (1) and (2) shows the bivariate and full-controls specifications respectively.[21]

[20]For each household-neighbour pair, Fearon's measure is calculated, then this is averaged across the entire neighbour set.

[21]Note that 27 observations have been dropped from this specification. The measure of cultural similarity results in a minority of tribes having close to zero similarity with their neighbours (these are typically tribes/language groups from outside Tanzania) with the rest of the sample having values between 0.6 and 0.95. Those below 0.6 are excluded from this specification.

Table 3.7: Alternate specification - average cultural similarity

	(1) OLS	(2) OLS	(3) t < 2005	(4) Restricted	(5) Final
Avg cultural similarity (50m)	-1.330***	-1.301**	-1.728**	-1.835**	-1.718*
	(0.357)	(0.653)	(0.748)	(0.762)	(0.906)
Avg similarity at arrival					-0.165
					(0.673)
Standard controls	No	Yes	Yes	Yes	Yes
Tribe f.e.	No	Yes	Yes	Yes	Yes
R^2	0.007	0.156	0.165	0.169	0.169
Obs	2220	2220	1866	1764	1764

Conley standard errors in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Dependent variable is a dummy for RL takeup

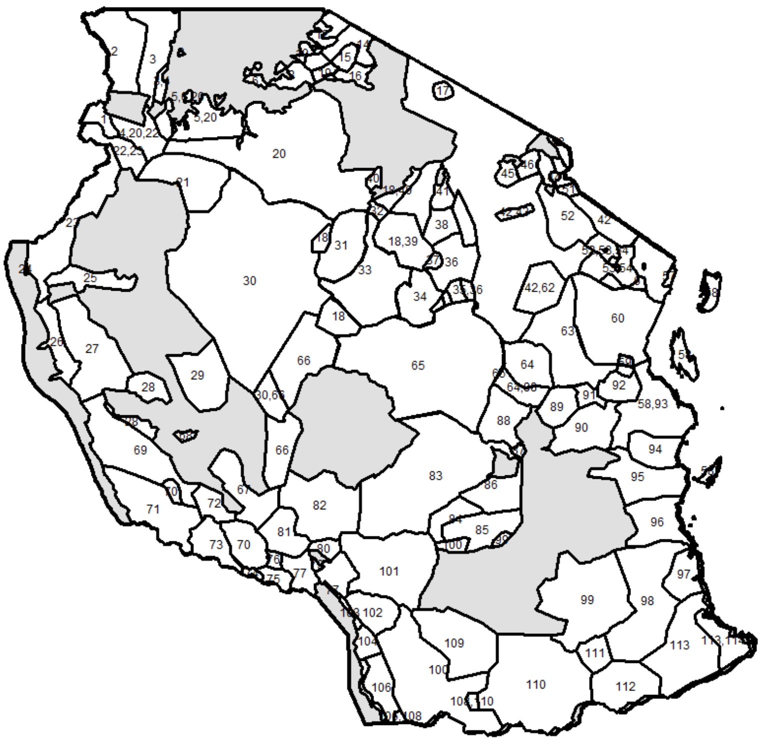
Table 3.8: Alternate specification: average tribal distance

	(1) OLS	(2) OLS	(3) t < 2005	(4) Restricted	(5) Final
Avg tribal distance	0.0212***	0.0496	0.0677**	0.0703**	0.0850*
	(0.00807)	(0.0313)	(0.0340)	(0.0353)	(0.0454)
Previous C	No	No	No	No	Yes
HH char	No	Yes	Yes	Yes	Yes
Parcel char	No	Yes	Yes	Yes	Yes
Neighbor char	No	Yes	Yes	Yes	Yes
Tribe f.e.	No	Yes	Yes	Yes	Yes
Location char	No	Yes	Yes	Yes	Yes
R^2	0.003	0.156	0.163	0.167	0.168
Obs	2230	2230	1875	1771	1771

Conley standard errors in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Dependent variable is a dummy for RL takeup

Figure 3.7: GIS language map of Tanzania (constructed using maps from *Ethnologue*)



Columns (3) and (4) make the same restrictions found in Table 3.4, where the sample is restricted first by year of arrival, then to those that have neighbours arriving more than five years later. While the inclusion of c_{i_B} results in a less precise estimate of θ, which is only significant at the 10% level, the estimate itself does not change much, and is still negative.

My other measure of ethnic similarity uses maps provided in *Ethnologue* detailing the approximate geographic origins of each language group in Tanzania (Lewis 2009). Figure 3.7 shows the map, which not only shows locations, but also the size of these tribal 'regions'. To approximate the exact coordinates and sizes of these territories, I overlaid *Ethnologue*'s map on a geographic information system (GIS) map of Tanzania, then calculated the average distance, in spherical coordinates, between each tribal region.

The result is a measure of ethnic similarity which captures actual physical distance between the tribes, which may capture cultural differences not covered in purely linguistic measures. Table 3.8 shows the results when c_{i_N} is recalculated using the average tribal distance between a household and its neighbours. While a lack of precision has knocked significance to the 10% level once restrictions are made, the results are still strikingly similar to that of the previous specifications, with the point estimate growing larger in magnitude with the inclusion of c_{i_B}. Note that because c_{i_N} now measures distance, rather than similarity, the point coefficients are reversed in sign.

3.5.3 Different neighbour sets

In this section I will consider whether changing the value of d substantially affects the main results. Table 3.9 shows estimates of θ when d is set at both 25 and 75 meters respectively, both before and after c_{i_B} is included as a control. Accordingly, the cutoffs used in calculating the Conley standard errors are set at the same distance d in each specification. Observation counts are different for the two specifications, as these samples are restricted to household that experience changes in their neighbour set after five years. When larger geographic cutoffs are used, each household as a higher probability of having new neighbours move in, hence the 75m specification has a larger sample.

While the coefficients presented in this table are broadly similar to the main specification, their precise values are not immediately comparable. Recall that the original estimate of θ could be interpreted as "the effect of an increase in the share of coethnic neighbours within 50 meters." Using larger (smaller) values of d allows more (less) neighbours into the neighbour set, so a 50% increase in the share of coethnics would result in a larger increase in the *number* of coethnic neighbours for larger values of d. Thus, we might expect estimates of θ to be larger, as they correspond to the impact of a larger number of people. However, as d increases, the neighbours it brings in are further away and so are probably less relevant to the household's decision to purchase a land title, which would mitigate the upward pressure on θ.

Despite this, the results seem to be robust under these perturbations of d, as Table 3.9 shows significant, negative effects for both cutoffs.

3.5.4 Spatial fixed effects

Despite the multitude of controls for land quality and geographic characteristics (listed in Section 3.3.6) which were included in the previous specifications, it is still possible that

Table 3.9: Main specification using different distance cutoffs

	25m		75m	
	(1)	(2)	(3)	(4)
% same stem within distance d	-0.268**	-0.334**	-0.432**	-0.725***
	(0.128)	(0.155)	(0.206)	(0.280)
% same stem at arrival		0.0595		0.269
		(0.0796)		(0.180)
Standard controls	Yes	Yes	Yes	Yes
Tribe f.e.	Yes	Yes	Yes	Yes
R^2	0.144	0.144	0.146	0.147
Obs	1628	1628	1819	1819

Conley standard errors in parentheses, cutoffs used set to 25 and 75 respectively.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Dependent variable is a dummy for RL takeup

coethnicity is correlated with unobserved geographic characteristics which drive residential license take-up. Recall the original empirical residential license adoption equation:

$$RL_i = \gamma + \theta c_{i_N} + \beta \mathbf{X}_i + \delta \overline{\mathbf{X}}_{i_N} + L_i + u_i + \varepsilon_i$$

If u_i contains neighbourhood specific effects which are also correlated with c_{i_N}, estimates of θ will be biased, even after controlling for the household's decision to sort with coethnics. This would be a problem if, for example, groups of coethnics reside in areas with poorer land quality.[22] While many studies are able to overcome this by including some sort of geographic fixed effects, the density of the households in this sample make it difficult to identify distinct areas where assumptions about the independence of unobserved characteristics are likely to hold.

One approach is to assume that unobserved heterogeneity in land quality is similar over spatially-proximate households, identifying the effect of interest off of the spatial discontinuity in residential license adoption and in coethnicity. Ignoring household-level unobservables for a moment, this assumption implies that for two observations, i and j, within some critical distance, $u_i = u_j = u$. This approach was first introduced in

[22]In an agricultural setting, quality often refers to unobserved soil characteristics. In this urban setting, where very little agriculture is performed, it comprises a range of attributes which factor into the household's valuation of land, such as geographic distance to infrastructure and services, risk of flooding or hazard, etc.

Goldstein and Udry (2008) and Conley and Udry (2010) to control for unobserved soil characteristics, and has been used by several studies to control for unobserved spatial heterogeneity (Ayalew Ali and Goldstein 2011; Magruder 2012a; Magruder 2012b). In this 'spatial fixed effects' (SFE) specification, the original empirical equation (3.1) is transformed by subtracting the average values for spatially-proximate neighbours:

$$RL_i - \sum_{j \in R(i)} \frac{RL_j}{N_p} = \theta \left(c_{i_N} - \sum_{j \in R(i)} \frac{c_{j_N}}{N_p} \right) + \beta \left(\mathbf{X}_i - \sum_{j \in R(i)} \frac{\mathbf{X}_j}{N_p} \right) + \delta \left(\overline{\mathbf{X}}_{i_N} - \sum_{j \in R(i)} \frac{\overline{\mathbf{X}}_{j_N}}{N_p} \right) + \left(L_i - \sum_{j \in R(i)} \frac{L_j}{N_p} \right) + \left(u_i - \sum_{j \in R(i)} \frac{u_j}{N_p} \right) + \left(\varepsilon_i - \sum_{j \in R(i)} \frac{\varepsilon_j}{N_p} \right) \quad (3.3)$$

Where $R(i)$ is a set of neighbours within a critical distance of household i and N_p is the number of neighbours within that critical distance. For the sake of simplified notation, let $y_i - \sum_{j \in R(i)} \frac{y_j}{N_p} = \triangle y_{ij}$. If the above assumption over u_i holds, then the term $\triangle u_{ij}$ should vanish from the above equation. More precisely, for θ to be identified, then $\triangle c_{ij_N}$ must be uncorrelated with $\triangle \varepsilon_{ij}$, conditional on the other transformed controls.

How likely is this assumption to hold? It is plausible, given the geographic and parcel-level controls already included in the main specification, that the SPE specification would successfully difference out any remaining geographic unobservables, so long as spatially-proximate households share a similar environment. However, u_i contains not only parcel/geographic-level unobservables, but also unobserved household characteristics which may be correlated with c_{i_N}. If households sort on these unobservables, then proximate neighbours will share similar characteristics, and it is likely that these will also be differenced out. Yet, if households primarily choose their coethnic mix c_{i_N} due to some unobserved preference also correlated with residential license take-up, and this preference is not continuous over space, then spatially-proximate difference in coethnicity $\triangle c_{ij}$ might still reflect differences in preferences, even after applying the SFE estimator. Thus, while this approach provides another opportunity to reject the null hypothesis that the identification assumptions in previous sections are correct, it is unlikely that it convincingly deals with the problem of endogenous sorting.

Table 3.10: Main results, spatial fixed effects specification

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	OLS	OLS	OLS	t < 2005	Restricted	Final
% same stem (50m)	-0.219***	-0.163**	-0.173***	-0.292*	-0.397**	-0.449**	-0.471**
	(0.0561)	(0.0636)	(0.0637)	(0.162)	(0.178)	(0.184)	(0.229)
% same stem at arrival							0.0201
							(0.122)
HH char		Yes	Yes	Yes	Yes	Yes	Yes
Parcel char		Yes	Yes	Yes	Yes	Yes	Yes
Neighbor char			Yes	Yes	Yes	Yes	Yes
Location char			Yes	Yes	Yes	Yes	Yes
Tribe f.e.				Yes	Yes	Yes	Yes
R^2	0.007	0.074	0.081	0.094	0.093	0.098	0.098
Obs	2247	2247	2247	2247	1886	1784	1784

Columns 1-4 replicate the results shown in Table 3.3 using spatial differencing (spatial fixed effects) **Column 5** restricts the sample to parcels acquired prior to 2005. **Column 3** further restricts the sample to those who who had neighbours move in more than 5 years later. **Column 4** introduces a control for the percentage of neighbors at arrival (and within 5 years later) from the same stem. Dependent variable is a dummy for RL takeup.
Conley standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.10 shows the results of applying the spatial fixed effects estimator to the main specification presented in Tables 3.3 and 3.4. Cutoffs for the SFE difference set $R(i)$ are set equal to 50m, the same as used for calculating c_i and $\overline{\mathbf{X}}_{i_N}$. The structure mimics that of the previous tables, with the first four columns gradually introducing several controls, the fifth and sixth column making sample restrictions before introducing a control for the coethnic choice in column (7). The results show a negative, significant impact of coethnic residence on residential license take-up, although they are of slightly lower magnitude (roughly 80% of the original point estimates).

3.6 Conclusion

Despite a deceleration in rates of urban growth in the last few decades, the rise of unplanned settlements in many cities remains a conundrum for policymakers. While providing landowners with recognised tenure is seen as one of several tools for ushering these settlements into formality, the success of demand-driven approaches such as Tanzania's will hinge upon the ability of policymakers to understand what drives this demand.

To take a small step in that direction, I have examined the effect of living with coethnic neighbours on the demand for residential licenses, one of the two most recent urban tenure instruments offered by the Tanzanian government. Despite controlling for a number of factors which might have determined both a household's coethnic sorting and the demand for tenure, there is a robust negative relationship between the two. The results also seem reasonably robust to small changes in the definition of coethnic and in neighbour, and do not appear to be proxying for other types of social connections, such as religion or direct kinship.

While the result that small ethnic enclaves might have a depressed demand for formal tenure might be disconcerting to those concerned with formalisation, there are a number of caveats which should be taken into account. Firstly, the effects revealed here are not the main drivers of the residential license program's lacklustre performance: the most insulated households exhibit over 30% less demand than the most isolated, yet most households are somewhere in between. While clustering is associated with lower demand, the intense and unchecked fractionalisation of the slum seems to have prevented these neighbourhoods from becoming too clustered. In this light, heterogeneity might complement formalisation, as a fractured population which is unable to self-provide crucial public goods will be more likely to accept state provision. It is unclear, however, whether such fractionalisation is more desirable from a welfare perspective.

Next, I have only presented a trade-off between ethnic ties and a single tenure instrument, one which is particularly weak along many dimensions (collateral value, transferability, durability) and so may be a particularly poor substitute for the protection afforded by those of a similar cultural background. This has been a significant concern in the academic literature on land tenure, that formal programmes focused on title and registration strip away many of the desirable, flexible features of informal/cutomary tenure systems (Platteau 2000; Udry 2012). Discerning which dimensions matter most for substitutability should be of particular concern here, and might be a useful path for future research. More substantive forms of tenure, such as the prohibitively expensive certificate of right of occupancy, may be strictly superior to residential licenses along the dimensions which matter most to residents of unplanned settlements.

It should be emphasised that the results in this chapter are highly context-dependent. As discussed before, tribal affiliation does not appear to have the same salience in Tanzania as it does in many neighbouring countries. The general lack of ethnic sorting in Dar es Salaam stands in stark contrast to countries where ethnicity is more polarising, such as Kenya.[23] While ethnicity does appear to matter for formal tenure decisions in Dar es Salaam, it is possible that this relationship would become more complex in contexts where groups are more polarised.

Finally, this work reveals that, while overall levels of ethnolinguistic heterogeneity are still important, there is still much to be learned about how heterogenous societies transition from a framework of segregation and self-provision to one of integration and state-provision.

[23] A brief analysis of ethnic sorting using geo-referenced data from Kibera, Nairobi's largest slum, reveals both lower levels of heterogeneity (a Herfindahl Index of 0.6639) and stronger clustering of ethnic groups

3.A Chapter 3 Appendix

3.A.1 Additional tables/figures

Table 3.11: Ethnic co-location and RL take up, full results

	(1)	(2)	(3)	(4)	(5)
% same stem (50m)	-0.253***	-0.242***	-0.210***	-0.348***	-0.380***
	(0.0535)	(0.0611)	(0.0611)	(0.107)	(0.146)
% adjacent in same tribe		0.100*	0.0646	0.0346	0.0667
		(0.0565)	(0.0546)	(0.0586)	(0.0549)
# of neighbors (50m)		0.00298***	0.00202*	0.00260**	0.00205*
		(0.000934)	(0.00118)	(0.00128)	(0.00120)
HH size		0.0147***	0.0131***	0.0139***	0.0141***
		(0.00493)	(0.00480)	(0.00533)	(0.00486)
Log(assets)		-0.00475	0.00449	0.00401	0.00463
		(0.00991)	(0.00954)	(0.0100)	(0.00948)
Monthly income (tsh)		-0.00000414	-0.00000326	-0.00000343	-0.00000387
		(0.00000346)	(0.00000347)	(0.00000375)	(0.00000345)
Avg years of schooling		0.00788*	0.00664	0.00653	0.00605
		(0.00420)	(0.00409)	(0.00425)	(0.00421)
Average age		0.000332	0.00131	0.00152	0.00149
		(0.00130)	(0.00129)	(0.00138)	(0.00127)
Muslim		-0.0142	0.00116	0.0159	-0.00235
		(0.0229)	(0.0230)	(0.0274)	(0.0248)
Tribal language fluency		0.00920	0.0129	0.00821	0.0110
		(0.00864)	(0.00836)	(0.00877)	(0.00836)
Network measure		0.0306***	0.0320***	0.0293***	0.0324***
		(0.00838)	(0.00877)	(0.00919)	(0.00870)
Log (parcel size m^2		-0.00742	-0.0133	-0.0123	-0.0161
		(0.0197)	(0.0207)	(0.0224)	(0.0206)
Parcel was inherited		-0.161***	-0.146***	-0.128***	-0.146***
		(0.0303)	(0.0302)	(0.0326)	(0.0300)
Toilet = pit latrine		0.0274	0.0279	0.0262	0.0292
		(0.0255)	(0.0246)	(0.0261)	(0.0246)
Power connection?		0.145***	0.153***	0.142***	0.150***

Table 3.11: Ethnic co-location and RL take up, full results - *continued*

	(1)	(2)	(3)	(4)	(5)
		(0.0234)	(0.0224)	(0.0242)	(0.0222)
Business on parcel?		0.0539	0.0420	0.0325	0.0486
		(0.0395)	(0.0383)	(0.0419)	(0.0384)
Parcel rented out?		-0.00863	0.00235	0.0165	0.00112
		(0.0222)	(0.0213)	(0.0223)	(0.0216)
Year parcel obtained		-0.00102	-0.00180*	-0.00264**	-0.00203**
		(0.00104)	(0.00102)	(0.00110)	(0.00103)
Avg N Age			-0.0532	-0.0688	-0.0587
			(0.0462)	(0.0492)	(0.0462)
Avg N Network			-0.0112	0.0388	-0.0130
			(0.0406)	(0.0454)	(0.0412)
Avg N schoolng			0.00352	-0.00187	0.00248
			(0.0313)	(0.0331)	(0.0315)
Avg N age			0.00953	0.00704	0.0111*
			(0.00624)	(0.00664)	(0.00626)
Avg N income			-2.42e-08	-3.75e-08	-2.86e-08
			(4.41e-08)	(4.59e-08)	(4.50e-08)
Avg N HH size			0.0310	0.0293	0.0362
			(0.0325)	(0.0358)	(0.0323)
Avg N Muslim			-0.0534	-0.0456	-0.0509
			(0.123)	(0.132)	(0.125)
Avg N fluency			-0.131**	-0.126**	-0.129**
			(0.0530)	(0.0568)	(0.0537)
Avg N renting			-0.256	-0.293*	-0.284*
			(0.164)	(0.173)	(0.165)
Avg N business			0.104	0.173	0.148
			(0.266)	(0.291)	(0.269)
Avg N inherited			-0.567**	-0.543**	-0.585**
			(0.243)	(0.258)	(0.248)
Avg N toilet			-0.242	-0.224	-0.198
			(0.161)	(0.174)	(0.163)
Avg N power			-0.193	-0.297**	-0.214
			(0.140)	(0.148)	(0.141)

Table 3.11: Ethnic co-location and RL take up, full results - *continued*

	(1)	(2)	(3)	(4)	(5)
Avg N p size			-0.118	-0.0717	-0.110
			(0.0888)	(0.0957)	(0.0891)
Kigogo Kati			0.122*	0.126	0.125*
			(0.0738)	(0.0774)	(0.0734)
D to church			-0.0000324	-0.0000595	-0.0000369
			(0.000118)	(0.000128)	(0.000117)
D to field			0.0000547	0.0000371	0.0000853
			(0.000175)	(0.000181)	(0.000177)
D to hazard land			-0.000275	-0.000243	-0.000318
			(0.000291)	(0.000308)	(0.000289)
D to mosque			0.000520	0.000544	0.000549
			(0.000415)	(0.000431)	(0.000416)
D to footpath			-0.0000442	0.0000264	-0.00000449
			(0.000346)	(0.000364)	(0.000341)
D to main road			-0.000334*	-0.000249	-0.000333
			(0.000200)	(0.000219)	(0.000204)
D to river			-0.000357	-0.000484	-0.000399
			(0.000280)	(0.000295)	(0.000283)
D to road			-0.0000545	-0.000184	-0.0000333
			(0.000338)	(0.000347)	(0.000335)
D to school			-0.0000364	-0.0000431	-0.0000471
			(0.000295)	(0.000316)	(0.000293)
D to mtaa office			-0.000572**	-0.000629**	-0.000629**
			(0.000269)	(0.000285)	(0.000272)
Tribal distance to Dar				-0.0000365	
				(0.0000688)	
Size of tribal homeland				-0.000000491	
				(0.00000107)	
# of coethnics in mtaa				0.470**	
				(0.207)	
Descent type (major)				0.0204	
				(0.0205)	
Constant	0.536***	2.213	5.599**	7.369***	6.113***

Table 3.11: Ethnic co-location and RL take up, full results - *continued*

	(1)	(2)	(3)	(4)	(5)
	(0.0177)	(2.098)	(2.178)	(2.344)	(2.229)
Tribe f.e.	No	No	No	No	Yes
R^2	0.010	0.079	0.127	0.126	0.139
Obs	2247	2247	2247	1978	2247

Conley standard errors in parentheses: $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Variable of interest is the percentage of other hhs within 50m from the same language stem

Dependent variable is a dummy for RL takeup

Table 3.12: Controlling for previous coethnic choice, full results

	(1) OLS	(2) t < 2005	(3) Restricted	(4) Final
% same stem (50m)	-0.380***	-0.434***	-0.468***	-0.576***
	(0.146)	(0.165)	(0.171)	(0.219)
% adjacent in same tribe	0.0667	0.0656	0.0771	0.0818
	(0.0549)	(0.0575)	(0.0600)	(0.0601)
# of neighbors (50m)	0.00205*	0.00270**	0.00303**	0.00299**
	(0.00120)	(0.00135)	(0.00137)	(0.00136)
HH size	0.0141***	0.0107**	0.0102*	0.0103*
	(0.00486)	(0.00526)	(0.00541)	(0.00540)
Log(assets)	0.00463	0.00360	0.00199	0.00184
	(0.00948)	(0.0102)	(0.0104)	(0.0104)
Monthly income (tsh)	-0.00000387	-0.00000387	-0.00000201	-0.00000214
	(0.00000345)	(0.00000534)	(0.00000659)	(0.00000656)
Avg years of schooling	0.00605	0.00386	0.00317	0.00317
	(0.00421)	(0.00465)	(0.00475)	(0.00476)
Average age	0.00149	0.00143	0.00153	0.00153
	(0.00127)	(0.00137)	(0.00139)	(0.00139)

Table 3.12: Controlling for previous coethnic choice, full results - *continued*

	(1)	(2)	(3)	(4)
	OLS	t < 2005	Restricted	Final
Muslim	-0.00235	-0.00183	-0.000817	-0.000436
	(0.0248)	(0.0271)	(0.0274)	(0.0274)
Tribal language fluency	0.0110	0.00336	0.00490	0.00476
	(0.00836)	(0.00919)	(0.00945)	(0.00945)
Network measure	0.0324^{***}	0.0295^{***}	0.0300^{***}	0.0299^{***}
	(0.00870)	(0.00918)	(0.00917)	(0.00916)
Log (parcel size m^2	-0.0161	-0.0158	-0.0201	-0.0202
	(0.0206)	(0.0221)	(0.0230)	(0.0230)
Parcel was inherited	-0.146^{***}	-0.164^{***}	-0.182^{***}	-0.183^{***}
	(0.0300)	(0.0344)	(0.0359)	(0.0358)
Toilet = pit latrine	0.0292	0.0232	0.0173	0.0174
	(0.0246)	(0.0273)	(0.0286)	(0.0286)
Power connection?	0.150^{***}	0.155^{***}	0.155^{***}	0.155^{***}
	(0.0222)	(0.0243)	(0.0248)	(0.0248)
Business on parcel?	0.0486	0.0525	0.0375	0.0389
	(0.0384)	(0.0406)	(0.0412)	(0.0412)
Parcel rented out?	0.00112	-0.00596	-0.0120	-0.0118
	(0.0216)	(0.0234)	(0.0242)	(0.0242)
Avg N Age	-0.0587	-0.0873^{*}	-0.0961^{*}	-0.0963^{*}
	(0.0462)	(0.0513)	(0.0526)	(0.0526)
Avg N Network	-0.0130	-0.00883	-0.00903	-0.00761
	(0.0412)	(0.0459)	(0.0463)	(0.0462)
Avg N schoolng	0.00248	0.0219	0.0212	0.0197
	(0.0315)	(0.0334)	(0.0342)	(0.0342)
Avg N age	0.0111^{*}	0.0128^{*}	0.0130^{*}	0.0130^{*}

Table 3.12: Controlling for previous coethnic choice, full results - *continued*

	(1) OLS	(2) t < 2005	(3) Restricted	(4) Final
	(0.00626)	(0.00699)	(0.00712)	(0.00712)
Avg N income	-2.86e-08	-1.27e-08	-1.69e-08	-1.63e-08
	(4.50e-08)	(4.91e-08)	(4.86e-08)	(4.83e-08)
Avg N HH size	0.0362	0.0514	0.0445	0.0455
	(0.0323)	(0.0356)	(0.0365)	(0.0366)
Avg N Muslim	-0.0509	-0.158	-0.180	-0.191
	(0.125)	(0.138)	(0.141)	(0.142)
Avg N fluency	-0.129^{**}	-0.138^{**}	-0.163^{***}	-0.162^{***}
	(0.0537)	(0.0591)	(0.0599)	(0.0598)
Avg N renting	-0.284^{*}	-0.135	-0.139	-0.133
	(0.165)	(0.184)	(0.189)	(0.189)
Avg N business	0.148	0.226	0.181	0.196
	(0.269)	(0.303)	(0.308)	(0.306)
Avg N inherited	-0.585^{**}	-0.444	-0.427	-0.433
	(0.248)	(0.277)	(0.281)	(0.281)
Avg N toilet	-0.198	-0.222	-0.221	-0.215
	(0.163)	(0.177)	(0.183)	(0.183)
Avg N power	-0.214	-0.255	-0.245	-0.246
	(0.141)	(0.157)	(0.161)	(0.160)
Avg N p size	-0.110	-0.117	-0.0846	-0.0866
	(0.0891)	(0.0967)	(0.0990)	(0.0988)
Kigogo Kati	0.125^{*}	0.187^{**}	0.166^{*}	0.168^{*}
	(0.0734)	(0.0873)	(0.0919)	(0.0919)
D to church	-0.0000369	-0.0000182	-0.0000578	-0.0000555
	(0.000117)	(0.000129)	(0.000130)	(0.000130)

Table 3.12: Controlling for previous coethnic choice, full results - *continued*

	(1) OLS	(2) t < 2005	(3) Restricted	(4) Final
D to field	0.0000853	0.000179	0.000131	0.000130
	(0.000177)	(0.000198)	(0.000204)	(0.000204)
D to hazard land	-0.000318	-0.000535*	-0.000479	-0.000497
	(0.000289)	(0.000324)	(0.000333)	(0.000333)
D to mosque	0.000549	0.000634	0.000669	0.000668
	(0.000416)	(0.000450)	(0.000467)	(0.000465)
D to footpath	-0.00000449	0.000196	0.0000556	0.0000675
	(0.000341)	(0.000382)	(0.000392)	(0.000391)
D to main road	-0.000333	-0.000352	-0.000329	-0.000330
	(0.000204)	(0.000225)	(0.000228)	(0.000228)
D to river	-0.000399	-0.000617*	-0.000589*	-0.000599*
	(0.000283)	(0.000322)	(0.000328)	(0.000328)
D to road	-0.0000333	0.0000478	-0.0000653	-0.0000696
	(0.000335)	(0.000380)	(0.000382)	(0.000382)
D to school	-0.0000471	-0.000151	-0.0000727	-0.0000546
	(0.000293)	(0.000325)	(0.000333)	(0.000334)
D to mtaa office	-0.000629**	-0.000769**	-0.000734**	-0.000750**
	(0.000272)	(0.000300)	(0.000307)	(0.000307)
Year parcel obtained	-0.00203**	0.000943	0.000540	0.000725
	(0.00103)	(0.00123)	(0.00131)	(0.00134)
% same stem at arrival				0.0972
				(0.123)
Constant	6.113***	0.425	1.327	0.997
	(2.229)	(2.702)	(2.851)	(2.900)
Tribe f.e.	Yes	Yes	Yes	Yes
R^2	0.139	0.141	0.142	0.142

Table 3.12: Controlling for previous coethnic choice, full results - *continued*

	(1)	(2)	(3)	(4)
	OLS	t < 2005	Restricted	Final
Obs	2247	1886	1784	1784

Conley standard errors in parentheses: $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Column 1 is the full OLS model from the previous table

Column 2 restricts the sample to parcels acquired prior to 2005. **Column 3** further restricts the sample to those who who had neighbors move in more than 5 years later.

Column 4 introduces a control for the percentage of % of neighbors at arrival (and within 5 years later) from the same language stem.

Dependent variable is a dummy for RL take-up

Table 3.13: Coethnic location and possible channels of tenure demand

	Expropriation risk		HH has borrowed		Given to public goods?	
	(1)	(2)	(3)	(4)	(5)	(6)
% same stem (50m)	-0.241**	-0.271**	0.0200	0.0678	-0.0188	0.00748
	(0.0979)	(0.137)	(0.0772)	(0.0953)	(0.150)	(0.199)
% same name	0.00741	0.00855	-0.0193	-0.0211	-0.116	-0.117
	(0.110)	(0.110)	(0.102)	(0.103)	(0.168)	(0.168)
% same religion	0.0731	0.0730	-0.00298	-0.00281	0.133	0.133
	(0.0544)	(0.0544)	(0.0522)	(0.0523)	(0.0844)	(0.0843)
% same stem at arrival		0.0273		-0.0436		-0.0240
		(0.0826)		(0.0545)		(0.115)
Standard controls	Yes	Yes	Yes	Yes	Yes	Yes
Tribe f.e.	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.117	0.118	0.134	0.134	0.086	0.086
Obs	1715	1715	1715	1715	1715	1715

Conley standard errors in parentheses: $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Dependent variables: (i) the perceived probability of expropriation in the next five years, (ii) indicator variable = 1 if the household has borrowed in the past five years, and (iii) indicator variable = 1 if the household has contributed anything to neighborhood public goods in past year

3.A.2 Fearon's measure of cultural similarity

First established by Fearon (2003) as an extension to standard Herfindahl-based measures of ethnolinguistic fractionalization, the measure of 'cultural similarity' is calculated as:

$$\tau_{ij} = \left(\frac{l_{ij}}{m}\right)^{\delta} \tag{3.4}$$

Where l is the number of shared branches between languages i and j and m is the maximum number of branches in the set of languages being considered. δ is a tolerance parameter, determining how much weight is given to earlier over later shared branches.[24] I use the same formula, with a slight modification to the way m is calculated.[25]

[24]Fearon (arbitrarily) sets this parameter = 0.5, although other authors have suggested alternative values (Desmet, Ortuño-Ortín, and Wacziarg (2012) use a value of $\delta = .05$). I use Fearon's preferred value for consistency. The resulting measure (3.4 takes a value between zero and one.

[25]Instead of setting m equal to the maximum number branches in the total set of languages, I set it equal to the maximum number of branches for each of the two languages being considered, as I feel this results in a better localized measure of cultural similarity.

Chapter 4

The price of empowerment: land titling and female inclusion in urban Tanzania

4.1 Introduction

Across sub-Saharan Africa, one stubbornly persistent sign of gender inequality is the low rate of land ownership among women. The Food and Agriculture Organization's (FAO) Gender and Land Rights database reveals the dire position most women face in this regard: of the countries surveyed women made up, on average, no more than 20% of all landowners.[1]

Meanwhile, as large-scale titling programmes continue to rise in popularity, researchers have primarily been concerned with identifying the aggregate effects these schemes have on households. However, significantly less attention has been given to the *intrahousehold* effects of land titling and similar property rights interventions, particularly their impact on women's asset ownership and marital bargaining power. This is a growing concern, as land titling programmes often produce female inclusion rates which are lower than the desired level, despite many schemes making joint-titling between husband and wife

[1] Own calculations. The FAO's database only covers a select group of countries, with data ranging from the early 1990s to the early 2000s.

a requirement (Deere and León 2001; Payne, Durand-Lasserve, and Rakodi 2007). If women are not being included as owners during the titling process, it is possible that these programmes might at best maintain an already unfavourable status quo and at worst strip women of the informal, customary claims to land that they might already enjoy.

This short chapter presents preliminary results from a unique experiment in the unplanned settlements of Dar es Salaam, a context in which formal land titles have only been available for a short time and self-reported *de facto* female ownership is quite low.[2] The main goal of the experiment was to drastically reduce the cost of obtaining a land title for a randomly-chosen group of land-owning households. However, the experiment also had a second level of randomisation, where a lottery system was used to introduce further random variation in the price that households faced to buy a land title, with some households being assigned discounts which were made conditional on including a woman as owner on the title application.

Using data on which households went on to purchase a land title and whether or not they included a woman in the application, we[3] show that not only do vouchers have a positive impact on purchase of land titles, but households receiving conditional subsidies are just as likely to purchase as those receiving unconditional subsidies, indicating that conditionality does not depress demand. We go on to show that, for those households purchasing a land title, receiving a conditional subsidy substantially and significantly increases the probability that a woman's name is included on the title. The overall result is that offering conditional discounts will increase, in net terms, the number of women listed as landowners. While these results are encouraging, the fact that households are so easily nudged into co-titling[4] raises concerns that they might not be treating the decision as if it has had significant implications for household bargaining power. To investigate this further, we investigate whether voucher assignments are more or less effective in households with higher levels of ex-ante bargaining power, as measured using baseline

[2]Only 13% of dual-headed households in our sample report a woman as being an owner of their land, with less than 50% reporting that a woman must be consulted in the event of sale, transfer or rental.

[3]This chapter is part of a co-authored work (see declaration at start of thesis).

[4]For the remainder of the chapter, we will use 'co-titling' to indicate any situation where are woman is included on a land title.

household characteristics.

To our knowledge, this is the first research to introduce randomised variation in women's access to property. It shows that not only are these interventions relatively easy to design and implement, but that they can have substantial effects on women's legal claims to ownership. This chapter is structured as follows: in Section 4.2, we discuss the motivation for such an experiment by drawing on existing evidence for gender and bargaining power impacts of property rights and land titling interventions. This section also covers the Tanzanian context, where recently-introduced land tenure reforms have created an opportunity for the intrahousehold status quo to change substantially. In Section 4.3, we discuss the experiment in more detail, specifically the conditionality of the vouchers, balance, and household characteristics at baseline. Section 4.4 covers the main results on demand for title, co-titling, and discusses the implications for effectiveness of gender conditionality. Finally, we conclude the chapter with Section 4.5.

4.2 Background and motivation

4.2.1 Titling, land ownership and bargaining power outcomes

While evidence of the impact of formal joint-titling on women's outcomes is limited, there are several studies which associate improvements in women's property rights with other desirable outcomes such as measures of female empowerment, child health, education and women's welfare, all of which are associated with increases in bargaining power. For example, self-reported ownership of land is positively correlated with child health status and various measures of empowerment in Nepal (Allendorf 2007) and with expenditure on 'gendered' goods in both China and Ghana (Wang 2011; Doss 2005). Inheritance rights, in particular, appear to matter: Peterman (2011) shows that women in rural Tanzania who enjoy improvements in inheritance rights are more likely to enter the labour market and earn higher wages. Telalagic (2012) shows that women from villages practicing matrilineal descent, whose improved inheritance rights result in a better outside-option, are less likely to utilise domestic labour as a source of bargaining power. Both Roy (2008) and Deininger, Goyal, and Nagarajan (2010) have found a positive impact stemming from India's Hindu

Succession Act, which extended inheritance rights to women, on outcomes such as female education and self-reported autonomy.

There is also growing evidence that formal land titling itself can be advantageous to women, irrespective of their state of ownership. Using data from a Peruvian titling programme with a distinct focus on joint-titling, Field (2003) demonstrated a link between title acquisition and subsequent reduction in household fertility. Galiani and Schargrodsky (2010) show that titling in Buenos Aires resulted in a reduction in household size and higher levels of child education. Preliminary evidence from Rwanda has also shown that titling programmes can be successful at increasing perceived female ownership and the recording of inheritance rights (Ayalew Ali and Goldstein 2011).

While it is clear that land titling has the capacity to improve the lot of women in developing countries, most studies are unable to distinguish the overall impact of titling from the *additional* impact of joint-titling (what we will call co-titling in this chapter). This distinction might seem less crucial in contexts where land titling is compulsory, but in the face of large costs for formalisation governments are often resorting to demand-driven approaches (Payne et al. 2007). In these settings, if households see co-titling as a cost, then policymakers might find that convincing households to purchase property titles and getting them to co-title are conflicting goals. If making co-titling a requirement depresses a household's demand for a title, we should be concerned with identifying the 'price of empowerment', the subsidy required to offset that reduction in demand.

4.2.2 Female land ownership in urban Tanzania

To get a sense of the current state of women's *de jure* ownership in Tanzania, we return to the 1999 Land Act, which has previously been hailed as being one of the first pieces of land legislation to explicitly recognise the rights of women as landowners (Sundet 2005). The Land Act also established two forms of urban land tenure, a short term lease known as a residential license, and a longer-term title known as a certificate of right of occupancy (CRO), the latter being our focus in this chapter. The Land Act has several provisions relating directly to both the default ownership status and the rights of spouses. The default ownership state for spouses is known as occupancy-in-common, which provides each spouse with an equal share which can be sold (conditional on the agreement of the other shareholder) or left as inheritance. Legally, this implies that each 'occupier' or owner has substantial control over the land: one occupier will be unable to sell, rent or mortgage the property without consent of the other.

This would suggest that women actually have significant legal control over land, yet the Land Act suggests that spouses of landowners might not be considered to be occupiers-in-common if they are not listed as an owner on a CRO:

> "Where a spouse obtains land under a right of occupancy for the co-occupation and use of both spouses or where there is more than one wife, there shall be a presumption that, unless a provision in the certificate of occupancy or certificate of customary occupancy clearly states that one spouse is taking the right of occupancy in his or her name only... the spouses will hold the land as occupiers in common and, unless the Presumption is rebutted in the manner stated in this subsection, the Registrar shall register the spouses as joint occupiers accordingly."

The law appears to be ambiguous enough to allow landowners to register land in their name only, excluding their spouses as legal occupiers. In practice this appears to be quite common: as we observed in Chapter 1, records from Dar es Salaam's Kinondoni Municipality suggest that this is the case, with an estimated 75% of properties in the municipal land registry having only a single male owner listed.[5] It then appears that being included on a being included on a CRO application is a crucial, if difficult step to cementing legal ownership.

However, while giving a woman status as an occupier-in-common appears to give her substantial *de jure* control over land, it is unclear whether or not this would actually lead to *de facto* changes in her outside options or bargaining power. For example, while Tanzania's 1971 Law of Marriage Act gives Tanzanian courts some flexibility in assigning

[5] The Kinondoni Municipality does not record the gender of the landowner. To estimate the proportion of male landowners, household data from the Tanzanian Land Rights Survey was used to identify a series of forenames which were unique to each gender (approximately 90% of all names from the roster were gender-unique). These 'gendered' names were then merged with the property register, of which 77% of all owner names could be identified using a gender-specific name.

assets at marriage, it still maintains that customary law should, whenever possible, take precedent. Frustratingly, the 1999 Land Act does not specifically discuss divorce, so it is unclear whether or not there is actually a legal basis for a co-titled wife having better outside options, as households may still use customary law to deal with marriage dissolution. However, as the Law of Marriage Act specifically mandates that property assigned to only one spouse cannot be claimed by the other spouse, co-titling seems to be the only way to ensure that a wife has a nonzero probability of making a successful claim.

Finally, the Land Act already contains a few provisions which are intended to provided spouses with some basic rights over land, irrespective of their inclusion in formal documents. Subsequent provisions, listed in Table 4.8 in the Appendix, allow for spouses who have invested in the land to be considered co-occupiers by default. There are also provisions which stipulate that owners who decide to sell or mortgage the land must first obtain the consent of their spouses. It is unknown whether or not these particular provisions are actually enforced or upheld during conflicts over ownership.

Despite these concerns, data on current *de jure* suggests there is substantial room for improvement. Figure 4.1 gives a sense of the state of *de facto* ownership: it is constructed using baseline data from the experimental intervention, which is discussed in more detail in the following section. Households in two unplanned settlements in Dar es Salaam were asked a series of questions about the de facto ownership of land, including the rights of household members over the sale, rental and transfer of land, as well as who would be include in a CRO application if one was made. The results, which are restricted to dual-headed households, suggest that women have limited *de facto* rights over land: roughly 13% of households report that a woman is one of the "default" owners of the land. Despite this, women fare better in 'use' rights, with just over 40% of households reporting that at least one woman in the household must agree before the land can be sold, transferred or rented out.[6] Finally, when households were asked who would be included on a CRO, only 25% indicated that a woman would be included. While this is certainly better than the

[6]To avoid priming, households were not asked directly about female ownership. Instead, they were asked to list all members of the household that were default owners, must be consulted before a sale, or would be included on a CRO.

Table 4.1: Female land ownership in Dar es Salaam

Variable	Mean	(Std. Dev.)	Min.	Max.	N
One of default owners is female	0.132	(0.339)	0	1	606
Woman has rights over land sale	0.449	(0.498)	0	1	602
Woman has rights over transfer	0.437	(0.496)	0	1	602
Woman has rights over rental	0.42	(0.494)	0	1	602
Household would include woman on CRO	0.253	(0.435)	0	1	600

Notes: data are from Tanzanian Land Rights survey. Sample restricted to dual-headed households in treatment blocks.

de facto state, it reinforces concerns that titling may help cement a status quo in which women are excluded from owning land.

4.3 The experiment, baseline data collection and empirical framework

While the experiment and its general context were presented in some length in Chapter 2, it is worth reviewing the main details here, as the voucher scheme was not discussed in much detail. We return in the same setting as the previous chapters, where we focus on two adjacent communities in Dar es Salaam's Kinondoni Municipality. As described before, Mburahati Barafu and Kigogo Kati are unplanned, informal settlements with markedly low levels of access to infrastructure and public utilities, even by the relatively low benchmark set by other communities in the municipality. Both of these *mitaa* also appear to have noticeably lower levels of female land ownership: investigating the gender breakdown of land ownership in the Kinondoni land registry reveals that Barafu and Kati have female ownership rates of 17% and 22% respectively, compared to the municipal average of 25%.

The main purpose of the experiment was to induce households in both communities to purchase certificates of right of occupancy (CROs), in order to subsequently study their impact. This involved several levels of randomisation:

1. **Cadastral survey and repayment programme**: blocks of land parcels were identified and randomly selected into treatment and control groups. All parcels in treatment blocks were subject to cadastral surveying, with residents given the

option to repay the heavily-subsidized cost (100,000 TSh) in exchange for a land title, drastically bringing down the cost of a CRO for residents.

2. **Random price variation within treatment blocks**: households within treatment blocks were randomly allocated vouchers redeemable for different levels of discount on the final price of a CRO.

3. **Random voucher conditionality**: roughly half of these vouchers were made conditional, where households were only allowed use them if a female household member was included as an owner on the CRO application.

Next, we will discuss these interventions in more detail, including the timing of their introduction in both communities.

4.3.1 Main intervention and voucher distribution

In the summer of 2010, prior to the intervention, the University of Oxford conducted a complete census of land parcels in Barafu and Kati, known as the Tanzanian Land Rights Survey (TLRS). These data were used in forming baseline controls in Chapters 2 and 3. Households were identified using records and maps from the Kinondoni Municipality, which had created a listing of all households in the area to assist with the creation of the land registry. Using this listing, parcel-owning households were identified and interviewed,[7] resulting in detailed data on household and parcel characteristics.

Following this survey, a ward-level meeting was held by a local NGO, the Women's Advancement Trust (WAT), to explain the overall intervention and process of selection into treatment and control blocks. Using a town plan recently drawn up as a prerequisite for CRO distribution, we then divided land parcels into 'blocks' (contiguous groups of parcels), randomly assigning half of these intro treatment and control groups.[8] All parcels in treatment blocks were subject to a cadastral survey and owning households were invited to participate in the programme to obtain a land title, which required them to repay the cost of 100,000 TSh over roughly a six month period.

[7]The survey team was agnostic about *which* household member was the actual member. The only condition for interview was that one of the household members was considered to be the owner.

[8]For Barafu, the total number of blocks was 10, for Kati it was 15.

Table 4.2: Intended general and conditional voucher allocations

	Conditional Discount					
General Discount	0	20k	40k	60k	80k	Total
0	0.067	0.067	0.067	0.067	0.067	0.333
20k	0.067	0.067	0.067	0.067	.	0.267
40k	0.067	0.067	0.067	.	.	0.2
60k	0.067	0.067	.	.	.	0.133
80k	0.067	.	.	.	.	0.067
Total	0.333	0.267	0.2	0.133	0.067	1

Notes: The baseline price was 100,000 TSh for a CRO, per parcel, regardless of size or other characteristics. Each cell shows the probability a household is assigned a given general/conditional voucher discount. Blank cells were not used to avoid offering a negative net price.

The second and third dimensions of the intervention were cross-cutting and randomised at the individual parcel level within treatment blocks. After treatment parcels were selected, owners were to be given up to two types of discounts on the price of a CRO, both redeemable at WAT's office. The first type was an unconditional voucher, a simple discount on the 100,000 TSh price. The second was a conditional voucher, which could only be applied if one of the names registered on the CRO application form was a female household member. These conditions were carefully explained in Swahili on each type of voucher. If households elected to use a conditional voucher, names were checked at the time of application to ensure compliance with the requirements. Vouchers were assigned to a parcel, rather than to a particular owner, so as to remain impartial to the identity of the actual owner within the household and to prevent vouchers from being exchanged between households. Examples of both types of vouchers can be found in Figure 4.4 in the appendix for this chapter.

Vouchers could take on values ranging from zero to 80,000 TSh, in iterations of 20,000, so households could face subsidies between 0% and 80% of the total cost of a CRO. This variation will be crucial for our ability to estimate the price-elasticities of demand for both unconditional and conditional 'prices' of CROs. As shown in Table 4.2, every feasible combination of vouchers was given equal weighting in the randomisation.[9]

While there were ex-ante concerns that a randomised top-down voucher allocation

[9]The net price of a title was restricted to be strictly greater than zero, so any voucher combination which would violate this restriction was excluded from the randomisation.

might be perceived as unfair by participants, block-level public lotteries were deemed to be too impractical and problematic for ensuring balance and compliance. To balance these two concerns, we performed the voucher randomisation in the following manner for each block:

1. We randomly drew a distribution of general/conditional voucher pairs, repeating the draw 100 times.

2. Balance was then tested for each draw using a vector of observable parcel-level characteristics and the three draws that were the most balanced (defined by *average* t-stat values) were kept.

3. These three outcomes were then presented to residents at the block-level information sessions. Each attendee as made aware of the three possible distributions, each labeled with a designated number. One of the attendees was selected by the rest to draw a number out of a hat, each number corresponding to a voucher distribution outcome. Whichever number was chosen determined the draw that would be used for the voucher distribution.

Thus we were able to maintain control over the broad aspects of the randomisation while still allowing residents some perceived agency in choosing the outcome. Following the voucher distribution, households were free to sign up with WAT and begin repayment.

Both the block and the parcel-level randomisations in Barafu and Kati were performed at different times and thus represent independent draws. Due to delays in the government provision of the maps necessary to identify treatment and control households, the programme was first introduced in Barafu in late 2010, but not in Kigogo Kati until approximately a year later. In Barafu, block-level information and voucher sessions were held in late October, 2010, with participating landowners paying their net price to WAT between November and the summer of 2011. Following repayment, landowners in Barafu have been filling out and turning in CRO applications, to then be checked and sent on to the local government by WAT. In Kigogo Kati, the voucher sessions were held in early November, 2011, with repayment continuing until the summer of 2012. Due to excessive flooding in Kati, overall participation and take up has been significantly lower than in Barafu. The data presented in this chapter comprises the latest take up and application data available from the project.

4.3.2 Balance and summary statistics

Table 4.3 shows summary statistics for a select group of baseline characteristics, as well as a series of balance tests. To test whether there is a significant correlation between assigned voucher values and baseline characteristics, we estimate the following specification for each characteristic using ordinary least squares:

$$x_i = \alpha + \pi_1 v_G + \pi_2 v_C + \pi_3 kati + \varepsilon_i \tag{4.1}$$

Where x_i is the characteristic of interest, v_G is the general voucher value, v_C is the conditional voucher value, expressed in thousands of shillings and *kati* is a dummy equal to one if the household is part of the Kigogo Kati randomisation. While it is more common to test the bivariate relationship between baseline characteristics and a single treatment, this method is most-closely approximates the specification we will be using in the next section. Furthermore, as general and conditional voucher values were drawn as part of a joint distribution, it is more appropriate to test for the partial correlation between each voucher value while holding the other constant.

In Table 4.3, column (1) shows the mean and standard deviation for each baseline characteristic. These include the year the parcel was acquired, whether or not it is currently being rented out, whether it was inherited, if the parcel has electricity access, whether there has been recent investment in the parcel and the log of the parcel size in square meters. Household characteristics include whether the household is Muslim, monthly income and total assets, the household's average schooling and size, and whether the household live in the parcel. While these are the characteristics we will be using as controls in the next section, we might also be interested in whether the intervention is balanced along a range of measures of female empowerment. These include whether the household is a single-female headed household, whether a woman in the household has any use rights, whether or not there is a default female owner, if the household would hypothetically include the woman on a CRO, and the percentage of total household income contributed by the female household head.

Columns (2) and (3) show estimates of π_1 and π_2, respectively. Column (4) displays the point estimate of a bivariate regression of the baseline characteristic on the net price faced by the household ($100-v_G-v_C$). In general, there is good balance across the range of baseline characteristics. There are a few significant differences: households with a higher likelihood of having access to electricity had higher general and conditional voucher values, inherited parcels were assigned slightly lower voucher values. There is also a slight lack of balance between household size, parcel size, the female household head's share of income and general voucher values. On the whole, these differences are small, but do imply that these characteristics should be used as control in the main specification. In the next section, we will include most of these baseline characteristics as controls.

Table 4.3: Summary statistics and balance

	Mean/SD	General	Conditional	Price
	(1)	(2)	(3)	(4)
Year parcel was acquired	1992.487	-.009	-.024	0.018
	(13.505)	(0.017)	(0.019)	(0.015)
Parcel is rented out	0.388	-.001	-.0006	0.0008
	(0.512)	(0.0007)	(0.0007)	(0.0006)
Parcel was inherited	0.113	-.0004	-.0008	0.0006
	(0.332)	(0.0005)	(0.0004)*	(0.0004)
Electricity access?	0.398	0.001	0.001	-.001
	(0.514)	(0.0007)**	(0.0007)**	(0.0006)**
Recent investment in parcel	0.214	0.0004	0.0007	-.0005
	(0.43)	(0.0006)	(0.0005)	(0.0005)
Muslim hh	0.569	-.0004	-.0004	0.0004
	(0.522)	(0.0007)	(0.0007)	(0.0006)
Monthly income (tsh '000)	387.497	-.915	-.817	0.86
	(686.831)	(0.957)	(0.744)	(0.71)
Log(total assets (tsh '000))	7.518	-.003	0.0005	0.0008
	(1.238)	(0.002)	(0.002)	(0.001)
Average schooling of hh	12.219	-.002	-.002	0.002
	(2.895)	(0.004)	(0.004)	(0.003)
Household size	5.044	0.007	0.003	-.005
	(2.711)	(0.004)*	(0.003)	(0.003)
Log(Parcel Area m^2)	5.115	-.002	-.0008	0.001
	(0.579)	(0.0008)**	(0.0008)	(0.0007)
HH lives on parcel	0.794	0.0004	0.0005	-.0004
	(0.425)	(0.0006)	(0.0006)	(0.0005)
Single female-headed household	0.189	-.0003	-.0007	0.0005
	(0.413)	(0.0006)	(0.0005)	(0.0005)
Woman has rights over sale	0.582	-.0004	-.0002	0.0003
	(0.593)	(0.0008)	(0.0008)	(0.0007)
De facto female owner	0.266	-.0005	-.0006	0.0005
	(0.464)	(0.0007)	(0.0006)	(0.0005)
Would hypothetically cotitle	0.355	-.0004	0.00005	0.0001
	(0.507)	(0.0007)	(0.0007)	(0.0006)
Women's share of hh income	0.307	-.001	-.001	0.001
	(0.546)	(0.0008)	(0.0007)*	(0.0006)*
Obs	1148	1148	1148	1148

Column (1) displays the mean and standard deviation for each variable. Columns (2)-(3) display the mean and standard error of β_2 and β_3 from the linear regression of each variable $var = \beta_1 + \beta_2 * G + \beta_3 * C$, where G and C are the general and conditional voucher values. Column (4) shows the results of a single bivariate regression of each variable on the overall price households faced, net of all vouchers. Voucher values are measured in ('000 TSh). Robust standard errors $^{*}(p < 0.10),^{**}(p < 0.05),^{***}(p < 0.01)$

4.3.3 Empirical framework

Consider a linear probability model of a similar form presented in previous chapters, where the household's decision to purchase a title T_i is a function of household characteristics $\mathbf{x}_i$ and the price faced p_i:

$$T_i = \alpha^* + \beta_p p_i + \mathbf{x}_i\delta + \varepsilon_i \tag{4.2}$$

Given the project has exogenously-varied the price p_i through the voucher-subsidy lottery, equation (4.2) is an empirical representation of a linear demand function. However, because we might be interested in the differential effects of the voucher subsidies, we can separate p_i into its constituent components:

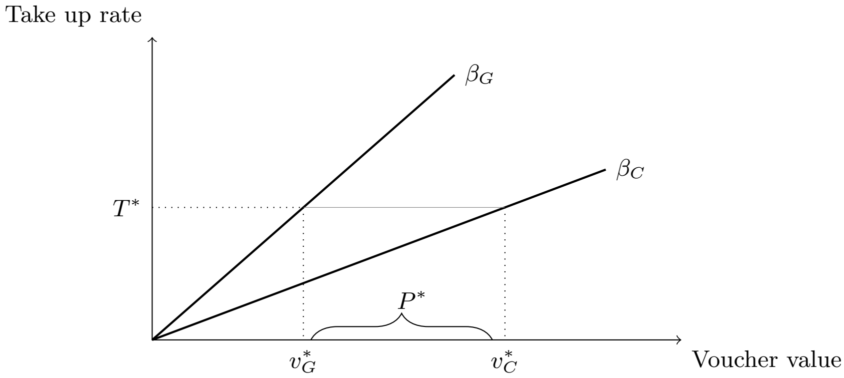
$$T_i = \alpha^* + \beta_p(100 - v_G + \gamma v_C) + \mathbf{x}_i\delta + \varepsilon_i \tag{4.3}$$

Note that increasing the general voucher amount v_G is equivalent to directly reducing the price, so that $\frac{\partial Pr(T_i=1)}{\partial v_G}$ is simply the negative of the price coefficient β_p. However, we might expect the effect of a conditional voucher to differ. Without making any assumptions on their structure, imagine a simple context where households receive some disutility from co-titling M, with γ measuring the inverse of that disutility $\gamma = \frac{1}{1+M}$.[10] In this case, providing a conditional voucher does two things: it increases the probability of take-up through the demand effect of a price reduction, but due to the conditionality, this effect will be strictly less than $\frac{\partial Pr(T_i=1)}{\partial v_G}$. This drives a wedge between the demand functions for general and conditional titling, so that, as long as $M > 0$, conditional vouchers will be less effective that general vouchers. As M gets very large, γ approaches zero, with conditional vouchers having no effect on the probability of take-up. If there is no disutility from co-titling, then $\gamma = 1$ and conditional vouchers will be indistinguishable from general vouchers in their effect on take-up.

To test the relative impact of these two randomised vouchers and households' subsequent take-up, we can estimate a linear probability of the following form:

[10] This is an arbitrary functional form, which could easily be captured by any function which is decreasing in M, but equal to one when $M = 0$.

Figure 4.1: The price of "empowerment"



$$T_i = \alpha + \beta_G v_G + \beta_C v_C + \mathbf{x}_i \delta + \varepsilon_i \tag{4.4}$$

In this equation, v_G and v_C are the levels of general and conditional vouchers which household i has been allocated, expressed in thousands of Tanzanian shillings. For all demand estimates, we restrict the effect of voucher values to be linear. Table 4.9 in Appendix 4.A.1 displays the results from a series of tests which suggest that this linear restriction is reasonable for take-up. The matrix $\mathbf{x}_i$ indicates household and parcel-level characteristics from the baseline survey, which will be included in some specifications. The outcome measure, T_i, is a dichotomous variable, equal to one if the household has fully paid for a CRO through the programme.

The coefficients β_G and β_C capture the impact of general and conditional vouchers on a household's propensity to take-up. Comparing equations (4.4) and (4.3) reveal that a test of $\beta_G = \beta_C$ is equivalent to testing $\gamma = 1$. If β_C is discernably smaller than β_G, this would reveal that households, on average, show distaste for co-titling. If β_C is substantially lower than β_G, households with conditional titles would need to receive higher subsidies to ensure a given level of titling. This is the "price of empowerment", the amount that would be need to transferred to households to offset the decline in demand caused by conditionality. Figure 4.1 illustrates this relationship: for a desired level of CRO take-up T^* and linear demand effects of general and conditional vouchers βG and βC, the extra discount needed to offset the conditionality of the vouchers is given by $v_C^* - v_G^* = P^*$. This price is crucial for policymakers weighing the benefits of co-titling against the extra costs associated with the reduction in demand.

While the implications for conditional vouchers in take-up are relatively straight forward, how might we expect the vouchers to affect observed co-titling outcomes? When there is zero disutilility from cotitling ($M = 0$), households should be, at worst, indifferent to co-titling, so will either all co-title, not co-title, or do so randomly. In this instance, conditional vouchers should push all receiving households into the co-titling state, conditional on taking up.

If M is high enough, then not only will conditional vouchers have zero effect on take-up (as households will opt not to use them) but there should be no observed co-titling, nor any positive effect of conditionality on co-titling. However, at some intermediate levels of M, households should be responsive to conditional vouchers, as there will be some level of conditional subsidy which, by offsetting the disutility, is enough to induce the household to switch from single-titling to co-titling. Thus, for as long as $\beta_G > \beta_C > 0$, conditional vouchers will be less effective at inducing households to adopt, but receiving households will be more likely to co-title.

In the next Section, we consider the basic demand results, where we test the effectiveness of both voucher types in inducing households to take-up a CRO. In the following subsection, we will examine how effective conditional vouchers are on inducing households to co-title.

4.4 Results

4.4.1 Demand results

Table 4.10 displays the results from the estimation of equation (4.4), first for the two communities Barafu and Kati separately, then pooling both together. In each case, the sample is restricted to households in treatment blocks.[11] Column (1) shows the results from estimating equation (4.4) without baseline controls. In Column (2), the sample is restricted to households with baseline data and in column (3) we include baseline controls.

[11] As discussed in Chapter 2, households in control blocks were excluded from purchasing through the NGO, and local records suggest that none have gone on to purchase CROs through the municipal government.

This process is repeated for Kigogo Kati in columns (4), (5) and (6).

In Barafu, general vouchers have a large, positive effect on take-up of CROs, with a point coefficient of 0.00471, indicating that each TSh 20,000 of subsidy results in an increase in the predicted probability of take-up by 9.4 percentage points. This effect increases slightly when baseline controls are included, but not substantially so. While it is still significant, the effect of general vouchers is much smaller in Kigogo Kati, where each TSh 20,000 reduction in the price of a CRO leads to roughly a three percentage point increase in the probability of take-up. In this specification, the constant can be interpreted as the take-up rate for households receiving no vouchers of either type. The results indicate that not only do general vouchers appear to be less effective in Kati, but overall adoption rates as well.

Across both neighbourhoods, households appear to be equally responsive to conditional vouchers. At the bottom of Table 4.10, "Test 1" reports the p-value from the linear test of $\beta_G = \beta_C$, revealing that we can comfortably accept the null that these two coefficients are equal across all specifications.[12] The results here strongly suggest that households in both communities treat conditional vouchers as 'cash': that is, they do not appear to be any demand effects of imposing conditionality. This implies that, on average, gender conditionality can be imposed without excluding households averse to co-titling. This indicates that households do not appear, on average, to suffer any expected disutility from co-titling. We will discuss the implications for bargaining power effects shortly. Figure 4.2 displays estimated take-up levels for each voucher type for Barafu and Kati separately. While the pattern of take-up across each value value differs slightly between general and conditional vouchers, they do not appear to be significantly different.

Columns (7) and (8) of Table 4.10 pool the data from both communities, reporting average voucher effects across both *mitaa* while controlling different level effects with a dummy for Kigogo Kati. Again, general and conditional effects appear to be almost identical, with lower overall take-up in Kati. Interacting the voucher values with the Kati dummy reinforces the result that demand response is lower in Kati, but that households in both locations do not discern between general and conditional vouchers.

[12] The failure to reject the null is not driven by imprecision, as the coefficients displayed here are precisely estimated.

Table 4.4: Effect of voucher distribution on CRO adoption

	Barafu			Kati			Pooled	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
General voucher value (tsh ’000)	0.00471***	0.00420***	0.00501***	0.00152**	0.00161**	0.00153**	0.00278***	0.00496***
	(0.00110)	(0.00117)	(0.00119)	(0.000650)	(0.000692)	(0.000667)	(0.000611)	(0.00118)
Conditional voucher value (tsh ’000)	0.00480***	0.00511***	0.00541***	0.00174***	0.00166**	0.00167**	0.00320***	0.00533***
	(0.000941)	(0.000957)	(0.000944)	(0.000651)	(0.000691)	(0.000673)	(0.000561)	(0.000941)
General × Kati								-0.00341**
								(0.00135)
Conditional × Kati								-0.00360***
								(0.00116)
Kati dummy							-0.460***	-0.273***
							(0.0279)	(0.0639)
Constant	0.349***	0.375***	0.346***	0.0754**	0.0754**	0.0710**	0.461***	0.346***
	(0.0538)	(0.0559)	(0.0559)	(0.0297)	(0.0315)	(0.0307)	(0.0367)	(0.0556)
Baseline controls	No	No	Yes	No	No	Yes	Yes	Yes
Test 1: $\beta_G = \beta_C$	0.929	0.386	0.703	0.752	0.947	0.849	0.482	0.723
Test 2: $\beta_G + \beta_{G\times K} = \beta_C + \beta_{C\times K}$								0.801
R^2	0.0576	0.0621	0.104	0.0130	0.0132	0.0702	0.283	0.291
Obs	461	422	422	684	612	612	1034	1034

Notes: Linear probability model. Dependent variable = 1 if household has fully paid for a CRO. The first three columns show results using the Barafu sample with column (1) using no unreported controls. Column (2) restricts the sample to households with non-missing controls. Column (3) includes baseline controls. The three columns for Kati sample follow this same pattern. Columns (7) and (8) pool both samples together, with and without Kati dummy interactions. Test 1 displays the p-value from a linear test of the hypothesis that the general and conditional voucher coefficients are equal. Test 2 tests the same hypothesis in the fully-interacted model. Robust standard errors $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Figure 4.2: Voucher values and take-up rates, by mtaa

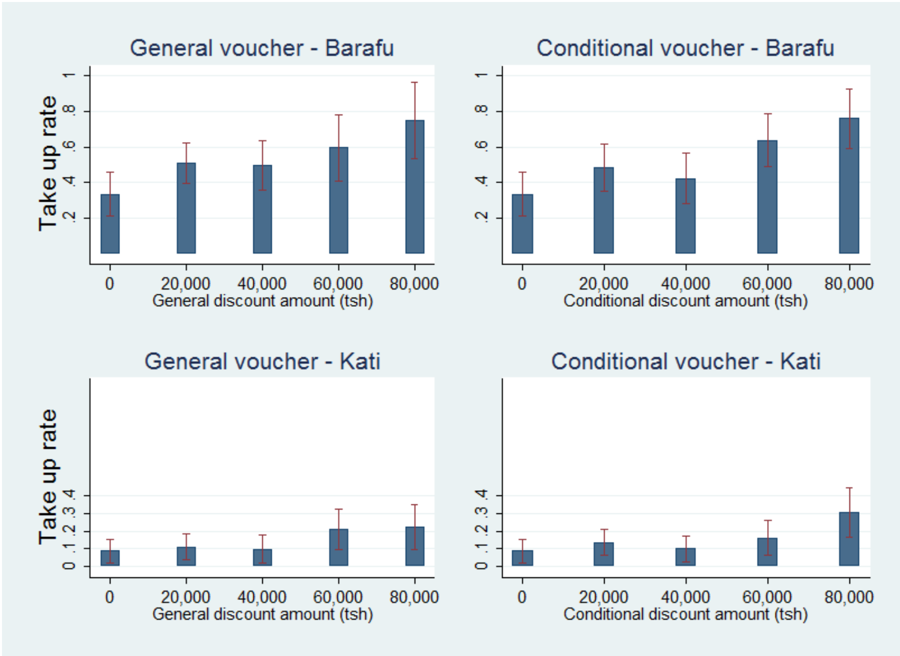


Figure shows estimates of take-up probability, conditioning on general conditional voucher values. Red bars indicates 95% confidence interval.

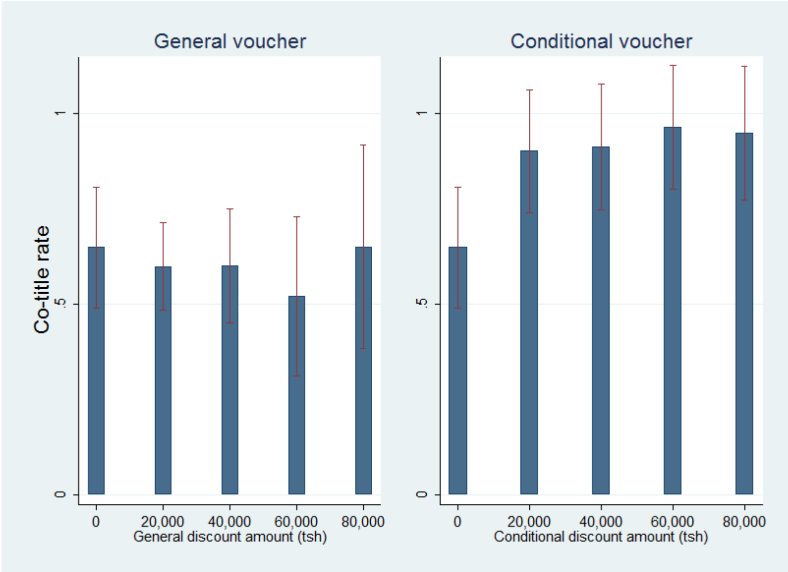
4.4.2 Co-titling results

While the results in the previous subsection encouragingly suggest that applying conditionality does not deter households from purchasing land titles, it is not yet clear that this conditionality actually leads to an increase in co-titling. Households might be indifferent to listing women as owners or might have all planned to co-title irrespective of any conditionality.

To investigate whether households respond to price incentives by co-titling, we rely on data from the household's CRO application, where women from the household were identified an recorded. Define $cotitle_i$ as a binary outcome equal to one if the household has included *any* woman from the household on the CRO application, *conditional* on the household having chosen to purchase a CRO. We then wish to re-estimate (4.4), using this variable as our outcome of interest:

$$cotitle_i = \alpha + \beta_G v_G + \beta_C v_C + \mathbf{x}_i \delta + \varepsilon_i \tag{4.5}$$

Figure 4.3: Voucher values and female co-title rates



Note: Figure shows estimates of co-titling (conditional on submission of an application probability), conditioning on general/conditional voucher values. Red bars indicates 95% confidence interval. Sample is pooled across both mitaa.

Ideally, equation (4.5) should be estimated over the full sample of households who have chosen to purchase a CRO. However, to date, application data is not available for approximately 30% of households who have taken up. While we will proceed as if the determinants of application data being observable are random, it is possible that non-random selection could lead to bias of our estimates. We investigate this further in Appendix 4.A.2, where we use a basic sample selection model to suggest that selection is not resulting in any significant bias of our estimates of β_G and β_C.

Table 4.5 shows the results from estimating (4.5) for Barafu and Kati separately in columns (1) and (2), then as a pooled sample, all with baseline control included. For Barafu, while the general voucher only has a small, insignificant negative effect on co-tilting, the conditional voucher has a large, positive significant impact, with each 10,000 TSh subsidy resulting in an increase in the predicted probability that a woman is included by 3.5 percentage points. The results are even stronger in Kigogo Kati, where each 10,000 TSh of conditional subsidy increases the probability of co-titling by 7.2 percentage points. Column (3) shows the results, pooling both samples together, and in column (4) we de-

Table 4.5: Effect of voucher distribution on female co-titling, conditional on CRO application

	Barafu	Kati	Pooled	
	(1)	(2)	(3)	(4)
General voucher	-0.00103	0.000544	-0.000828	
	(0.00156)	(0.00316)	(0.00122)	
Conditional voucher	0.00348***	0.00724**	0.00372***	
	(0.00126)	(0.00283)	(0.00106)	
General = 20				-0.0250
				(0.0657)
General = 40				-0.0491
				(0.0773)
General = 60				-0.144
				(0.101)
General = 80				0.0559
				(0.114)
Conditional = 20				0.288***
				(0.0697)
Conditional = 40				0.297***
				(0.0763)
Conditional = 60				0.324***
				(0.0807)
Conditional = 80				0.304***
				(0.0962)
Constant	0.748***	0.540***	0.733***	0.645***
	(0.0807)	(0.167)	(0.0649)	(0.0749)
Baseline controls	Yes	Yes	Yes	Yes
R^2	0.130	0.176	0.126	0.171
Obs	211	53	264	264

Notes: Linear probability model. Dependent variable = 1 if household included a woman on their CRO application, conditional on having paid for a CRO. Sample is restricted to households with application data. First three columns use linear measures of voucher values. Last column introduces dummy for each voucher value. Robust standard errors. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

Table 4.6: Net effect of voucher distribution on co-titling

	Barafu	Kati	Pooled	
	(1)	(2)	(3)	(4)
	All	All	All	Dual-headed
General voucher	0.00243**	0.000416	0.00109**	0.00158**
	(0.00116)	(0.000467)	(0.000509)	(0.000698)
Conditional voucher	0.00498***	0.000318	0.00228***	0.00288***
	(0.000981)	(0.000410)	(0.000489)	(0.000649)
Constant	0.214***	0.0428**	0.326***	0.324***
	(0.0530)	(0.0208)	(0.0319)	(0.0424)
Baseline controls	Yes	Yes	Yes	Yes
Test: $\beta_G = \beta_C$	0.0219	0.835	0.0227	0.0652
R^2	0.0849	0.0185	0.214	0.231
Obs	422	615	1037	603

Notes: Linear probability model. Dependent variable = 1 if household has fully paid for a CRO and submitted an application with a woman listed on it. First two columns restrict the sample to Barafu and Kati, respectively. Column (3) pools both samples. Column (4) restricts sample to dual-headed households. Robust standard errors. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

part from the usual linear specification and introduce the individual voucher values as dummies. Households which receive *any* conditional voucher are 29-30 percentage points more likely to co-title than those that receive no voucher (the omitted category). This effect is persistent and statistically indistinguishable across all voucher values, indicating that households are effectively nudged into co-titling by conditional vouchers. This is illustrated in Figure 4.3, where co-titling rates are graphed against voucher values, indicating that households receiving any conditional voucher are highly likely to include a woman as an owner on the CRO application. In contrast, general vouchers do not appear to have any significant effect on co-titling.

As households receiving conditional vouchers are no less likely to purchase a CRO, but are almost certain to co-title, this suggests that imposing conditionality can only increase the total number of women on land titles. To test this, we define an unconditional co-titling outcome, equal to one if the household purchases a CRO, submits an application and includes a woman as an owner on the application, and equal to zero otherwise. We then repeat the standard specification with this "net co-titling" outcome to see if, in aggregate, conditional vouchers are more successful at moving households into a co-titled state. Columns (1) and (2) of Table 4.6 display the results for Barafu and Kati separately, with column (3) showing the pooled result. Column (4) shows the pooled result, but with the sample restricted to dual-headed households, where we might expect bargaining effects to be at play. Below the main results, 'Test' reports a linear test of the null hypothesis that the two vouchers have equal effects. Across all the columns, general vouchers have a positive, strong and significant effect in all but the Kati specification.[13] General vouchers have a positive effect because they induce households to purchase CROs, many of whom go on to co-title even without conditional incentives. Again, in all columns but (2), conditional vouchers have a strong positive effect, large enough that we are able to reject the null hypothesis of $\beta_G = \beta_C$ in three out of the four columns.

So it seems that, while the basic intervention was itself successful on improving the status quo, imposing conditionality can take it even further. However, while policymakers might see this as an easy, simple way to get women on land titles, the results might make us question the subsequent impact of getting women onto land titles.

4.4.3 Discussion and heterogenous effects

Reconsider the linear probability model (4.4) used to estimate the demand results in the previous subsection:

$$T_i = \alpha + \beta_G v_G + \beta_C v_C + \mathbf{x}_i \delta + \varepsilon_i$$

The results from the randomised voucher intervention have shown us that, given our estimates of β_G and β_C are indistinguishable, the cost of conditionality P^* is effectively zero: in the context of this intervention, small price incentives are sufficient to overcome any resistance to co-titling. This is encouraging from a simple policy perspective, as it seems particularly easy to nudge women onto land titles.

However, the fact that households are so easily nudged into including women suggests that either co-titling does not result in any substantial shifts in bargaining power or

[13] In Kigogo Kati, due to extremely low take-up, only 6% of households purchase a CRO, fill out an application, and include a woman as an owner, suggesting that there may just not be enough observations in this category to provide the necessary precision to reject the null.

that households do not believe that it will, as the results are consistent with little-to-no disutility from titling ($M = 0$). To better understand whether or not households are behaving as if co-titling will have substantial bargaining power effects, we can explore heterogeneity in take-up and co-titling, using baseline characteristics that might proxy for women's ex-ante bargaining power. This also allows us to investigate whether or not conditional vouchers are more successful at inducing certain types of households to co-title (in essence, allowing M and γ to vary across households).

Table 4.7 displays the results from re-estimating the three specifications used before (CRO take-up, conditional co-titling and net co-titling) with the sample restricted to dual-headed households, to focus on households where bargaining power is likely to be a concern. We consider two dummy variables which might proxy for women's current bargaining power: whether or not a woman is considered a default owner of the property, and the share of total household income the female household-head provides. Column (1) shows the aggregate result for take-up and column (2) displays the same specification, but with interactions between the default owner dummy and both voucher values. The results indicate that properties where women are already considered co-owners are significantly less likely to adopt CROs, but are not significantly more or less responsive to voucher allocations, nor do they treat general or conditional voucher values differently. However, the picture changes when we observe conditional co-titling outcomes in column (3), where households with de facto female ownership are substantially more likely to co-title, but are not responsive to conditional vouchers. While conditional vouchers appear to still have a strong positive effect on households without default female ownership, a linear test cannot reject the hypothesis the the two vouchers have an equivalent impact for households with default ownership (Test 2 under column three). Column (4) displays the unconditional, net co-titling outcomes, indicating no substantial differences between households with de facto ownership in either average outcomes nor responsiveness to vouchers.

Columns (5), (6) and (7) repeat this exercise, interacting the head's share of total household income with voucher values. Households in which women provide a greater share of household income are slightly less likely to purchase a CRO, although this effect is not significant at the 10% level. There is also no concrete evidence that these households respond differently to either voucher. However, column (6) indicates households where women provide greater shares of income are significantly more likely to co-title, conditional on purchasing a CRO and are less responsive to gender vouchers.

What are we to make of these results? Households where women appear to have less bargaining power are, overall, more likely to buy CROs and seem no less responsive to general or conditional price incentives. Conditional on purchasing a CRO, these households appear less likely to co-title, but are more responsive to conditional vouchers. Under the assumption that co-titling leads to a substantial shift in bargaining power, it is puzzling that male-dominated households appear happy to sign up women when asked to. Again, this suggests that households are not acting as if they expect co-titling to lead to substantive changes in the status quo. This might be because, as discussed in Section 4.2.2, the legal basis for a change in women's outside options is fairly murky, or because households are indifferent to these changes anyway, perhaps because they are better characterised by a unitary framework. Individual data on each spouse's desires over co-titling outcomes would make it easier to investigate whether or not co-titling is actually a desired outcome by female spouses, but as the conditional voucher intervention was conceived and designed after the baseline survey, these preferences were never recorded.

Finally, even if households do not expect co-titling to have substantial bargaining power impacts, if the co-titling decisions are short-sighted and the legal implications of co-titling are strengthened down the road, then these results are promising: households which have high female-bargaining power appear to be co-titling by default, where male-dominated households are successfully induced to co-title by conditional vouchers, with no observable reduction in demand.

4.5 Conclusion

In this chapter, we presented preliminary results from a land titling experiment in Dar es Salaam, Tanzania, where we use targeted subsides to induce random variation in the price that land-owning households faced when purchasing a land title. In addition, we also created randomised variation in the conditionality of these subsidies, requiring some

Table 4.7: CRO adoption and co-titling, interaction effects - dual-headed households

	(1) Take up	(2) Take up	(3) Co-titling	(4) Net women	(5) Take up	(6) Co-titling	(7) Net women
General voucher	0.00363***	0.00325***	-0.000964	0.00138*	0.00386***	0.000522	0.00180**
	(0.000827)	(0.000898)	(0.00202)	(0.000745)	(0.00106)	(0.00213)	(0.000888)
Conditional voucher	0.00386***	0.00349***	0.00488***	0.00289***	0.00334***	0.00724***	0.00331***
	(0.000742)	(0.000815)	(0.00156)	(0.000719)	(0.000953)	(0.00194)	(0.000820)
Default female owner		-0.217**	0.475***	-0.0637			
		(0.0988)	(0.176)	(0.0932)			
General × default female owner		0.00288	-0.00378	0.00171			
		(0.00230)	(0.00352)	(0.00214)			
Conditional × default female owner		0.00267	-0.00938**	-0.000116			
		(0.00189)	(0.00418)	(0.00171)			
Women's share of hh income					-0.136	0.820**	0.0253
					(0.138)	(0.335)	(0.116)
General × w. share					0.00171	-0.0156**	-0.00123
					(0.00333)	(0.00626)	(0.00217)
Conditional × w. share					0.00449	-0.0116**	-0.0000545
					(0.00274)	(0.00498)	(0.00245)
Constant	0.436***	0.466***	0.680***	0.333***	0.431***	0.545***	0.295***
	(0.0493)	(0.0526)	(0.108)	(0.0454)	(0.0613)	(0.128)	(0.0506)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Test 1: $\beta_G = \beta_C$	0.780	0.782	0.0000439	0.0451	0.596	0.0000212	0.0818
Test 2: $\beta_G + \beta_{G\times X} = \beta_C + \beta_{C\times X}$		0.992	0.936	0.878	0.449	0.00295	0.161
R^2	0.279	0.283	0.190	0.233	0.275	0.275	0.244
Obs	603	603	166	603	519	143	519

Notes: Linear probability model. Dependent variable = 1 if household has fully paid for a CRO and submitted an application with a woman listed on it. First two columns restrict the sample to Barafu and Kati, respectively. Column (3) pools both samples. Column (4) restricts sample to dual-headed households. Robust standard errors. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

households to include a woman on the land title application in order to apply the discount.

Our results strongly suggest that, on average, both general and conditional subsidies have identical impacts on CRO adoption, revealing that households are not deterred by conditionality. Conditional on purchasing a CRO, households which were allocated a conditional voucher were much more likely to include a woman on their title application. These two results, taken together, indicate that small price incentives are an effective means of encouraging *de jure* empowerment of women in the implementation of land titling schemes. However, it remains to be seen whether or not these strictly legal improvements in women's land ownership will result in actual *de facto* improvements in the lives of urban landowners, in particular for the lives of women. The fact that the "price of empowerment" appears to be very low is troubling, as households might be co-titling under the belief that *de jure* improvements in women's land ownership will not translate into real changes in women's household bargaining power. Future iterations of this research will take advantage of follow-up data to determine whether or not co-titling results in any palpable changes in women's welfare.

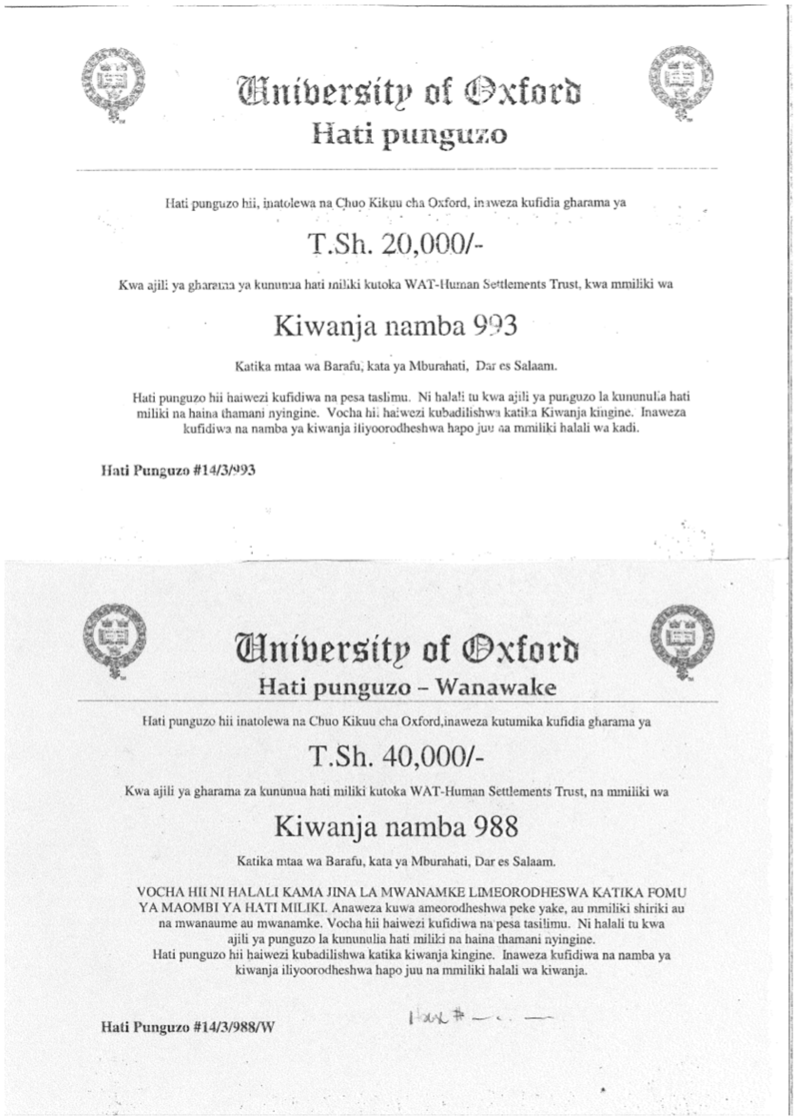
4.A Chapter 4 Appendix

4.A.1 Extra figures and tables

Table 4.8: Tanzania's Land Act of 1999 - provisions relating to spouses

161.-(1)	Where a spouse obtains land under a right of occupancy for the co-occupation, and use of both spouses or where there is more than one wife, there shall be a presumption that, unless a provision in the certificate of occupancy or certificate of customary occupancy clearly states that one spouse is taking the right of occupancy in his or her name only or that the spouses are taking the land as occupiers in common, the spouses will hold the land as occupiers in common and, unless the Presumption is rebutted in the manner stated in this subsection, the Registrar shall register the spouses spouses as joint occupiers accordingly.
161.-(2)	Where land held for a right of occupancy is held in the name of one spouse only but the other spouse or spouses contribute by their labour to the productivity, upkeep and improvement of the land, that spouse or those spouses shall be deemed by virtue of that labour to have acquired an interest in that land in the nature of an occupancy in common of that land with the spouse in whose name the certificate of occupancy or customary certificate of occupancy has been registered.
161.-(3)	Where a spouse who holds Act No. 5 land or a dwelling house for a right of occupancy in his or her name alone undertakes a disposition of that land or of 1971 dwelling house, then- 1. Where that disposition is a mortgage, the lender shall be under a duty to make inquiries of the borrower has or as the case may be, have consented to that mortgage accordance with the provisions of section 59 of the Law of Marriage Act, 1971 2. Where that disposition is an assignment or a transfer of land, the assignee or transferee shall be under a duty to Make inquiries of the assignor Or transferor as to whether the spouse or spouses have consented to that assignment or transfer in accordance with section 59 of the Law Of Marriage and where the aforesaid spouse undertaking the disposition deliberately misleads the lender or, as the case may be, the assignee or transferee as to the answers to the inquiries made in accordance with Paragraphs (a) and (b), the disposition shall be voidable at the option of the spouse or spouses who have not consented to the disposition.

Figure 4.4: Example vouchers, general and conditional



University of Oxford

Hati punguzo

Hati punguzo hii, inatolewa na Chuo Kikuu cha Oxford, inaweza kufidia gharama ya

T.Sh. 20,000/-

Kwa ajili ya gharama ya kununua hati miliki kutoka WAT-Human Settlements Trust, kwa mmiliki wa

Kiwanja namba 993

Katika mtaa wa Barafu, kata ya Mburahati, Dar es Salaam.

Hati punguzo hii haiwezi kufidiwa na pesa taslimu. Ni halali tu kwa ajili ya punguzo la kununulia hati miliki na haina thamani nyingine. Vocha hii haiwezi kubadilishwa katika Kiwanja kingine. Inaweza kufidiwa na namba ya kiwanja iliyoorodheshwa hapo juu na mmiliki halali wa kadi.

Hati Punguzo #14/3/993

University of Oxford

Hati punguzo – Wanawake

Hati punguzo hii inatolewa na Chuo Kikuu cha Oxford,inaweza kutumika kufidia gharama ya

T.Sh. 40,000/-

Kwa ajili ya gharama za kununua hati miliki kutoka WAT-Human Settlements Trust, na mmiliki wa

Kiwanja namba 988

Katika mtaa wa Barafu, kata ya Mburahati, Dar es Salaam.

VOCHA HII NI HALALI KAMA JINA LA MWANAMKE LIMEORODHESWA KATIKA FOMU YA MAOMBI YA HATI MILIKI. Anaweza kuwa ameorodheshwa peke yake, au mmiliki shiriki au na mwanaume au mwanamke. Vocha hii haiwezi kufidiwa na pesa taslimu. Ni halali tu kwa ajili ya punguzo la kununulia hati miliki na haina thamani nyingine.
Hati punguzo hii haiwezi kubadilishwa katika kiwanja kingine. Inaweza kufidiwa na namba ya kiwanja iliyoorodheshwa hapo juu na mmiliki halali wa kiwanja.

Hati Punguzo #14/3/988/W

Figure shows two examples of vouchers which households might have received, indicating the conditionality ("wanawake" is Swahili for "woman"), the amount the voucher was worth, as well as the parcel number for which the voucher would apply (*kiwanja namba*).

Table 4.9: Test of linearity assumption of voucher impacts

	20	40	60	80
General voucher				
20	.	.0678622	.2939097	.3077594
40	.0678622	.	.3028473	.2676995
60	.2939097	.3028473	.	.9613195
80	.3077594	.2676995	.9613195	.
Conditional voucher				
20_	.	.0493371	.3282069	.6497055
40	.0493371	.	.1019254	.0138067
60	.3282069	.1019254	.	.304582
80	.6497055	.0138067	.304582	.

Results taken from regression of take up on a dummy for each general and conditional voucher value. Each cell contains the p-value from a test of linearity between two coefficients. For example, cell (20,40) tests that $2 * \beta\_20 = \beta\_40$ Results are from pooled sample with no baseline controls.

Table 4.10: Effect of voucher distribution on CRO adoption, female-headed households

	Barafu			Kati		
	(1)	(2)	(3)	(4)	(5)	(6)
General voucher (tsh '000)	0.00542***	0.00547***	0.00604***	0.00197**	0.00197**	0.00203***
	(0.00130)	(0.00130)	(0.00132)	(0.000801)	(0.000801)	(0.000782)
Conditional voucher (tsh '000)	0.00602***	0.00618***	0.00631***	0.00221***	0.00221***	0.00230***
	(0.00102)	(0.00103)	(0.00102)	(0.000776)	(0.000776)	(0.000762)
General × female-headed	-0.00607**	-0.00611**	-0.00526*	-0.00183	-0.00183	-0.00252*
	(0.00296)	(0.00296)	(0.00295)	(0.00153)	(0.00153)	(0.00141)
Conditional × female-headed	-0.00566**	-0.00582**	-0.00504*	-0.00313*	-0.00313*	-0.00367**
	(0.00280)	(0.00280)	(0.00271)	(0.00161)	(0.00161)	(0.00157)
Female (only) headed hh	0.344**	0.353**	0.317**	0.0745	0.0745	0.104
	(0.147)	(0.147)	(0.144)	(0.0780)	(0.0780)	(0.0727)
Baseline controls	No	No	Yes	No	No	Yes
Test 1: $\beta_G + \beta_{G\times F} = 0$	0.810	0.810	0.771	0.910	0.910	0.678
Test 2: $\beta_C + \beta_{C\times F} = 0$	0.889	0.889	0.612	0.519	0.519	0.318
R^2	0.0738	0.0769	0.115	0.0221	0.0221	0.0791
Obs	423	421	421	608	608	608

Notes: Linear probability model. Dependent variable = 1 if household has fully paid for a CRO. The first three columns show results using the Barafu sample with column (1) using no unreported controls. Column (2) restricts the sample to households with non-missing controls. Column (3) includes baseline controls. The three columns for Kati sample follow this same pattern. Test 1 displays the p-value from a linear test of the hypothesis that the general voucher effect for female-headed households is zero and Test 2 tests the same hypothesis for conditional vouchers. Robust standard errors $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

4.A.2 Selection model for applications

Several regressions in this chapter rely on CRO application data which is only observable for a subset of CRO-purchasing households who have submitted an application. Consider two equations, the co-titling specification from Subsection 4.4.2 and a selection equation, where A_i is a binary variable equal to one if the household has submitted an application (conditional on purchasing a CRO):

$$cotitle_i = \alpha + \beta_G v_G + \beta_C v_C + \mathbf{x}_i \delta + \varepsilon_i \tag{4.6}$$

$$A_i = \gamma + \mathbf{z}\beta_z + \upsilon_i \tag{4.7}$$

The vector $\mathbf{z}$ comprises observable household characteristics which affect the probability a household submits an application and ordinarily contains all the covariates included in (4.6). If the error terms of these two equations are uncorrelated, then the determinants of selection are random (conditional on the covariates in (4.6)). However, if the unobserved determinants of selection are correlated with the unobserved determinants of co-titling, $cov(\varepsilon_i, \upsilon_i) \neq 0$, then estimates of β_G and β_C will be subject to sample selectivity bias. For example, if male-dominated households are less likely to turn in an application, conditional on purchase, and male-dominated households are less likely to co-title, then coefficient estimates in the co-titling equation are likely to be biased.

To account for this bias, we use a standard Heckman selection model, in which we first estimate equation (4.7) using a probit, then use the predicted values to construct the estimated inverse mills ratio $\frac{\phi(\mathbf{z}\widehat{\beta}_z)}{\Phi(\mathbf{z}\widehat{\beta}_z)}$. If equation (4.6) is subject to sample selection bias, inclusion of the IMR should correct for it (Heckman 1979). However, in practice, if $\mathbf{z}$ only comprises observable characteristics already included equation (4.6), then the inverse mills ratio will be strongly collinear with the pre-existing covariates in the outcome equation. For robust identification, we require an observable characteristic which can be included in the selection equation, but reasonably be excluded from the outcome equation. In this case, our 'instrument' of choice is a dummy equal to one if the household resides on the owned parcel. The NGO tasked with managing the repayment programme found it substantially more difficult to reach and follow up with households living away from their parcels, as these households were often located outside of the neighbourhood. Unsurprisingly, households living off of their owned parcel were much less likely to purchase a land title or submit an application. However, conditional on the purchase decision, we argue that the household's residence status can reasonable be excluded from the *co-titling* equation, as there is no reason to believe that households living on their owned parcel will be more or less likely to include a woman as a landowner.

Table 4.11 displays the results from four separate specifications. Column (1) displays the results from the main OLS specification, estimating the probability of co-titling on the voucher values and a vector of baseline controls. Column (2) uses the Heckman 2-step method of correcting for sample selection bias.[14] Estimated coefficients from the outcome equation are shown in the top half of the table and those from the first-stage selection estimation are shown in the bottom half (baseline controls are not reported). In the selection specification, the estimated coefficients on the voucher dummies are broadly similar to those in the OLS specification. Indeed, the estimated coefficient of the inverse mills ratio is not statistically significant, indicating that selection is not biasing the OLS results. To test for differences using a different function form, columns (3) and (4) display the results from a probit estimation of the outcome equation and a probit model with Heckman sample selection.[15] Again, the results are very similar across the two specifications, and a test of independence between the two fails to reject the null of no difference.

[14]The two-stage procedure is more robust to violation of the assumption of bivariate normal error terms.
[15]This procedure is described in Section 17.4.3 in (Wooldridge 2002)

Table 4.11: Effect of voucher distribution on co-titling, sample selection specification

	(1) OLS	(2) Selection	(3) Probit	(4) Selection Probit
Main Equation				
General voucher	-0.000824	-0.00115	-0.00244	-0.00328
	(0.00122)	(0.00152)	(0.00432)	(0.00438)
Conditional voucher	0.00371***	0.00429***	0.0181***	0.0177***
	(0.00106)	(0.00146)	(0.00539)	(0.00477)
Constant	0.727***	0.896***	0.579**	0.767***
	(0.0644)	(0.135)	(0.242)	(0.241)
Pr(Application submitted)				
General voucher		-0.000943		-0.000500
		(0.00369)		(0.00369)
Conditional voucher		-0.00331		-0.00303
		(0.00339)		(0.00342)
HH lives on parcel		0.386*		0.459**
		(0.204)		(0.193)
Constant		0.608**		0.525*
		(0.281)		(0.277)
λ		-0.435		
		(0.337)		
χ^2 (test of indep)				1.87
Prob > χ^2				0.1711
Baseline controls	Yes	Yes	Yes	Yes
R^2	0.125		0.146	
Obs	264	368	264	368

Notes: Dependent variable = 1 if household includes a woman on CRO application. Column (1) shows results from linear probability model. Column (2) uses Heckman 2-step to account for the selection of submitting an application. All controls are included in outcome and selection equation but not reported. A dummy for whether the household resides on the owner parcel is excluded from the second-stage equation. Column (3) shows the raw coefficients from a probit model. Column (4) shows the results from a selection probit with the same exclusion restriction. Main equation shows results from second stage and Pr(application) shows results from first stage.
Robust standard errors $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Chapter 5

Birth order and child development in the Philippines

5.1 Introduction

In a recent article about promoting social mobility, James Heckman suggested that it was the "accident of birth" that was responsible for most inequality today (Heckman 2012). While Heckman was actually lamenting the repercussions of being born into a disadvantaged family, his comments could equally apply to 'accidents' which lead to inequality *within* households, as many children face relative deprivation solely due to random, exogenous differences between them and their siblings. Within developing countries, gender is frequently considered to be the most salient of these differences, often leading to starkly divergent outcomes, the result of a mix of parental preferences and economic realities (Rosenzweig and Schultz 1983; Garg and Morduch 1998; Qian 2008).

This chapter is concerned with another, less acknowledged source of sibling heterogeneity: birth order, the sequence in which children enter and transition through the household. Birth order has long been noted in the psychology literature as a source of inequity between siblings (Slater 1962; Zajonc and Markus 1975), but has only more recently been seriously considered by economists working on child outcomes. While psychologists have been primarily preoccupied with cognitive development and the interactions between siblings of different ages, economists have largely considered birth order effects to be the result of differing endowments, household constraints or parental investments (Behrman and Taubman 1986).

In this chapter, I investigate birth order effects using a longitudinal survey which followed children in the Philippines from birth until their early twenties, gathering a rich set of data on their health and educational outcomes. Thus, this is the first paper to examine birth order effects over the childhood of a single cohort, rather than using household cross-sectional or panel data to estimate average birth order effects across children of different ages. This not only allows for the existence of age *trends* in birth order effects, but also allows me to investigate whether or not birth order effects are permanent or transitory as children reach adulthood. The existence of data across such a long time period also allows me to construct more accurate, fixed measures of birth order, where most studies are forced to derive a measure based on the household size at the time of interview. This is largely a descriptive chapter: my aim is to precisely estimate differences in outcomes between children of different birth order, although I will consider some of the root causes of these differences.

First, I find that birth order differences in health and educational attainment are largely transitory: while they are apparent for children at a fairly young age, differences appear to be negligible as children reach adulthood. This suggests that studies which estimate average birth order effects *across* ages may be overestimating the impact of birth order on final outcomes. Second, I also find some evidence for both nonlinear and, in some cases, non-monotonic birth order effects. Economic theories of birth order effects largely posit that these differences will be linear and monotonic (what is good for a firstborn child will always be bad for a lastborn child). Finally, I find limited support for liquidity-constraint hypotheses of birth order constraints, where later-born children suffer more because they are born into periods with greater competition for resources. I find little evidence for popular competing theories of birth order effects.

In the next section I will discuss the current empirical evidence behind birth order effects in health and educational outcomes, as well as the prevailing theories for these effects. In Section 5.3 I will discuss the source of the data (taken from the Cebu Longitudinal Health and Nutrition Survey), the construction of birth order measures and my empirical model for estimating birth order effects. In Section 5.4 I will discuss the results over both health and educational outcomes, first using linear measures of birth order, then using nonlinear measures. Section 5.5 covers potential mechanisms behind the birth order effects revealed. Section 5.6 presents some robustness checks to these main results and I conclude with Section 5.7.

5.2 Background and conceptual framework

5.2.1 What do we know about birth order effects today?

Within the economic literature, the study of birth order outcomes began with Peter Lindert's work on 'sibling position' in the US, which revealed that later-born children tended to have lower levels of educational attainment (Lindert 1977). Since then, empirical studies from developed countries have consistently confirmed this result (Behrman and Taubman 1986; Black, Devereux, and Salvanes 2005; Kantarevic and Mechoulan 2006; Conley and Glauber 2006; Booth and Kee 2009; De Haan 2010; Hotz and Pantano 2011).

The evidence gets discernably more mixed when we begin examining birth order outcomes in developing countries. While several studies have confirmed an aggregate negative effect of birth order on educational attainment (Quisumbing 1995; Afridi 2005; Tenikue and Verheyden 2010), slightly more have found the opposite result (Ejrnæs and Pörtner 2004; Ota and Moffatt 2007; Emerson and Souza 2008; Rammohan and Dancer 2008; De Haan, Plug, and Rosero 2012).

The picture becomes a little more consistent when we consider research on birth order differences in health and nutrition in developing countries, where inequities in these outcomes are typically more salient. The general consensus is that later-born children are generally worse off: Horton's 1988 study of children under fifteen in the Bicol region of the Philippines found that children of higher birth order were both more likely to be stunted and wasted than their siblings. Senauer, Garcia, and Jancito (1988) studied caloric intake in three rural Filipino districts, finding that higher birth order children were significantly disadvantaged after controlling for family size. Outside of the Philippines, negative birth order effects on anthropometric outcomes have also been demonstrated by

Behrman (1988) using ICRISAT data, Haughton and Haughton (1997) in Vietnam and Kebede (2005) in Ethiopia.

It is worth noting that, while most developing country studies use retrospective data on family composition to look at outcomes in adulthood, the sort of high quality administrative data needed to discern birth order effects in adulthood is often lacking in developing countries. As a result, nearly all birth order studies in poor countries use household surveys to look at education and health outcomes of children at a relatively early age, thus most research is unable to discern whether or not observed impacts are permanent or transient.[1]

5.2.2 What are birth order theories telling us?

Differences in environment

The most common explanation for negative birth order effects is that they are driven by differing family environments during childhood. Almost by definition, later-born children are born into larger families and spend more time surrounded by their siblings. For a given type of parental investment, such as time or expenditure on food or education, children of a higher birth order must spend more of their childhood competing for these resources with their brethren (Sulloway, Hertwig, and Davis 2002).

This is not a particularly satisfying explanation, as it relies on the assumption that parents either have very little agency, a very simple distribution method (such as dividing total resources by total number of children present) or very high discount rates. Many models of parental preferences allow for concerns over equality between children (Behrman, Pollak, and Taubman 1982; Behrman and Taubman 1986), in which case we might expect parents to equalise child outcomes by shifting resources intertemporally.[2] Indeed, birth order effects tend to be present in contexts where it is difficult for parents to transfer resources across time, such as in child care time (Lindert 1977; Birdsall 1991;

[1] Ejrnæs and Pörtner (2004) are one of the few exceptions, as they look at birth order impacts on completed eduction of children of the Philippines.

[2] For example, parents could save for subsequent children (at the immediate expense of earlier-borns) in order to equalize per-child resources across periods. Or, if fertility is unpredictable and parents receive too many fertility shocks, then a functioning credit system would allow them to borrow to smooth per-child expenditures.

Price 2008). If liquidity constraints exist for other resources, then we might expect birth order discrepancies to be stronger in settings where they are binding. Finally, while the change in sibling structure is the main difference in environment that children of different birth orders face, they also encounter parents at different ages, levels of experience and wealth, all of which are likely to affect child outcomes.

Discrimination between children

What birth order theories based on differences in environment have in common is an implicit assumption that there are no active incentives for parents to discriminate among children. Differences in child outcomes are solely the result of parents' inability to equalise education and health inputs between children across time. Yet, many theories of birth order allow for the returns to investment in children of different birth orders to differ, allowing for parents to explicitly discriminate between them.

For example, early-born children are the first to become old enough to participate in the labour market. As these children age, parents gain the ability to relieve tightening liquidity constraints by pulling these children out of school and sending them to work. Several birth order models have combined Basu and Van's (1998) luxury axiom with age differences between siblings to predict that constrained households will send older, earlier-born children to work at the expense of their educational attainment (Chesnokova and Vaithianathan 2008; Tenikue and Verheyden 2010). This earlier-born work-bias has been verified empirically in a number of different developing-country settings (Edmonds 2006; Emerson and Souza 2008; De Haan, Plug, and Rosero 2012), although a direct connection to household liquidity constraints has yet to be established.

Being the older child in a family might also lead to positive discrimination if children are required to remit to parents in their old age. With wages determined by educational attainment, parents can use their children as a mechanism for transferring income to their old age. If children are, in a sense, retirement funds, then older children will begin paying out at an earlier date, making the returns to educational investment greater for this cohort. Ethnographic work by Medina (1991) suggests that parents in the Philippines give preference to firstborns in exchange for old-age security:

> “The eldest may also be given preference in education and inheritance, with the expectation that he or she will later take care of the aged parents and the younger siblings.”

Tenikue and Verheyden (2010), who explicitly introduce old-age security concerns into a model which also incorporates the luxury axiom, find evidence that early-born children in sub-Saharan Africa enjoy an educational advantage when households are asset-rich, arguing that once liquidity constraints are not a problem, the investment effect dominates.

Predictions from standard theories of birth order effects

What sort of patterns should we expect to observe, given these models of birth order effects? One element that all of these models have in common is that they predict birth order differences in final outcomes, rather than transitory shocks. All birth order effects, be they due to liquidity constraints in parental investment, luxury axiom effects, or direct parental investment due to old age concern, should be reflected in observable differences in outcomes at adulthood. Thus any result showing that birth order effects are transitory might suggest that a simple one or two period parental investment framework is not adequately capturing the channels through which birth order effects emerge.

Also, by frequently relying on stylised models of two-child families, these models imply that birth order effects will be monotonic: if a firstborn child is better off than a third-born child, then a third-born child will not be better off than a second-born child. In part, this is due to limitations in the theoretical models used: most papers rely on two-child setups, which only allow for a bias towards firstborns or lastborns, but no middleborn effects. I will show in Section 5.4 that not only can birth order differences vary across time, but that there is some evidence that these effects are nonlinear and non-monotonic, suggesting that two-child models may not be capturing the full relationship.

Finally, while much of this chapter will be dedicated to testing the permanence and monotonicity of standard birth order models, it is also possible to test the more basic predictions made by these models over which children are disadvantaged (early or later-born), both over different types of outcomes and across different kinds of families. When liquidity constraints are driving the results, children in households which are more likely to be constrained should suffer from larger birth order differences, as parents are unable to transfer resources from low-sibling periods to high-sibling periods, which places a greater burden on later-born children. The Basu and Van-style argument also has a very clear prediction: early-born children should, at any age, be more likely to work and less likely to attend school than later-born children, and that these differences should also be exacerbated for poorer households. I will discuss specifically what results we should expect to see in Section 5.3.5, after I have discussed the data and setting a little further.

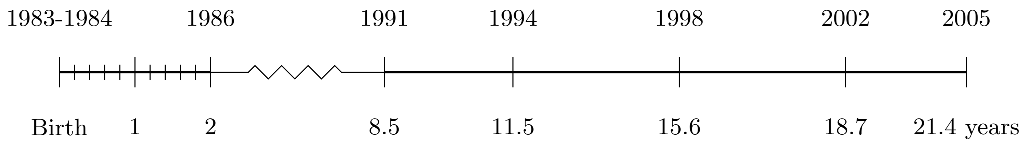
5.3 Data and empirical model

5.3.1 The CHLNS

The data I will be using are from the Cebu Longitudinal Health and Nutrition Survey (CLHNS), an ongoing longitudinal study based in the Filipino island of Cebu. Within the metropolitan area the survey team randomly selected thirty-three barangays (local administrative units) and performed a census of all pregnant women living there. Out of the initial sample of 3,327 women interviewed during the start of their third trimester, 3,080 had pregnancies resulting in single live births, the entirety occurring within 1983 and 1984. After a check-up within the first few days, each child was visited every two months for the first two years of life (to be referred to as rounds 0-12). The CHLNS team then performed follow-up surveys of both mothers and the index children in 1991, 1994, 1998, 2002 and 2005, beginning a few years after the start of schooling and ending a little over the age of twenty-one. Figure 5.1 shows a timeline of the data collection and the average age for the index children during each round. While the CHLNS did gather limited data on the next-youngest sibling in both the 1991 and 1994 rounds, for most rounds the observations are restricted to one child per mother. As a result the analysis will focus on identifying birth order effects across children from different households, rather than within-households.

The survey team took detailed anthropometric measurements of the index children each round of the survey, as well as occasional data on nutritional intake. Other outcomes

Figure 5.1: Timeline of data collection and average age of children



were measured intermittently: schooling and some work information is available from the first follow-up round in 1991.[3] A non-verbal IQ test designed specifically for children in the Philippines (Guthrie et al. 1977) was administered in the 1991 and 1994 rounds, along with standardised math, English and Cebuano tests in the latter.

Attrition in the CHLNS has been significant, in line with other longitudinal studies of this scope (Outes-Leon and Dercon 2008). By the 2005 round, 61% and 67% of the children and mother from the original live-birth cohort remained. While refusal levels were relatively higher in the initial years of the survey (nearly 20% of the sample was lost during the first two-year period), it is primarily migration outside of the sample area which has been driving the attrition. Rather than try and explicitly model the attrition, I will restrict the bulk of my analysis to a sub-sample of those children who are still observed in the final round (as I will explain shortly, this will also become necessary for the construction of relative birth order and several controls). I will revisit the issue of this restriction in the robustness section of this chapter.

5.3.2 Outcomes of interest and birth order

For measures of health status, I use two anthropometric measures: height-for-age and BMI-for-age (body mass index). The former, which I will use for my main analysis, will be useful for examining long term height-outcomes while the latter indicates short-term nutritional status. For the first twelve rounds of data, covering the first two years of life, these measures are standardised using growth curves derived from the 2006 update of the WHO reference curves (WHO 2006). For the subsequent rounds, the 2007 WHO reference is used, which is an amalgamation of data from the Multicentre study and the original NCHS/WHO reference. Each observation is standardised using the growth curves: $z_i = \frac{(a_i - \mu)}{\sigma}$, where μ and σ are the reference mean and standard deviation, respectively.

[3]Children in the Philippines do not typically start school until age 6.

This standardisation is intended to correct for natural gender and age differences in height and weight. Despite this, there is some evidence that the WHO reference may not be entirely appropriate for the Filipino cohort, as the CLHNS data reveals gender differences at birth.[4] While weight-for-height is the more traditional method for gauging short-term nutritional status, BMI-for-age has the advantage of being calculable across all ages of childhood,[5] and corrects for age-dependent changes in body mass (Flegal et al. 2002).

By the penultimate round approximately half the sample is over the age limit for the WHO reference tables, with all children exceeding the age limit in the final round. To ensure than relative height and weight estimates are comparable across rounds, I calculate z-scores for all individuals over 19 using the last available reference mean and standard deviation for each gender.

Figure 5.4 shows the height growth in centimeters of the CLHNS cohort in relation to the WHO reference line. The right vertical axis shows the average height-for-age z-score across the same period. Using the recommended z-score cutoff for stunting of -1.64 (WHO 1995), the average index child is stunted by the age of one. This collapse in relative height continues until the end of first two years, then appears to almost converge with the cutoff by the end of childhood.[6]

Figure 5.2 disaggregates this decline by three sibling position categories: first, last and middleborn, which are any children with a birth order greater than one and less than the total number of children. Firstborns seem to be mildly disadvantaged at birth, a result which is consistent with the work of Miller et al. (1994) which finds that a woman's first child tends to be born lighter and shorter. However, firstborns gain a small, distinct advantage during the first two years of life. As lastborns appear to catch up during the late-childhood recover period, middleborns are at a particular disadvantage.[7]

[4]A t-test for differences in the height-for-age at birth shows that boys are -0.182 (sd .039) standard deviations shorter relative to their reference mean than girls. While it is possible that boys could suffer more from relative deprivation in utero, the negative "boy" effect does not appear to be any stronger in richer/poorer families or those where the mother has higher/lower BMI levels. This suggests that Filipino boys enjoy less of a natural height advantage over girls than the reference population. Consequently, the point estimates of gender differences in the CLHNS cohort should interpreted with some caution.

[5]The WHO weight-for-height standard is restricted to pre-pubescent children.

[6]Recent research has revealed that 'catch-up' growth is actually a common phenomenon, and has been documented before in the CEBU cohort (Hirvonen 2013) and several other studies (Boersma and Wit 1997; Crookston et al. 2010; Dercon et al. 2012; Outes and Porter 2012).

[7]It should be noted that, without family size controls, $E(n_i \mid i = \textit{middleborn}) > E(n_i \mid i \neq \textit{middleborn})$. That is, middleborns are more likely to come from larger families (three or greater by

Figure 5.2: Height-for-age and sibling position across childhood

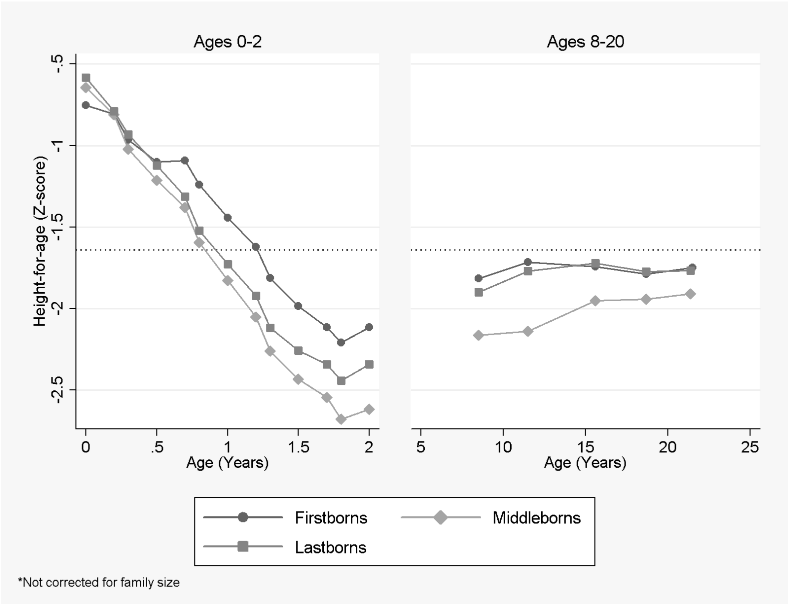


Figure 5.3: Grade attainment and sibling position across childhood

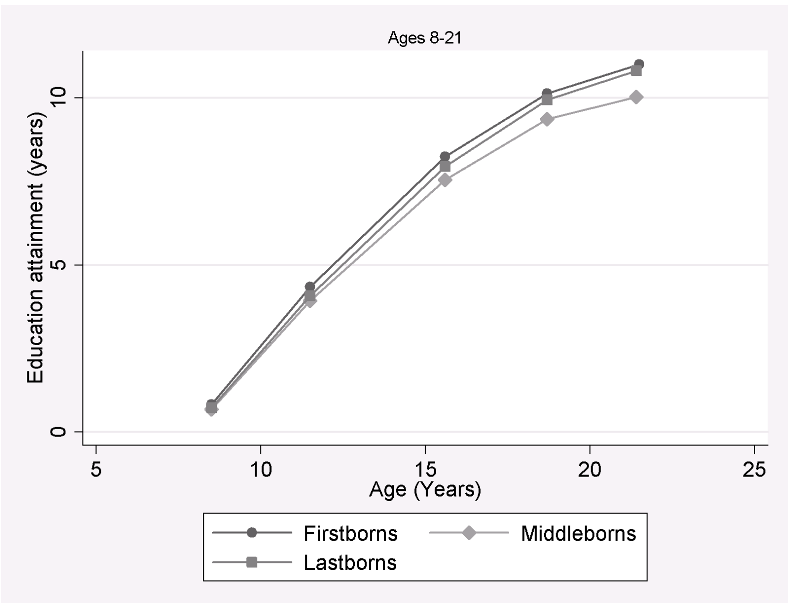
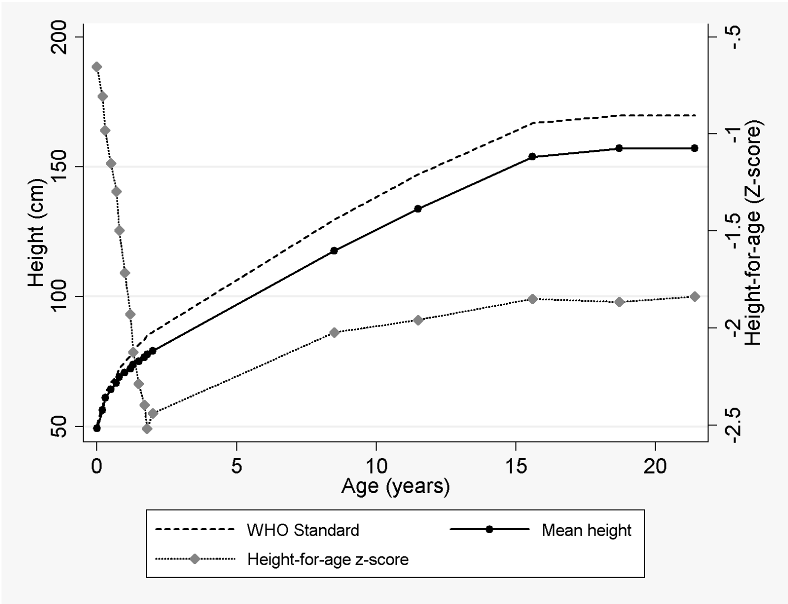


Figure 5.4: Height growth vs WHO standard



My other outcomes of interest include educational attainment, measured by years of completed schooling,[8] and current work/school status, which I will discuss in Section 5.5. Education in the Philippines normally begins between ages six and seven. By the time of the 1991 round, over 90% of the index cohort have started school. By the final round, the average index child has completed 10 years of schooling, which is typically the end of secondary education. Figure 5.3 disaggregates the progress in years of schooling across the five rounds of data for each birth order category. The differences are minute, but again there is a slight middleborn disadvantage. I will investigate these differences further in the next section, where I will specify an empirical model designed to consider factors that are likely to confound the simple story being told in Figures 5.2-5.3.

5.3.3 Empirical model

The empirical specification I will be using to measure order effects is:

$$y_{ijt} = \omega + bo_i\theta + \mathbf{X}_{ijt}\beta_x + \mathbf{D}_j\beta_d + \mathbf{W}_{jt}\beta_w + \gamma_i + \alpha_j + \varepsilon_{ijt} \tag{5.1}$$

default), where first and lastborns can also come from families of size 2. I will discuss this issue further in section 5.4.2.

[8]This does not include repeated grades and counted skipped grades as full years.

Where y_{ijt} is the outcome of interest (health status, educational attainment, work status, etc) for child i in household j at time t. The scalar bo_i is our measure of birth order, which will I will describe in detail below. $\mathbf{X}_{ijt}$ is a vector of child characteristics, including age[9] and gender. $\mathbf{D}_j$ is a vector of time-invariant household characteristics, including the mother's completed fertility (known as sibling or sibship size), average birth spacing, mother's height (cm) and years of education at the beginning of the panel. Average birth spacing is calculated as the average monthly spacing between birth dates for all live births recalled by the mother at the end of the 2005 round.

$\mathbf{W}_{jt}$ comprises time-varying household characteristics, including household size, total monthly household income, land and assets values. Household size is calculated in adult equivalence units used in Dercon and Krishnan (1998), which are based on WHO estimates of age and gender-specific caloric needs. Household income is aggregated from data on individual wage income, crop and livestock income,[10] as well as other transfers such as remittances. As all post-1991 rounds only gathered information on asset stocks, following Filmer and Pritchett (2001) I derive an asset index using principal-component analysis. In the results tables, income, asset values and the asset indices are all standardised to make marginal effects comparable across rounds.

θ is the estimated 'impact' of a change in birth order, and its meaning depends heavily on the measure of birth order used. I will discuss the interpretation of θ more in the next subsection. γ_{ij} is the unobserved child-specific effect, α_j the unobserved household fixed effect, and ε_{ijt} is the child-specific, time-varying error term. For identification of θ, the aggregate birth order effect, the following assumptions need to hold:

$$cov(bo_i, \gamma_i) = 0 \tag{5.2}$$

$$cov(bo_i, \alpha_j) = 0 \tag{5.3}$$

[9]While the CLHNS cohort were all born within a year of each other, the follow-up rounds which began in 1991 took some time to complete, which means that observed varation in ages is somewhat greater than in the initial sample. Because age trend effects can be conflated with birth order effects, it is important that I control for them here.

[10]Crop, livestock and fishing income are calculated after netting expenses and imputing the forgone income from household consumption. A small number of households reported negative farm/livestock income, most likely due to an underestimate of the imputed income measures. These values have been censored at zero.

$$cov(bo_i, \varepsilon_{ijt}) = 0 \tag{5.4}$$

The aim here is to identify a birth order effect which is driven by changes in the household environment or parental investment, but not by innate characteristics of the child or the household. Condition (5.2) requires that the birth order measure be uncorrelated with any innate child endowments. One reason (5.2) might not hold would be the increased likelihood of birth defects in later borns, which would bias estimates of θ downwards. Another possibility is that parents use a fertility stopping rule similar to the one described in Ejrnæs and Pörtner (2004), where parents cease childbearing when they have a child with a high enough endowment 'shock'. If this is the case then $E(\gamma_i \mid bo_i) \neq E(\gamma_i)$, as lastborns would have higher endowments, even after controlling for fertility. However, auxiliary regressions using IQ outcomes (not shown here) suggest there is no lastborn 'bump' in cognitive ability. It is also possible that birth defects, generally thought to be increasing with the mother's age, might be more likely in later-born children.

For health outcomes, another effect that might fall into γ_i is the observed relationship between birth order and birth outcomes in the Philippines, where children of higher order are typically born longer and heavier (Miller et al. 1994). Whether or not this impact is biological is debatable, as older mothers have a higher BMI in the sample. If this contribution to γ_i is orthogonal to household decision-making, then it would be prudent to control for birth outcomes. However, if households systematically divert more resources to mothers who are preparing to give birth to later-borns (perhaps to buffer against being born into a larger family), then this might eliminate an important part of the aggregate impact of birth order on child outcomes. For comparison, I will include a health specification with birth length as a control. Under ideal conditions, (5.3) should hold, as a child's birth order is theoretically uncorrelated with time-invariant household characteristics,[11] an assumption that should hold even without vectors $\mathbf{D}_i$ and $\mathbf{W}_{jt}$ included. However, the CHLNS was not a random sampling of children across metropolitan Cebu, but a within-cluster census of all pregnant women observed during the baseline, so variation in birth order may be correlated with factors that make it more likely that a given child is

[11]Note, birth order *should* be correlated with time-varying characteristics, as birth order is itself a proxy for entering the household at different stages.

observed in the sample. For example, if two households are equal in every way, including the age of the mother, but one of the mothers has ten more years of education, we might expect the latter to have delayed fertility and thus is more likely to be giving birth to a child of lower birth order. While the controls in $\mathbf{D}_i$ and $\mathbf{W}_{jt}$ should bring us closer to satisfying (5.3), is is still possible that observed birth order is the result of selection on household unobservables which are also correlated with y_{ijt}. In the robustness section, I will explore a few methods of eliminating the fixed effect α_j at the expense of cleanly identifying trends over time.

5.3.4 Measuring birth order, interpreting θ and implications for identification

One particular concern for identification is that observed birth order is structurally dependent on the number of children in the family. By default, birth order is top-censored by the number of siblings born in a family, n_j, so that ($bo_{ij} <= n_j$). If we observe a child of birth order one, the probability that child has no siblings is significantly higher than for a child of birth order two. That is, the conditional expectation of the number of children increases with observed birth order: $\frac{\Delta E(n_j|bo_{ij}=y))}{(\Delta y)} > 0$.

As birth order measures are often constructed from family size data, it is usually simple enough to control for the number of siblings in the specification. Despite this, concerns over the structural endogeneity of birth order have led many to use an alternate measure, known as relative birth order (RBO), initially introduced by Slater (1962). Relative birth order is defined as the proportion of siblings that are older than the index child: $\frac{bo_i-1)}{(n_j-1)}$. This transformation purges the birth order measure of its linear relationship with sibship size (the total number of children in the family). Table 5.1 shows the relationship between sibship size, birth order, and relative birth order, which is undefined for sibship sizes of one. When using relative birth order as a measure, the estimated coefficient $\hat{\theta}$ can be interpreted as the effect of moving from a firstborn position (RBO = 0) to a lastborn position (RBO = 1). I will use this measure as my primary indicator of birth order throughout the paper. Since birth order differences do not exist for only children, families which never have more than one child are dropped from all specifications (approximately

Table 5.1: Birth order, family size, and relative birth order

Sibship Size	**Birth order** 1	2	3	4	5	6	7	8	9	10
1										
2	0	1								
3	0	0.50	1							
4	0	0.33	0.67	1						
5	0	0.25	0.50	0.75	1					
6	0	0.20	0.40	0.60	0.80	1				
7	0	0.17	0.33	0.50	0.67	0.83	1			
8	0	0.14	0.29	0.43	0.57	0.71	0.86	1		
9	0	0.13	0.25	0.38	0.50	0.63	0.75	0.88	1	
10	0	0.11	0.22	0.33	0.44	0.56	0.67	0.78	0.89	1

Note: Each cell indicates the relative birth order for a given birth order/sibship size combination.

1.6% of the final sample).

In previous work birth order is usually constructed by ranking by age all observed siblings or all observed children in the household. This proxy of birth order can be problematic for several reasons: usually households have not completed fertility at the time of observation, leading to upward bias of the measure of relative birth order (middleborn children will be labeled as lastborns). If the discrepancy between current and complete fertility is correlated with household unobservables α_j, then estimates of θ will be biased. Furthermore, observed sibship sizes will be smaller than their true value when older children have left to work or form new households, a difference that might also be correlated with household unobservables.

In each follow-up round of the CHLNS, mothers are asked their entire fertility history, giving us complete information on all pregnancies. The measure of sibship size/completed fertility which I will use in the main specification is obtained from the last round in 2005, and is defined as the total number of live births the mother has had. By this point the median age of the remaining mothers was 46.7 years and less than 5% had children under the age of five, implying that observed fertility in this round is approaching completed fertility. For the very few cases where an index child is observed in the final round, but not the mother, the next-available round where the mother reported her fertility history is used. Final birth order is calculated by ranking the live births[12] chronologically and

[12]The main results of this chapter are also robust to using all pregnancies or all children currently alive

identifying the index child's position. All other measures, including relative birth order, are calculated using this method. This results in a unique, constant measure of birth order: a child that is born third into a family but later has two younger siblings will be coded as a middle-born (rbo = 0.5), even though she might have been the last born temporarily.

Later, I will also include dummy variables for firstborns and lastborns, as well as include a squared relative birth order term in order to check if birth order effects are concentrated at the ends of the relative birth order distribution and to test for non-linearity in birth order effects.

5.3.5 Conceptual framework, estimation and descriptive statistics

As it is currently structured, equation (5.1) allows us to test whether or not the predictions of current birth order models hold up in the CEBU context. Recall that liquidity constraint-based explanations typically assume that later-born children will suffer, as by definition they are born into families with larger family sizes. This indicates that estimates of θ will be negative for both health and educational outcomes, especially for resource constrained families. I will test whether or not θ varies across households at risk of being constrained in Section 5.5.1. For education, this would also allow me to discern whether or not birth order effects are being driven by constraints or by direct parental preferences, such as the desire to use early-born children for old-age security, as suggested by Tenikue and Verheyden (2010).

A positive estimate of θ would indicate that early-born children are actually worse off, a result consistent with Basu and Van-style models where these children are sent to the labour market early to relieve liquidity constraints, which should lead to lower grade attainment and possibly even lower nutritional status. There are a host of other reasons why earlier-born children might be worse off, such as inexperienced parents or lower family income,[13] yet we will see that including these factors as controls in (5.1) will not substantially alter the main results.

in the age ranking, instead of live births.

[13] Recall that early-born children arrive earlier in a family's life cycle, thus will face lower total earnings.

Table 5.2: Summary statistics: time invariant characteristics

Variable	**Mean**	**(Std. Dev.)**	**Min.**	**Max.**	**N**
Family size (final)	5.693	(2.79)	1	16	1865
Final birth order	3.383	(2.342)	1	14	1865
Final relative bo	0.508	(0.362)	0	1	1829
Male dummy	0.526	(0.499)	0	1	1888
Average birth spacing (mons)	35.275	(16.563)	10.5	182	1838
Mother's years of schooling	7.438	(3.711)	0	19	1888
Mother's height (cm)	150.798	(5.083)	128.8	169.5	1887

Note that, while θ is specified as time-invariant in (5.1), as well as in all preceding theoretical and empirical models of birth order effects, there is little reason to expect this to be the case. The structure of the CEBU data will allow me to lift this restriction and allow θ to vary across time. In order to do this without imposing a particular function form, I will be estimating specification (5.1) using ordinary least squares *separately* for each round of the panel. Due to the spacing of the survey rounds, there is almost no overlap in age across rounds. Thus, changes in birth order effects across rounds will be interpreted as changes in birth order effects across childhood. Changes in θ across time will approximate age-specific changes in θ as long as there are no cohort-specific time effects which also interact with birth order. For example, if all households in the sample were subject to a single, common shock in a given time period which affected children of a higher birth order more or less than children of a lower birth order, then age changes in θ might be conflated with changes due to time-varying shocks. For the 1991 round, the CLHNS has limited data on the index children's next-youngest siblings. When inspecting the height-for-age trends taken from these data, the same negative age/time trend in height-for-age for index children seems to be present (Figure 5.10 in the appendix). This suggests that the time trends observed in the CLHNS are representative of general age trends, rather than cohort effects.

What does letting θ vary across time accomplish? First, it will allow us to examine whether or not birth order effects are constrained to a particular period in a child's development. Second, it will permit us to test whether or not birth order effects are persistent: while previous studies have estimated birth order effects over children of all ages, keeping θ time-invariant, allowing this parameter to vary round-by-round will reveal whether or not these effects still exist as children enter adulthood. If θ is empirically indistinguishable from zero in the last round of the CEBU panel, theories driven by parental investment decisions (such as the old-age security motive) are less likely to hold water.

In most specifications, I include only index children who are observed at the end of the panel, but include those that drop out intermittently. This maintains a relatively large sample, but introduces concerns over non-random, temporary attrition. In the robustness section, I show that limiting the sample to households which never drop out of the sample does not change the main results of the chapter.

Table 5.2 displays summary statistics for all time-invariant variables, including the controls in D_j. The average completed number of siblings (including index children) is over five children. The average birth order for the index children is 3.3, which puts them in the center of the relative birth order distribution. Table 5.3 includes all time-varying characteristics for select rounds before 1991 and all rounds from 1991 onwards. On average, children obtain just over 10 years of schooling (measured by the number of grades successfully completed) by the time they reach early-adulthood. At this point, over 50% of the index children are working. Information on household assets and land value are not available between the baseline survey and the 1991 follow-up round, so these values are repeated for all pre-1991 rounds.

In the next section I will discuss the results from estimating (5.1), using relative birth order as my main measure of birth order and both height-for-age and grade attainment as my outcomes of interest.

5.4 Main results

5.4.1 Relative birth order, height-for-age and educational attainment

Height-for-age

Table 5.4 shows the results of estimating specification (5.1) for each round separately, using the child's height-for-age z-score as the outcome of interest. Each column indicates a given round, with the average age of the index children listed at the top of the column.

Table 5.3: Summary statistics: time varying characteristics by round

	1983-1986							
	1	6	12	1991	1994	1998	2002	2005
Age (years)	0.2	1.001	2.001	8.486	11.526	15.576	18.694	21.459
	(0.007)	(0.011)	(0.009)	(0.395)	(0.399)	(0.618)	(0.338)	(0.304)
Height-for-age z-score	-.767	-1.695	-2.428	-2.021	-1.970	-1.854	-1.867	-1.838
	(1.121)	(1.138)	(1.118)	(0.995)	(1.029)	(0.834)	(0.814)	(0.816)
BMI-for-age z-score	-.502	-.589	-.165	-.821	-1.091	-.858	-.710	-.374
	(1.093)	(1.004)	(0.943)	(0.892)	(1.102)	(1.055)	(0.995)	(1.030)
Highest grade obtained				0.734	4.082	7.850	9.714	10.412
				(0.518)	(1.065)	(1.961)	(2.582)	(3.146)
Currently in school?				0.933	0.953	0.794	0.374	0.145
				(0.25)	(0.211)	(0.404)	(0.484)	(0.352)
Working for pay?				0.013	0.116	0.189	0.408	0.547
				(0.115)	(0.32)	(0.391)	(0.492)	(0.498)
Monthly income ('000 pesos)	0.074	1.377	1.575	5.694	7.939	13.227	21.299	20.466
	(0.448)	(2.389)	(2.140)	(6.585)	(8.452)	(12.043)	(182.830)	(33.566)
Asset value*	3.708	3.708	3.708	0.121	0.107	0.025	0.033	0.054
	(12.552)	(12.552)	(12.552)	(2.343)	(2.480)	(2.320)	(2.997)	(3.056)
Adult equivalence units	4.890	4.860	4.866	5.521	5.863	5.892	5.873	5.482
	(2.317)	(2.250)	(2.193)	(1.872)	(1.957)	(1.960)	(2.171)	(2.256)
Land value ('000 pesos)	10.373	10.373	10.373	37.721	315.101	225.160	289.937	315.422
	(52.226)	(52.226)	(52.226)	(103.628)	(637.343)	(620.180)	(741.958)	(771.641)
Obs (min)**	1829	1780	1769	1874	1880	1850	1862	1885

* Asset values are defined as thousands of pesos for pre-1991 rounds and as an asset index for 1991 onwards. Standard deviations in (). ** Number of observations listed are the total number for which all time-varying data is available.

To keep the table concise, I have included only a subset of the first twelve rounds. Each row indicates a different set of controls, with row (1) including just a basic set of controls (age, gender and final number of siblings), row (2) extending the controls to mother's education and height, household size, monthly income, assets and the date of interview, and row (3) including the index child's birth length. For example, the point estimate in the third column, second row indicates the estimate of θ (-0.183) when children were, on average, one year old, using the basic set of controls. The point estimates for all controls in the 'full controls" specification can be found in Table 5.17 in Appendix 5.A.1. In each round, I use all observations with non-missing data for the round, conditional on having observable outcomes in 2005. I use this 'unbalanced' set to increase precision, but in Section 5.6 show that the results are robust to restricting the sample to those observed in every round. For each table, I report standard errors robust to heteroskedasticity.[14]

As it is easier to discuss the trends in birth order effects when they are presented graphically, I have used the esimates from Table 5.4, including the omitted rounds, to produce Figure 5.5, which shows how $\hat{\theta}$ changes over time. Point estimates which are darkened indicate those which are significant at the 10% level. The first thing to note is that later-born children start off with an advantage in height-for-age, a result which is consistent with previous studies on this cohort (Miller et al. 1994). This advantage quickly disappears as the children proceed into their first year of life, with later-born children faring worse and worse relative to firstborn children. The peak of this difference is around the first year of life, where estimates of θ fall between -0.2 and -0.3 standard deviations. This is comparable to the gender effects seen during this period, with boys having height-for-age z-scores (shown in Table 5.17 in the appendix) approximately -0.2 standard deviations lower than girls. However, as the children approach their second year of life, this later-born deficit begins to attenuate, and as children approach adolescence, it becomes negligible and insignificant, save for the birth length specification, where a small negative effect (-0.88) is present.

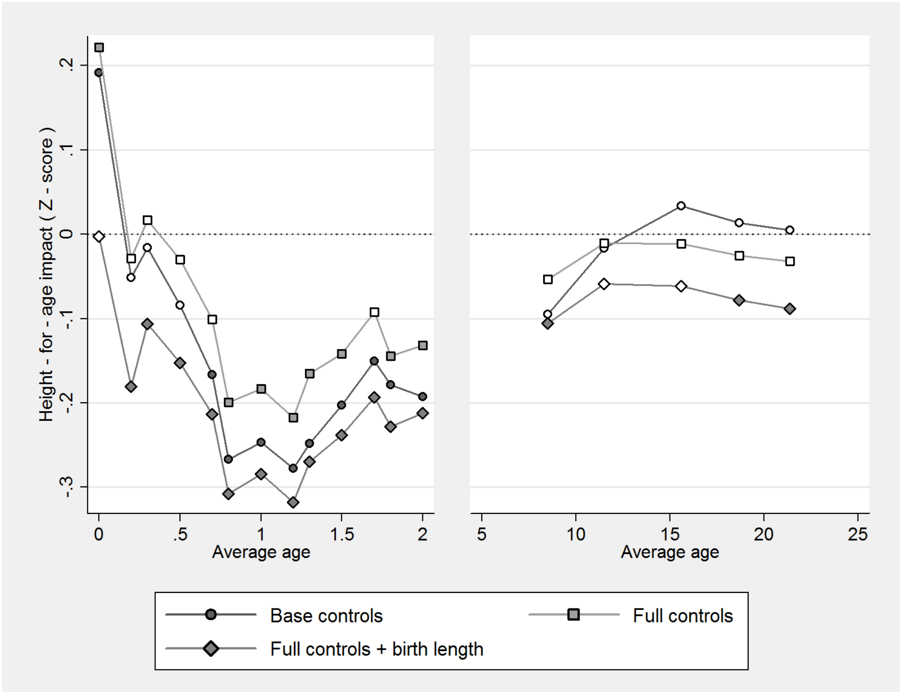
[14] While clustering at the district/barangay level might seem attractive, the limited number of clusters (33) could lead to biased standard errors (estimates of (5.1) with this clustering almost always produce less conservative standard errors). Standard methods of adjusting inference based on the small number of clusters (Moulton 1986) does not qualitatively change the inference presented here.

Table 5.4: The effect of relative birth order on height-for-age z-score

	Rounds: 1983-1985, average age listed below					Rounds: 1991-2005				
	0	.5	1	1.5	2	8.5	11.5	15.6	18.7	21.4
(1) Basic controls	0.191***	-0.0841	-0.247***	-0.203***	-0.193***	-0.0949	-0.0169	0.0335	0.0135	0.00511
	(0.0698)	(0.0712)	(0.0746)	(0.0764)	(0.0722)	(0.0593)	(0.0644)	(0.0535)	(0.0527)	(0.0530)
(2) Full controls	0.221***	-0.0299	-0.183**	-0.142*	-0.132*	-0.0533	-0.0108	-0.0113	-0.0253	-0.0320
	(0.0693)	(0.0688)	(0.0716)	(0.0725)	(0.0696)	(0.0557)	(0.0593)	(0.0485)	(0.0477)	(0.0484)
(3) Full controls + birth length	-0.00240	-0.152**	-0.284***	-0.238***	-0.212***	-0.105*	-0.0589	-0.0615	-0.0781*	-0.0880*
	(0.00407)	(0.0617)	(0.0663)	(0.0680)	(0.0660)	(0.0547)	(0.0582)	(0.0469)	(0.0462)	(0.0465)
Observations	1826	1738	1710	1703	1697	1812	1820	1789	1798	1823

Notes: Coefficients shown are taken from an OLS regression of child's height-for-age z-score on relative birth order. Each column is a separate CLHNS round (with column titles showing the mean age for the sample in that round). Each row is a separate specification: **Basic controls** = [Age, sibling size and gender]. **Full controls** = [Basic controls + mother's education and height, household size, monthly income, land value, asset index and date of interview. Robust standard errors $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Figure 5.5: Linear impact of relative birth order on height-for-age across childhood



Note: Darkened estimates are those which are statistically significant at the 10% level

Table 5.5: The effect of relative birth order on grade attainment

	Rounds: 1991-2005				
	8.5	11.5	15.6	18.7	21.4
(1) Basic controls	-0.109***	-0.214***	-0.165	-0.0717	-0.00684
	(0.0323)	(0.0608)	(0.114)	(0.149)	(0.183)
(2) Full controls	-0.111***	-0.180***	-0.126	-0.00397	0.0170
	(0.0329)	(0.0576)	(0.108)	(0.143)	(0.170)
Observations	1755	1755	1755	1755	1755

Notes: Coefficients shown are taken from an OLS regression of child's grade attainment on relative birth order. Each column is a separate CLHNS round (with column titles showing the mean age for the sample in that round). Each row is a separate specification: **Basic controls** = [Age, sibling size and gender]. **Full controls** = [Basic controls + mother's education and height, household size, monthly income, land value, asset index and date of interview. Robust standard errors $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Educational attainment

Table 5.5 shows the results of estimating the same specification, using the child's highest grade attained as the outcome of interest, and restricting the sample to children observed in every round from 1991 onwards. As children do not enter school until six or seven, the table is restricted to the five rounds between 1991 and 2005 (covering ages 8-22). Again, each column indicates a separate round, with the average age listed at the top, the two rows indicate two different sets of controls. Starting with the first round (age 8.5), later-borns are significantly behind in grade attainment, with lastborns on average having completed approximately -0.1 less grades than firstborns. These differences are even greater in the next round, at close to -0.2 less grades attained. These effect sizes are comparable to the gender effects seen in educational attainment, with boys lagging behind by -0.10 grades in the first round and -0.29 grades in the second. However, as the children enter adolescence, the birth order effect begins to attenuate, while the disparity between boys and girls grows to nearly 1.5 grades less (Table 5.18). Similar to the height-for-age results above, there is no discernable effect of relative birth order on grade attainment by the time children reach adulthood. Table 5.18 in Appendix 5.A.1 shows the results for the full set of controls for row (2).

So while significant, strongly-negative linear birth order effects emerge both in health outcomes and educational attainment, they do not appear to last. In Section 5.5 I will discuss possible reasons for the disappearance of these effects.

5.4.2 Nonlinear birth order measures

The results in the previous section heavily suggest that, while there is a divergence in both health and educational outcomes for children of different birth orders, these differences have largely vanished when children reach adulthood. In keeping with much of the literature, these results have relied on a linear measure of birth order, which only allows for later-born children to be better or worse off on average, but not for any threshold effects or non-monotonicity. In this subsection, I will instead examine effects at the ends of the birth order distribution by substituting relative birth order with two dummy variables equal to one if the child is the first or lastborn child, respectively. In this case the omitted category is a middleborn child. When doing this, I limit the sample to families with completed sibling groups of three or higher. I do this because the omitted category is undefined for families with only two children, and so a firstborn/lastborn dummy might also measure the effect of moving from a two-child family to a three-child or greater family, even after controlling for family size. To test for more general nonlinearities, I also introduce a specification where both relative birth order and its square are included as regressors. For this section, I will refer to the coefficient on the firstborn and lastborn dummies and relative birth order and its square as θ_{FB}, θ_{LB}, θ_{RBO} and θ_{RBO^2} respectively. It should be noted that, while relative birth order is, by construction, structurally independent of family size, nonlinear specifications do not generally satisfy this requirement. This is due to the fact that most observations will be clustered around the values of 0 and 1 in the birth order distribution, as these values are always present in sibling groups of three or more, while intermediate values are not. Similarly there will also be significant clustering around 0.5, as it present in all family sizes comprising an odd number of children, as is evident from Table 5.1. Since I am already controlling for sibling/family size in the specification, this correlation should not be an issue.

Height-for-age

Table 5.6 shows the results for height-for-age, using first and lastborn dummies. The structure of the table is otherwise identical to that of the linear case in Table 5.4 , with columns indicating rounds and pairs of rows indicating specifications with different sets of controls. At the bottom of the table are the results from three linear tests of the hypothesis $\theta_{FB} = \theta_{LB}$, one for each control specification. The results indicate that much of the birth order effect is due to a strong firstborn advantage in height-for-age, which is strongly positive and significant until roughly the age of 11 in rows (1) and (2), and on until adulthood in row (3), where birth length is controlled for.

Throughout the first two years, lastborns are not significantly worse off than middle-borns. However, there is a small, positive effect of being lastborn, which is significant in the 1994 round (at age 11.5). Table 5.19, shown in the Appendix, shows the same specification using relative birth order and its square. Again, while the does not appear to be much evidence for substantial nonlinearities in the first two years of life, the sign of θ_{rbo^2} is significant and negative in several of the subsequent round. However, by the time index children have reached adulthood, the coefficient sizes are small and insignificant in most specifications.

Grade attainment

Table 5.7 shows the results for grade attainment using firstborn and lastborn dummies. There is a distinct firstborn advantage which starts off small, but grows as the children age. While this advantage, equal to approximately 0.2-0.3 extra grades attained, is still present at adulthood, it is not statistically significant in the full control specification, nor is it statistically different from the lastborn effect θ_{LB} which arises in later rounds.

Table 5.8 repeats the specification using relative birth order and its square. During most rounds, the squared term θ_{rbo^2} is significant and of opposite sign of θ_{rbo}, indicating that (given the quadratic structure imposed in the specification) children in the middle of the birth order distribution fare worse in grade attainment, with lastborns catching up to firstborns by adulthood. While the firstborn bias seen in Table 5.7 suggests that birth order effects are nonlinear, this result suggests that effects might not even be monotonic. However, this result is contingent on the quadratic structure imposed here.

Table 5.6: Firstborn/lastborn effects of birth order on height-for-age z-score across childhood

	Dummy	0	.5	1	1.5	2	8.5	11.5	15.6	18.7	21.4
(1) Basic controls	Firstborn	-0.239***	-0.0506	0.178**	0.202**	0.246***	0.203***	0.233***	0.0737	0.0436	0.0488
		(0.0705)	(0.0753)	(0.0765)	(0.0797)	(0.0753)	(0.0603)	(0.0685)	(0.0533)	(0.0526)	(0.0524)
	Lastborn	-0.0106	-0.0438	-0.0953	-0.0323	0.0492	0.0951	0.163**	0.0934*	0.0620	0.0569
		(0.0706)	(0.0714)	(0.0771)	(0.0764)	(0.0711)	(0.0602)	(0.0636)	(0.0549)	(0.0541)	(0.0547)
(2) Full controls	Firstborn	-0.264***	-0.0662	0.145**	0.145*	0.190***	0.160***	0.202***	0.0613	0.0273	0.0450
		(0.0699)	(0.0729)	(0.0719)	(0.0746)	(0.0712)	(0.0551)	(0.0623)	(0.0480)	(0.0465)	(0.0463)
	Lastborn	-0.0138	-0.0154	-0.0767	-0.0433	0.0483	0.0753	0.123**	0.0313	0.00107	-0.000898
		(0.0704)	(0.0692)	(0.0741)	(0.0726)	(0.0682)	(0.0552)	(0.0587)	(0.0489)	(0.0485)	(0.0486)
(3) Full controls + birth length	Firstborn	0.00245	0.0673	0.253***	0.248***	0.285***	0.223***	0.263***	0.118**	0.0891**	0.105**
		(0.00378)	(0.0649)	(0.0666)	(0.0695)	(0.0662)	(0.0539)	(0.0610)	(0.0466)	(0.0445)	(0.0441)
	Lastborn	-0.000309	-0.00821	-0.0636	-0.0301	0.0659	0.0834	0.129**	0.0367	0.00669	0.00191
		(0.00210)	(0.0614)	(0.0680)	(0.0670)	(0.0645)	(0.0528)	(0.0564)	(0.0464)	(0.0466)	(0.0463)
Test (1): $\theta_{FB} = \theta_{LB}$		0.00944	0.939	0.00348	0.0135	0.0266	0.139	0.385	0.763	0.776	0.900
Test (2): $\theta_{FB} = \theta_{LB}$		0.00400	0.554	0.0130	0.0360	0.0980	0.212	0.288	0.611	0.651	0.434
Test (3): $\theta_{FB} = \theta_{LB}$		0.520	0.321	0	0.00100	0.00600	0.0330	0.0640	0.150	0.138	0.0640
Obs		1698	1617	1592	1587	1586	1685	1693	1664	1671	1695

Notes: Coefficients are taken from an OLS regression of child's height-for-age on dummies = 1 if the child is a firstborn or lastborn, first with a set of basic controls [age, sibling size and gender], then including a full set of controls [mother's education and height, household size, monthly income, land value, asset index and date of interview. Each column shows the results for a given round. The FB = LB columns show the p-values from a linear test of equality between the firstborn and lastborn coefficients. Robust standard errors $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

5.5 Mechanisms and discussion

In the previous discussion, I showed that there are strong, significant later-born deficits which emerge in both height and in educational attainment, results which are consistent both with liquidity-constraint theories and with direct parental-preferences for early-born children. However, these later-born deficits attenuate as children get older, which is inconsistent with basic models of parental investment, where adult outcomes should be the cumulative outcome of childhood inputs. While this calls into question some theories of direct investment, such as Tenikue and Verheyden's (2010) old-age security model, it might still be consistent with liquidity-constraint models if, for instance, later-born children suffer temporarily due to increased resource competition with other siblings,[15] but parents eventually offset this by directing more resources to these children later on, either directly or indirectly by sending older children to work.

In this section, I will investigate some of the possible channels for these effects a little further, tying them to existing theories of birth order along the way. First I will investigate whether or not these effects are strong or weaker for households which are *likely* to be liquidity constrained, using the household asset index as a proxy for wealth. Then I will investigate whether or not parental inputs to health appear to follow the same pattern seen in health outcomes. Finally, I will investigate whether or not early-born children are indeed being used to relieve household constraints and how later-born children eventually catch up.

5.5.1 Liquidity constraints and birth order effects

If the results in Section 5.4 are being driven by liquidity constraints, we would expect later-born (or middle-born) children to be at a greater disadvantage when they in households which are likely to be more constrained. To investigate whether or not this is the case, I will rely on the asset index described in Section 5.3.3 as a proxy for household wealth, itself a proxy for how constrained the household should be. I then proceed with both

[15] I also showed evidence that these effects might not be strictly linear in the long run, as there is some evidence that middleborn children fare worse in the post infancy rounds (after the age of two). Competition might be worse for middle-born children who have to share resources with both older and younger siblings.

Table 5.7: Firstborn/lastborn effects on grade attainment, rounds 1995-2005

	Dummy	8.5	11.5	15.6	18.7	21.4
(1) Basic controls	Firstborn	0.0756**	0.236***	0.345***	0.337**	0.338*
		(0.0337)	(0.0571)	(0.111)	(0.145)	(0.181)
	Lastborn	-0.0507	-0.0337	0.0695	0.129	0.123
		(0.0312)	(0.0637)	(0.117)	(0.150)	(0.184)
(2) Full controls	Firstborn	0.0685**	0.191***	0.295***	0.246*	0.211
		(0.0333)	(0.0519)	(0.106)	(0.134)	(0.160)
	Lastborn	-0.0533*	-0.00776	0.131	0.178	0.121
		(0.0302)	(0.0589)	(0.112)	(0.147)	(0.171)
Test (1): $\theta_{FB} = \theta_{LB}$		0.00126	0.000106	0.0351	0.223	0.314
Test (2): $\theta_{FB} = \theta_{LB}$		0.00173	0.00201	0.196	0.681	0.653
Obs		1634	1634	1634	1634	1634

Notes: Coefficients are taken from an OLS regression of child's grade attainment on dummies = 1 if the child is a firstborn or lastborn, first with a set of basic controls [age, sibling size and gender], then including a full set of controls [mother's education and height, household size, monthly income, land value, asset indx and date of interview. Each column shows the results For a given round. The FB = LB columns show the p-values from a linear test of equality between the firstborn and lastborn coefficients. Robust standard errors $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table 5.8: Nonlinear effects of birth order on grade attainment, rounds 1995-2005

	Coefficient	8.5	11.5	15.6	18.7	21.4
(1) Basic controls	RBO	-0.258**	-0.762***	-1.194***	-1.132**	-1.254*
		(0.122)	(0.228)	(0.434)	(0.559)	(0.689)
	RBO^2	0.148	0.540**	1.051**	1.081**	1.246*
		(0.115)	(0.225)	(0.425)	(0.544)	(0.666)
(2) Full controls	RBO	-0.246**	-0.684***	-1.167***	-1.003*	-0.956
		(0.121)	(0.206)	(0.410)	(0.518)	(0.607)
	RBO^2	0.132	0.499**	1.076***	1.049**	1.020*
		(0.112)	(0.203)	(0.403)	(0.511)	(0.591)
Observations		1634	1634	1634	1634	1634

Notes: Coefficients are taken from an OLS regression of child's grade attainment on relative birth order (RBO) and RBO^2 , first with a set of basic controls [age, sibling size and gender], then including a full set of controls [mother's education and height, household size, monthly income, land value, asset indx and date of interview. Each column shows the results for a given round. Robust standard errors $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

the basic linear birth order specification and the firstborn/lastborn dummy specification, interacted with the household asset index, to see if there is any heterogeneity in birth order effects.[16]

Table 5.9 reveals the height-for-age results from these asset index interactions, both for the linear birth order and the firstborn/lastborn specifications. The coefficients on the interactions are neither significant nor consistently positive or negative. This suggests that, at least for nutrition, liquidity constraints are not driving the birth order effects observed before, or that households of different wealth levels are all facing the same constraints with regards to nutritional inputs. This in itself presents a puzzle: if later-born children are at a temporary nutritional advantage, but are not any worse off in poorer families, when what is driving these transitory effects in the first place?

Table 5.10 repeats the same procedure using grade attainment data. Again, in the linear case, even though assets appear to be a reasonably strong predictor of grade attainment, there appears to be little consistent heterogeneity in birth order effects across households of different asset levels. However, in the nonlinear case, the firstborn advantage revealed previously in Table 5.7 is significantly reduced in high asset households. This is, again, at odds with Tenikue and Verheyden's (2010) model in which high asset households, undeterred by liquidity constraints, invest more in firstborn children. Instead, the result here is broadly consistent with a liquidity-constraint story, where firstborns have an advantage in education which is purely due to a household's inability to transfer resources from periods with just one child to periods where there are many siblings. Finally, this result also suggests that the luxury axiom is not at work here, or is dominated by other factors, as early-born children appear more likely to go to school. I will investigate this further in Section 5.5.3, where I will examine actual labour outcomes for the CEBU cohort.

[16]Following Tenikue and Verheyden (2010), I have also done this with asset tertiles, which reveal the same basic results.

Table 5.9: Height-for-age effects, asset interactions

	0	.5	1	1.5	2	8.5	11.5	15.6	18.7	21.4
Linear										
Relative birth order	0.228***	-0.0309	-0.184**	-0.148*	-0.136	-0.0444	0.000788	-0.00910	-0.0264	-0.0326
	(3.35)	(-0.45)	(-2.58)	(-2.05)	(-1.90)	(-0.79)	(0.01)	(-0.19)	(-0.55)	(-0.67)
HH asset index	0.0556	0.0472	0.0448	0.0135	0.0110	0.144***	0.256***	0.0923**	0.0580	0.0641
	(1.20)	(1.04)	(0.98)	(0.27)	(0.26)	(3.69)	(6.20)	(2.79)	(1.78)	(1.81)
RBO × Assets	-0.157	0.0505	0.0400	0.142	0.0837	-0.0650	-0.107	-0.000433	0.0431	0.0392
	(-1.62)	(0.52)	(0.40)	(1.42)	(0.86)	(-1.22)	(-1.89)	(-0.01)	(0.99)	(0.89)
Nonlinear										
Firstborn dummy	-0.276***	-0.0804	0.131	0.139*	0.181**	0.165**	0.201***	0.0793	0.0478	0.0605
	(-4.10)	(-1.20)	(1.90)	(2.01)	(2.74)	(3.09)	(3.49)	(1.69)	(1.04)	(1.33)
Lastborn dummy	-0.00338	-0.0301	-0.0680	-0.0353	0.0669	0.0812	0.130*	0.0312	-0.00775	0.000366
	(-0.05)	(-0.46)	(-1.02)	(-0.53)	(1.03)	(1.58)	(2.33)	(0.67)	(-0.17)	(0.01)
HH asset index	0.00456	0.0574*	0.0529	0.0593	0.0263	0.102**	0.189***	0.0912***	0.0554	0.0770*
	(0.15)	(2.05)	(1.84)	(1.76)	(0.98)	(3.19)	(5.43)	(3.30)	(1.85)	(2.45)
FB × assets	0.0254	0.0667	0.0452	-0.00152	0.0723	0.0492	0.0810	-0.0179	0.0153	-0.0171
	(0.33)	(0.86)	(0.58)	(-0.02)	(0.97)	(0.97)	(1.51)	(-0.41)	(0.36)	(-0.39)
LB × assets	-0.152	0.0401	0.0531	0.0925	0.128	0.000161	-0.0177	0.0266	0.0785	0.0390
	(-1.84)	(0.48)	(0.61)	(1.05)	(1.53)	(0.00)	(-0.34)	(0.62)	(1.87)	(0.97)
Observations	1826	1740	1712	1704	1697	1814	1822	1791	1800	1825

Notes: Linear columns show coefficients taken from OLS regression of child's height-for-age z-score on relative birth order, a household asset index, interactions between the two and the set of full controls listed in Table 5.4. **Nonlinear** columns show an equivalent specification using firstborn and lastborn dummies. Each column is a separate CLHNS round (with column titles showing the mean age for the sample in that round) Robust standard errors $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 5.10: Grade attainment effects by asset tertile

	8.5	11.5	15.6	18.7	21.4
Linear:					
Relative birth order	-0.114***	-0.192**	-0.143	-0.0155	0.0139
	(-3.39)	(-3.09)	(-1.22)	(-0.10)	(0.08)
Asset index	0.0686**	0.102*	0.538***	0.688***	1.208***
	(2.96)	(2.38)	(6.81)	(6.78)	(9.33)
RBO × assets	0.0321	0.0932	0.115	0.0218	-0.178
	(1.02)	(1.58)	(1.04)	(0.16)	(-1.11)
Nonlinear:					
Firstborn dummy	0.0523	0.174**	0.266*	0.232	0.191
	(1.65)	(2.89)	(2.38)	(1.64)	(1.15)
Lastborn dummy	-0.0690*	-0.0425	0.0580	0.126	0.109
	(-2.27)	(-0.73)	(0.53)	(0.89)	(0.66)
Asset index	0.0989***	0.192***	0.717***	0.940***	1.369***
	(5.19)	(5.29)	(10.87)	(10.15)	(11.99)
FB × assets	-0.0563	-0.174**	-0.369***	-0.501***	-0.399*
	(-1.89)	(-3.12)	(-3.55)	(-3.85)	(-2.52)
LB × assets	-0.0162	-0.0280	-0.199	-0.431***	-0.540***
	(-0.56)	(-0.52)	(-1.94)	(-3.31)	(-3.71)
Observations	1757	1757	1757	1757	1757

Notes: Linear columns show coefficients taken from OLS regression of child's grade attainment on relative birth order, a household asset index, interactions between the two and the set of full controls listed in Table 5.4. **Nonlinear** columns show an equivalent specification using firstborn and lastborn dummies. Each column is a separate CLHNS round (with column titles showing the mean age for the sample in that round) Robust standard errors $^{*}p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$

5.5.2 Short term nutritional status and breastfeeding

So far, the results indicate that later-born children fare worse in long term health outcomes, as measured by height-for-age. However, this tells us very little about the timing of short-term nutritional inputs. To investigate this further, I use data on BMI-for-age, as well as the mother's breastfeeding behaviour for the first two years of the child's life.

For BMI-for-age, I repeat the basic specification from Section 5.4.1, regressing BMI-for-age on relative birth order for each round. The table of these results results are available in Table 5.20 in the appendix, with the estimate of θ_{RBO} graphed across time in Figure 5.6 below. Later-born children experience a large drop in BMI-for-age in the first few months of life, which dovetails with the height-for-age results presented before. After this initial drop, later-born children appear to recover relative to early-born children, with no perceivable difference in BMI-for-age in the second year of life. The recovery might even lead to an advantage later in life as later-born children fare better than earlier born children around the age of 8-10 (when they are approximately 0.141-0.160 standard deviations above the reference mean). Again, there is convergence as children move into adolescence and adulthood. Liu, Mroz, and Adair (2009) recently showed that parents in the CLHNS panel identify and compensate children whose health outcomes are flagging. The weak boost in relative BMI-for-age seen here might be the result of such a parental compensation strategy.

Another mechanism worth exploring is breastfeeding behaviour, one of the methods used by Cebu parents in the Liu et al. (2009) study. Aside from acting as an input which parents target to improve child health, breastfeeding also has a more mechanical relationship with birth order. A number of studies have established an inverse relationship between breastfeeding and short-term fecundity (Rous 2001; Blackburn 2007), suggesting that early-weaning is a potential strategy for women wishing to get pregnant again. This trade-off can result in quite powerful health effects on children: Jayachandran and Kuziemko (2011) showed that Indian mothers weaned girls earlier to make room for another pregnancy (presumably to increase the chance of having a boy) and that breastfeeding increased as mothers approached ideal fertility. This leads to later-borns being weaned later and lastborns being breastfed the longest, as there are no subsequent children

Figure 5.6: Impact of relative birth order on BMI-for-age across childhood

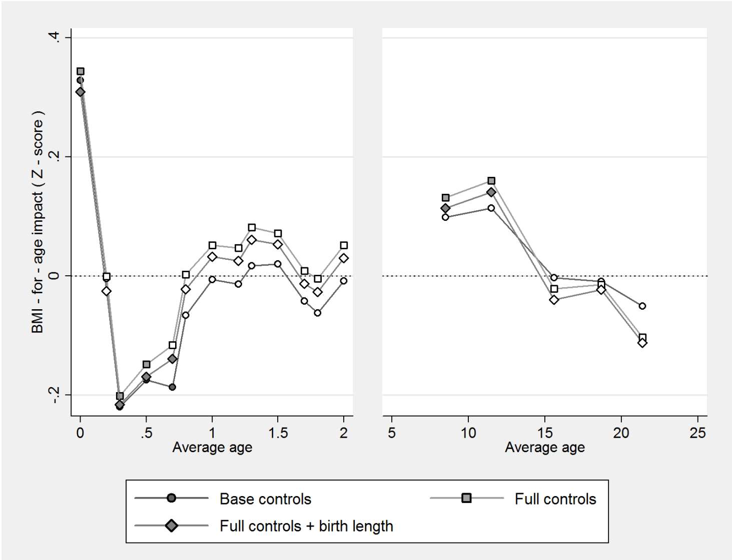
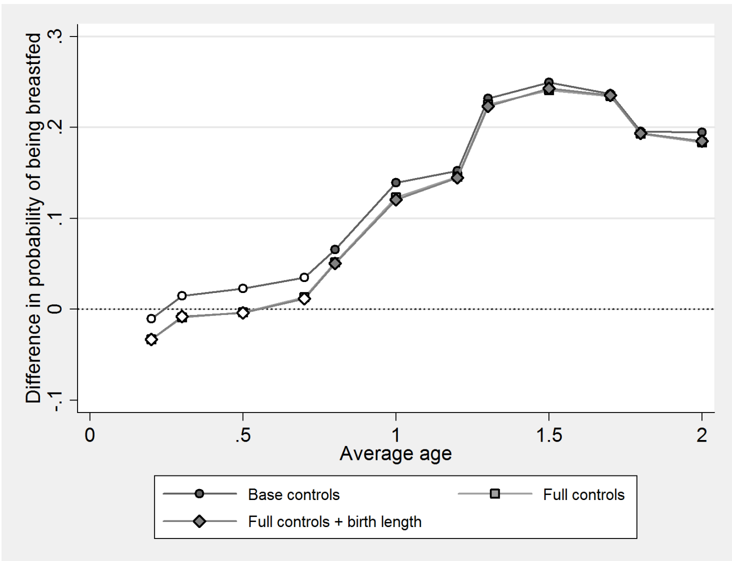


Figure 5.7: Probability of being breastfed in the last day, age 0-2 years



Note: Darkened estimates are those which are statistically significant at the 10% level to encourage an early weaning.

Table 5.21 in the appendix shows the results of estimating the probability[17] of an index child being breastfed in a given round on (1) relative birth order and (2) first and lastborn dummies. Figure 5.7 above graphs the estimates of θ across time. Beginning around the first year of life, later-born children become significantly more likely to be breastfed, with the predicted probability that a lastborn is breastfed approaching 20% by the time the index children are two. This is not purely a fertility-stopping effect, as middleborns are also more likely to be breastfed than firstborns, although the effect is the strongest for lastborn children. This might not only explain why linear birth order effects appear to decline as children get older, but also why nonlinear effects suggest that middleborns take longer to recover than lastborns. There is little evidence that the effects of birth order on either BMI or breastfeeding behaviour vary by household wealth, again suggesting that health effects might not be driven by liquidity constraints.

5.5.3 Delayed schooling and child labour

The results in Section 5.5 suggested that linear birth order effects in grade attainment emerge early, but are not present when children reach adulthood. This catchup is only possible if later-born children delay the start of school, but either progress more quickly through grades or stay in school longer. This is illustrated in Table 5.11, which displays the results of a probit regression on the index child's probability of attending school on relative birth order (specification 1) and first/lastborn dummies (specification 2) for each round separately.[18] The results indicate that early-borns are much more likely to be attending school in the first round that we observe schooling outcomes, but later/lastborns are much more likely to be attending school in the *last* round schooling outcomes are observed. This suggests that firstborns and other early-born children both begin and end schooling earlier. This is backed up by a series of auxiliary regressions of birth order on a variety of schooling attendance outcomes (Table 5.12), which indicate that later born children are less likely to begin kindergarten and are more likely to start school at a later

[17]This specification uses a simple linear probability model for ease of interpretation.
[18]Regression results show raw coefficients, rather than marginal effects.

Table 5.11: Birth order effects on school attendance/work probabilities, rounds 1995-2005

		8.5	11.5	15.6	18.7	21.4
(1) Attending school	RBO	-0.341*	0.0342	0.0153	-0.0531	0.209*
		(0.174)	(0.177)	(0.113)	(0.101)	(0.125)
(2) Attending school	Firstborn	0.413**	0.277	0.175	0.0240	-0.0101
		(0.203)	(0.212)	(0.112)	(0.0925)	(0.114)
	Lastborn	0.0232	0.0479	0.168	-0.0178	0.236**
		(0.142)	(0.156)	(0.113)	(0.0931)	(0.109)
(3) Working for pay	RBO	0.0346	-0.180	0.0675	0.00136	0.125
		(0.239)	(0.120)	(0.108)	(0.0952)	(0.0955)
(4) Working for pay	Firstborn	-0.231	0.0593	0.0280	0.155*	-0.124
		(0.265)	(0.111)	(0.102)	(0.0885)	(0.0877)
	Lastborn	0.0228	-0.158	0.125	0.111	-0.0288
		(0.211)	(0.114)	(0.101)	(0.0876)	(0.0869)
Observations		1585	1585	1585	1584	1585

Notes: Coefficients are taken from an probit regression of child's probability of attending school/working for pay in a given round. Specifications (1) and (3) use linear birth order, while specifications (2) and (4) use firstborn/lastborn dummies. Each column indicates a different round/average age. Full set of controls included (see Table 5.5) for list. Results shown are raw coefficients, not marginal effects.
$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

age, but are no more likely to repeat grades or miss years of schooling.

Rows (3) and (4) of Table 5.11 display the results of a probit regression of the probability that a child is working for pay in a given round. Except for one round in row (4), which suggests firstborns are more likely to work at roughly age 19, there is no consistent evidence of any birth order effects on working probabilities. This stands in contrast to the birth order theories and evidence presented in Section 5.2.2, which indicated that older siblings should be more likely to be sent into the labour force at any given point in time. Even when birth order is interacted with variables which might predict liquidity constraints, such as the asset index, it does not appear to have a consistent effect on the probability of working. If the luxury axiom is a factor in the Cebu context, it does not appear to be interacting with birth order. Thus, it appears that birth order differences in educational outcomes appear to be generating primarily through the tightening of household resources, rather than parents making strategic investments in different children.

Table 5.12: Relative birth order and other schooling outcomes: probit estimation

	Attended Preschool (1)	Started late (2)	Repeated grade(s) (3)	Missed year(s) (4)
main				
Relative birth order	-0.286***	0.330***	0.0879	0.0604
	(0.0937)	(0.105)	(0.0993)	(0.189)
Male dummy	-0.177***	0.0969	0.441***	0.337**
	(0.0657)	(0.0721)	(0.0697)	(0.142)
Family size (final)	-0.0363**	0.0576***	0.0489***	0.0570
	(0.0173)	(0.0170)	(0.0167)	(0.0363)
Age (mons)	-0.0154***	0.00467*	0.0138***	0.0158***
	(0.00278)	(0.00278)	(0.00284)	(0.00533)
Mother's education (years)	0.0830***	-0.0428***	-0.0493***	-0.0583**
	(0.0111)	(0.0125)	(0.0118)	(0.0236)
Birth spacing (mons)	0.00156	0.00232	-0.00411	-0.00220
	(0.00204)	(0.00262)	(0.00268)	(0.00448)
HH income (monthly)	0.0625	0.0308	-0.00558	-0.0366
	(0.0543)	(0.0386)	(0.0540)	(0.139)
Asset index	0.421***	-0.223***	-0.212***	-0.269*
	(0.0482)	(0.0538)	(0.0514)	(0.156)
Land/house value	0.0348	0.0105	0.0493	0.00251
	(0.0338)	(0.0383)	(0.0346)	(0.0767)
HH size (AEU)	-0.0328	0.0155	-0.00611	0.00195
	(0.0224)	(0.0250)	(0.0235)	(0.0574)
Constant	5.099***	-3.044***	-5.650***	-7.817***
	(0.993)	(1.002)	(1.024)	(1.956)
R^2	0.205	0.079	0.102	0.134
Obs	1819	1819	1819	1819

Notes: Table uses data from 1994 round (age 11.5) of survey. Raw coefficients shown, not marginal effects. $^{*}p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$

5.6 Robustness

5.6.1 Attrition and selection

The results presented in this chapter are defined over the sample of children observed in the final 2005 round of the CLHNS. This is necessary as my main measures of birth order and completed family size are calculated using retrospective data from this round, but results in roughly 40% of the original CEBU cohort being dropped from the main analysis. The primary reason for this is survey attrition, with most children dropping out of the sample during the first two years of life. If this attrition is purely random, then estimation of equation (5.1) for each round should be possible without concerns for sample selection bias. However, it is possible that attrition is correlated with child or household-level characteristics. Despite this, to date no other study working with CLHNS data has made a serious effort to correct for attrition bias.

For any given round in my analysis, consider two equations, the first being a simplified version of my main outcome equation (5.1):

$$y_i = \omega + \mathbf{X}_i\beta + \varepsilon_i \tag{5.5}$$

$$S_i = \gamma + \mathbf{Z}_i\delta + \upsilon_i \tag{5.6}$$

The second equation is a selection equation, with $S_i = 1$ if the child in question is observed at the end of the panel, thus being included in my sample. The vector $\mathbf{X}_i$ in the outcome equation comprises all characteristics that affect the outcome y_i. Note that the outcome equation (5.5) is only defined over children who are selected in $S_i = 1$. The vector $\mathbf{Z}_i$ comprises observable characteristics which are thought to affect selection into the sample, and usually includes all characteristics in $\mathbf{X}_i$, if not more. How we approach the attrition problem depends on our assumptions as to what belongs in $\mathbf{X}_i$ and $\mathbf{Z}_i$, and whether or not the error terms in equations (5.5) and (5.6) are correlated.

If the two are correlated ($cov(\varepsilon_i, \upsilon_i) \neq 0$), then the attrition problem is one of selection on unobservables (Fitzgerald, Gottschalk, and Moffitt 1998). Even if the characteristics in $\mathbf{X}_i$ are exogenous (independent of ε_i), correlation between the error terms will lead to bias in estimates of (5.5). Under the assumption that the errors are bivariate normal, the bias can be represented as an omitted variable problem, where the error term contains the inverse mills ratio $\frac{\phi(\mathbf{Z}_i\delta)}{\Phi(\mathbf{Z}_i\delta)}$. Because $\mathbf{Z}_i$ contains elements in $\mathbf{X}_i$, the latter will be correlated with the error term in the outcome equation. The standard method for dealing with this type of selection problem is to first estimate the inverse mills ratio (IMR) by recovering the predicted values from a probit estimation of equation (5.5), then including the estimated IMR as a control in the outcome equation (Heckman 1979).

This approach becomes difficult in the current setting for several reasons. First, due to attrition, time-varying characteristics used in the outcome equation $\mathbf{X}_i$ cannot be included in the selection equation, as they are not observed for those who have left the panel (e.g. household income in round 10 is only observed for households still in the sample at this period). We could include proxies for $\mathbf{X}_i$ in $\mathbf{Z}_i$, by including all time-varying characteristics observed in the baseline round, before any households have a chance to leave the sample. However, there are still two elements of $\mathbf{X}_i$ which cannot be included in the selection equation: both relative birth order and completed family size are constructed using data from the final round, thus they are only defined for households where $S_i = 1$. If birth order is correlated with exit from the sample, then excluding it from the estimation of (5.6) will lead to biased estimates of θ.

There is one reason to suspect that birth order can be reasonably excluded from the selection equation: recall that the average observed relative birth order for the selected sample is 0.508 (from Table 5.3). This is very close to relative birth order's population mean of 0.5.[19] If relative birth order was a correlate of attrition or selection out of the sample, we would expect the average relative birth order for the selected sample to deviate strongly from 0.5.

Despite this, there are two types of selection that might be of immediate concern: mortality and migration. In the first case, if, for instance, laterborn children are more likely to die in infancy, then this will induce a positive correlation between relative birth order in the selected sample with unobservable determinants of child survival, resulting in

[19]Assume we observe a child randomly chosen from a sibling set of size N_i. The expected birth order of a child chosen at random is $\frac{N_i+1}{2}$. It follows that the expected value of relative birth order is given by $E[rbo_i] = \frac{\frac{N_i+1}{2}-1}{N_i-1} = 0.5$

an upward bias in θ. To investigate this further, I turn to the fertility histories reported by mothers of index children in the 2005 round. As mothers list the outcomes of all pregnancies, I can compare average relative birth order for the sample of all index children born alive to the selected sample of those still alive in 2005. As this data requires that at least the mother be interviewed, it is still subject to selection of mothers out of the sample, but it is perhaps more reasonable to assume that mother's selection is uncorrelated with the index child's birth order. Table 5.13 shows the results of both OLS and probit regressions of the probability that an index child is reported by mothers as having died on relative birth order. Although the point coefficient is positive, indicating laterborn children are more likely to die during childhood, the effect is not statistically significant at standard levels of inference. Furthermore, even if laterborn children are more likely to die, this would work against the main result of the chapter.

The second type of selection which is of concern is out-migration of children of higher or lower birth order. Such migration could happen because families send older children away to work, or through primogeniture, with older children taking up possession of family land elsewhere. Again, I turn to the final round of the CLHNS, where whenever either a mother or an index child was reached by the survey team, data on the index child's whereabouts was recorded. While the results are not reported in this chapter, there appears to be no correlation between the index child's birth order and her propensity to either be living separately from the mother's household or leave the survey area live separately from the mother or having moved outside of the district. This suggests that, at least for migration, birth order does not appear to be a significant factor for selection.

Given that relative birth order is likely to be orthogonal to selection anyway, it might be reasonable to rely on a less restrictive set of assumptions. Under the assumption of selection on observables, ε_i and υ_i are assumed to be uncorrelated once we condition on a relevant set of observable characteristics. Following Wooldridge (2010), under this assumption we can weight observations in (5.5) using the inverse of the predicted probability that they remain in the sample. Redefine $\mathbf{Z}_i$ as the set of observable characteristics in the outcome equation, measured at baseline, as well a number of other baseline characteris-

Table 5.13: Birth order and the probability a child dies during course of CLHNS

	OLS		Probit	
	(1)	(2)	(3)	(4)
main				
Relative birth order	0.0216	0.0188	0.161	0.152
	(0.0164)	(0.0163)	(0.123)	(0.132)
Total number of siblings	0.00972***	0.00779***	0.0649***	0.0500***
	(0.00221)	(0.00246)	(0.0133)	(0.0165)
Male	0.0202*	0.0216*	0.134	0.155*
	(0.0118)	(0.0118)	(0.0854)	(0.0859)
Baseline controls	No	Yes	No	Yes
R-Square	0.0125	0.0203		
Obs	1964	1964	1964	1964

Dependent variable is a dummy variable = 1 if the index child died before the final round. Sample comprises index children reported by mother's interviewed in 2005 round. Baseline controls include assets, household income, family size and mother's height measured at birth. Robust standard errors in parentheses, $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

tics otherwise not included in (5.5).[20] I then estimate (5.6) using a probit regression and weight the observations in (5.5) using the inverse of the predicted values (this is known as inverse probability weighting). While this method is likely valid for children who drop out early, the assumption of ignorability of selection after conditioning on $\mathbf{Z}_i$ is less likely to hold for children who drop out of the sample later on.

Finally, for the main results in this chapter, I have restricted the sample to children with data observed in the final round, but do not exclude children who are dropped in any given round, but return by 2005. I do this primarily for the extra gains in precision, but it is possible that non-random selection in and out of the sample could be driving the trend in birth order effects. To check for this, I also restrict the sample in the basic OLS specification to children who are observed in each and every round across the entire panel.

Tables 5.14 and 5.15 show the estimation of the main specification for both height-for-age and grade attainment under (1) OLS, (2) inverse-probability weighting and (3)

[20]These include the mother's weight at birth, whether or not the child lives in an urban district, the child's birth length and BMI-for-age at birth.

Table 5.14: Linear birth order and height for age, robustness checks

	Rounds: 1983-1985, average age listed below					Rounds: 1991-2005				
	0	.5	1	1.5	2	8.5	11.5	15.6	18.7	21.4
(1) OLS - full controls	0.221***	-0.0287	-0.182**	-0.142*	-0.132*	-0.0489	-0.00618	-0.00912	-0.0232	-0.0305
	(0.0693)	(0.0687)	(0.0715)	(0.0725)	(0.0696)	(0.0557)	(0.0593)	(0.0485)	(0.0477)	(0.0483)
(2) IPW	0.228***	-0.0224	-0.183**	-0.139*	-0.127*	-0.0551	-0.0158	-0.0153	-0.0247	-0.0347
	(0.0717)	(0.0702)	(0.0724)	(0.0738)	(0.0707)	(0.0567)	(0.0606)	(0.0491)	(0.0479)	(0.0489)
(3) Balanced panel	0.214***	-0.0589	-0.163**	-0.132*	-0.143*	-0.0691	0.00668	-0.0291	-0.0566	-0.0503
	(0.0776)	(0.0756)	(0.0773)	(0.0783)	(0.0747)	(0.0656)	(0.0686)	(0.0553)	(0.0535)	(0.0552)
Observations (1)-(2)	1826	1740	1712	1704	1697	1814	1822	1791	1800	1825
Observations (4)	1427	1427	1427	1427	1427	1427	1427	1427	1427	1427

Notes: Coefficients shown are taken from an OLS regression of child's height-for-age z-score on relative birth order. Each column is a separate CLHNS round (with column titles showing the mean age for the sample in that round). Each row is a separate specification: **(1)** displays OLS results from row (2) of Table 5.4. **(2)** shows results using inverse-probability weighting. **(3)** shows OLS results when sample is restricted to obs that never drop out. Robust standard errors $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 5.15: Linear birth order and grade attainment, robustness checks

	Rounds: 1991-2005				
	8.5	11.5	15.6	18.7	21.4
(1) OLS - full controls	-0.106***	-0.195***	-0.129	-0.0439	-0.0222
	(0.0323)	(0.0572)	(0.108)	(0.142)	(0.171)
(2) IPW	-0.104***	-0.195***	-0.126	-0.0500	-0.0426
	(0.0322)	(0.0573)	(0.108)	(0.142)	(0.172)
(3) Balanced panel	-0.117***	-0.154**	-0.0189	0.0939	0.172
	(0.0373)	(0.0648)	(0.124)	(0.164)	(0.193)
Observations (1)-(2)	1815	1823	1792	1801	1827
Observations (3)	1429	1429	1429	1429	1429

Notes: Coefficients shown are taken from an OLS regression of child's grade attainment on relative birth order. Each column is a separate CLHNS round (with column titles showing the mean age for the sample in that round). Each row is a separate specification: **(1)** displays OLS results from row (2) of Table 5.5. **(2)** shows results using inverse-probability weighting. **(3)** shows OLS results when sample is restricted to observations that never drop out. Robust standard errors. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

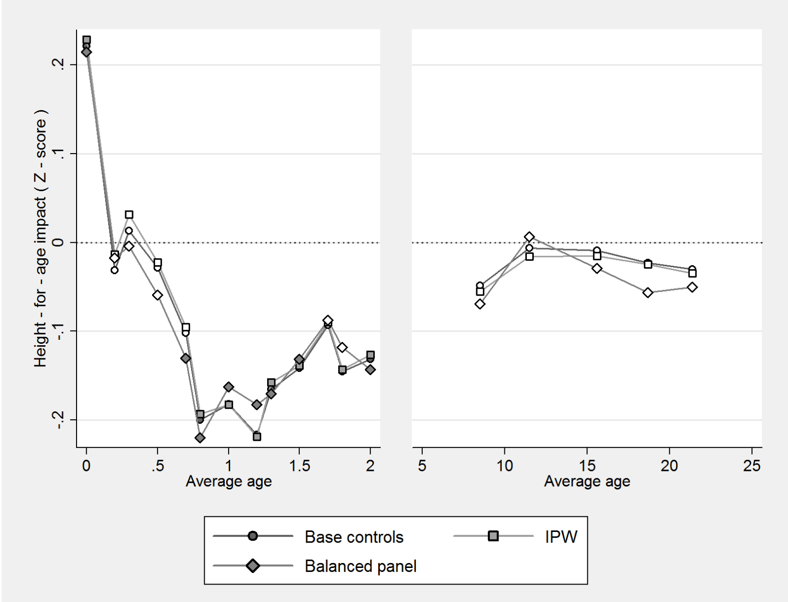
restricting the sample to a balanced panel. There are slight differences in the estimates of the birth order effect between the different methods, yet these differences are relatively small, with the overall trend remaining robust to these alternate specifications.[21] For comparison across all rounds 5.8 graphs the coefficients from the height-for-age specification, revealing a very similar trend across the three specifications. In the appendix, Tables 5.22 and 5.23 display the results using firstborn and lastborn dummies instead of linear birth order, again revealing that these alternative methods do not lead to substantially different estimates, suggesting that attrition is not a significant issue for estimates of birth order effects.

5.6.2 Sibling differences in completed education

In the final round of the CLHNS, mothers are asked about the educational attainment of all their surviving children. This allows me to use a standard household fixed-effects specification to remove time-invariant household unobservables α_j and see if the basic birth order results hold for grade attainment. Since the mother lists only her surviving

[21]The one exception are the last three rounds in Table 5.15, where the balanced panel deviates from the other two methods, although no specification produces statistically significant coefficients.

Figure 5.8: Impact of relative birth order on height-for-age, robustness



Note: Darkened estimates are those which are statistically significant at the 10% level

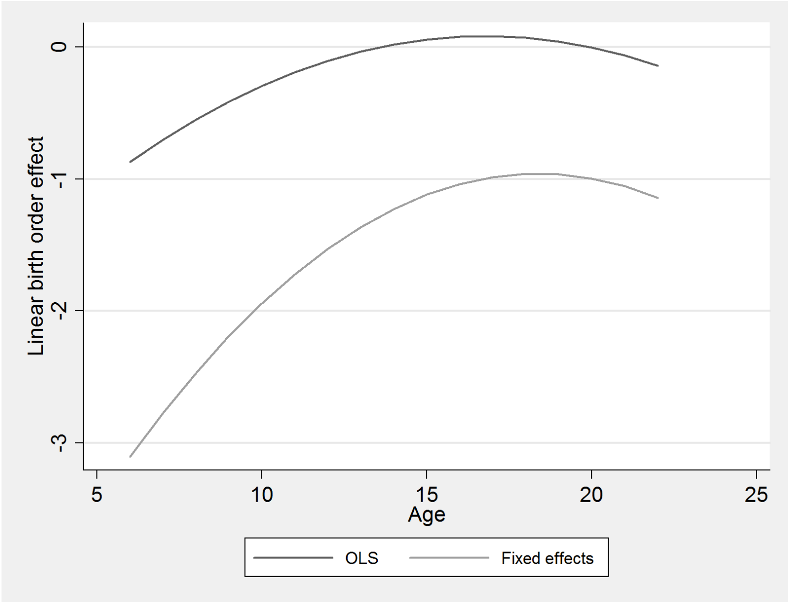
children, the resulting ranking is closer to an 'age ordering' than a birth order, which might still be a reasonable proxy for the birth order measures used in the previous section. Consider a within transformation of the original specification, equation (5.1), where we subtract from each observation the average value for all siblings from the household (at a given time t). Since we only observe the entire household in one period (2005), the t subscript can be dropped, as can any household characteristics:

$$
\begin{aligned}
y_{ij} - \overline{y}_j &= (bo_i - \overline{bo}_j)\theta + (\mathbf{X}_{it} - \overline{\mathbf{X}}_j)\beta_x + (\gamma_i - \overline{\gamma}_j) + (\varepsilon_{ij} - \overline{\varepsilon}_j) \\
y^*_{ij} &= bo^*_i\theta + \mathbf{X}^*_{it}\beta_x + \gamma^*_i + \varepsilon^*_{ij}
\end{aligned}
$$

Where $\overline{y}_{jt}$ is the mean value across other siblings in the household[22] j and y^*_{ij} is the demeaned value of y for sibling i from that household. As we only observe the household in one period, it is impossible to observe birth order effects changing over time as I have done in previous sections. I can, however, proxy for these long-term age trends by interacting

[22]The household fixed effect in this specification is defined over siblings sharing the same mother, so it is actually closer to a 'mother' fixed effect, as many of the siblings will now be living elsewhere.

Figure 5.9: Impact of relative birth order on grade attainment, fixed effects estimates



Note: Lines are calculated using point coefficients presented in Table 5.16 for ages 6-22.

the birth order measure with the age of the child.

Table 5.16 shows the results from four different specifications. The first column shows the results from a pooled OLS specification, in which all children between the ages of six (school starting age) and thirty are included without any household fixed effects.[23] The average effect of relative birth order is negative and insignificant. In column (2), I introduce interactions between the child's age and relative birth order. The coefficient on relative birth order is negative and significant, indicating that later-born children begin their school-aged years behind. However, the positive and significant coefficient on the interaction between birth order and age indicates that later-born children catch up as they get older. This catch up diminishes over time as children approach adulthood: Figure 5.9 graphs the estimated aggregate effect of relative birth order, utilising the estimated coefficients from Table 5.16.

Column (3) uses the same specification as (1) after introducing household fixed effects. In this specification, which is similar to those used in many cross-sectional studies of birth order effects, relative birth order has a strong, negative impact on educational attainment,

[23]To reduce the scope for cohort effects as opposed to age effects, I am only including children who should be experiencing changes in eduction as they age.

with firstborns estimated to have over a year more schooling than lastborns. Yet, when relative birth order is interacted with age, a similar trend to the OLS specification appears: while later-borns start off with a large deficit, they begin to recover by adulthood. Figure 5.9 also graphs the coefficient estimates for the fixed-effect specification. While it predicts larger birth order impacts overall, the same 'recovery' is seen here. These results are estimated over several different cohorts of children, which might explain why the effect sizes are much larger in comparison to those seen in previous sections. While these results are also dependent on the inclusion of a quadratic age term, they reinforce the result that grade inequality between children of different birth orders fades over time.

5.7 Final discussion and conclusion

In this chapter I set out to identify birth order effects in a single cohort of children from the Cebu province of Philippines. By estimating birth order effects for each round of the panel separately, I have shown that, while birth order effects appear in two salient child outcomes, height and educational attainment, they largely vanish by the time children reach adulthood. To date, although studies of birth order effects have shown positive, negative or null results during childhood or in adulthood, none have shown that these effects can appear temporarily. Given that many studies of birth order effects in poor contexts use data on children, it is possible that previous work is only capturing transitory effects, when true, final birth order effects are actually negligible. This might be due to parental compensation, either intentional or accidental, as later-born children see a brief, albeit weak improvement in their short-tern nutrition (as measured by BMI) and are subject to much later weaning periods. Later-born children also appear to start school later, but finish school at a later age while still completing an equal number of grades on average.

Given these results, future research in birth order effects should, whenever possible, either focus on adult-outcomes or at least estimate birth order effects with age interactions. The transitory nature of birth order effects should not be seen as an excuse to ignore them in future research or policy. When they appear, these effects are relative large and comparable to the gender effects seen in this cohort. Finding ways to reduce inequities, even temporary ones, will be welfare-enhancing.

There is also substantial evidence evidence that birth order effects are nonlinear, with firstborn enjoying a significant advantage over all other children in both height-for-age and educational attainment for most of their childhood. Several specifications also hint that lastborn children might be quicker to catch up than middleborns. There is also some evidence that lastborn children are quicker to catch up than middleborns, indicating that birth order effects might be non-monotonic. Both of these results suggest that current birth order models are incomplete, as a strictly firstborn and/or lastborn bias suggests that environmental effects are not a straightforward story of later-born children dealing with more siblings, and that the might be room for either direct parental preferences over certain children, or the incentive to have children of different birth order specialize. An ethnographic study of rural villages in the Philippines suggests that parents focus more attention and resources on both firstborns and lastborns (Mendez and Jocano 1974):

> "When it comes to inheritance, the panganay (oldest) and the bunso (youngest) are usually favored. The oldest deserves compensation for the help extended in looking after, and perhaps helping financially, too, the youngest siblings. The youngest child gets a bigger share, too, because she/he enjoyed the company of the parents for a shorter duration. The other children do not complain because this is the norm. As an informant put it, "I was a middle child and therefore got less than the oldest and the youngest in the family.""

I have also presented evidence that many of the popular theories of birth order effects do not appear to be operating in the Filipino context. For one, there is little support for the luxury axiom hypothesis suggested by studies such as Chesnokova and Vaithianathan (2008) and De Haan, Plug, and Rosero (2012), as early-born children do not have a significantly higher likelihood of entering the labour market at any stage of their life. Furthermore, the lack of persistence in any of the observed birth order effects suggest that parental investments not driving any observable inequities, or if they are then other, unobservable compensating factors are kicking in. While there is some limited evidence

Table 5.16: Relative birth order and grade attainment, household fixed effects

	OLS		Fixed effects	
	(1)	(2)	(3)	(4)
Relative birth order	-0.151	-2.232**	-1.108***	-5.694***
	(0.128)	(1.053)	(0.159)	(1.232)
Age (years)	1.596***	1.467***	1.561***	1.260***
	(0.0295)	(0.0917)	(0.0308)	(0.107)
Age^2	-0.0327***	-0.0292***	-0.0339***	-0.0266***
	(0.000843)	(0.00207)	(0.000854)	(0.00232)
$RBO \times Age$		0.276***		0.515***
		(0.0996)		(0.110)
$RBO \times Age^2$		-0.00824***		-0.0140***
		(0.00253)		(0.00263)
Male	-1.052***	-1.054***	-1.056***	-1.061***
	(0.0599)	(0.0599)	(0.0571)	(0.0571)
Sibship size	-0.136***	-0.132***		
	(0.0230)	(0.0235)		
Assets (std)	0.873***	0.876***		
	(0.0640)	(0.0640)		
Monthly income (std)	-0.0851	-0.0848		
	(0.0640)	(0.0642)		
Avg birth spacing	-0.00139	-0.00103		
	(0.00303)	(0.00304)		
M's education	0.192***	0.191***		
	(0.0140)	(0.0140)		
Mother height (cm)	0.0237***	0.0236***		
	(0.00883)	(0.00884)		
Interview date	-0.0000189	-0.0000621		
	(0.000947)	(0.000946)		
Land/house value	-0.119**	-0.119**		
	(0.0491)	(0.0489)		
Adult equiv units	0.0121	0.00987		
	(0.0231)	(0.0232)		
Constant	-11.59	-8.864	-6.528***	-3.492***
	(36.47)	(36.49)	(0.371)	(1.236)
Observations	8402	8402	8402	8402

Notes: Standard errors clustered at the household level $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

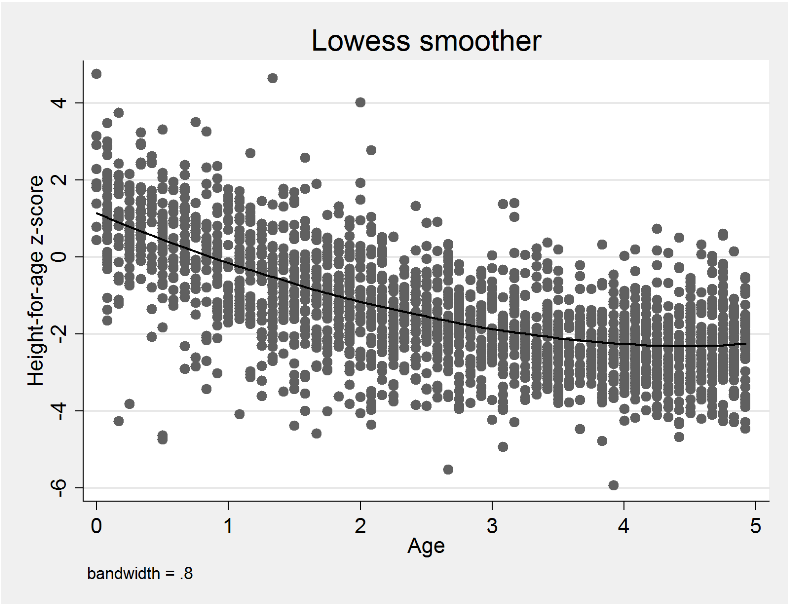
that liquidity constraints are driving the results in education, the result is not robust enough to make more general statements about the mechanisms behind birth order effects in this context. All of this this casts some doubt on the explanation that positive birth order effects in education are a common feature of developing countries (De Haan, Plug, and Rosero 2012). Given the heterogeneity in the estimated impact of birth order on education attainment across studies in poor societies, it is possible that birth order effects are highly context-sensitive.

5.A Chapter 5 Appendix

5.A.1 Extra tables and figures

5.A.2 Figures

Figure 5.10: Height-for-age z-score age trends, younger siblings (1991 round)



Note: Figure shows lowess smoothing plot of height-for-age z-scores for next-youngest siblings of index children.

Table 5.17: Height for age z-score results, full controls

	1983-1985					1991-2005				
	0	.5	1	1.5	2	8.5	11.5	15.6	18.7	21.4
Final relative bo	0.221***	-0.0299	-0.183**	-0.142*	-0.132*	-0.0533	-0.0108	-0.0113	-0.0253	-0.0320
	(0.0693)	(0.0688)	(0.0716)	(0.0725)	(0.0696)	(0.0557)	(0.0593)	(0.0485)	(0.0477)	(0.0484)
Family size (final)	-0.00684	-0.0369***	-0.0509***	-0.0547***	-0.0673***	-0.0288***	-0.0338***	-0.0247***	-0.0177**	-0.0232***
	(0.0104)	(0.0106)	(0.0110)	(0.0114)	(0.0108)	(0.00878)	(0.00999)	(0.00898)	(0.00804)	(0.00762)
Male	-0.180***	-0.264***	-0.234***	-0.192***	-0.108**	-0.111***	-0.0869**	-0.0638	-0.0311	-0.0230
	(0.0480)	(0.0483)	(0.0494)	(0.0494)	(0.0477)	(0.0386)	(0.0425)	(0.0956)	(0.0339)	(0.0338)
Age (months)	-0.505**	-0.0987	-0.112	-0.0825	0.798***	-0.0744***	-0.00423	-0.000346	0.00385	0.00619
	(0.229)	(0.187)	(0.163)	(0.163)	(0.146)	(0.00529)	(0.00629)	(0.00518)	(0.00508)	(0.00510)
Assets (std)	-0.00385	0.0661***	0.0595***	0.0681**	0.0415**	0.111***	0.199***	0.0913***	0.0807***	0.0860***
	(0.0194)	(0.0189)	(0.0188)	(0.0277)	(0.0196)	(0.0284)	(0.0281)	(0.0228)	(0.0238)	(0.0262)
Monthly income (std)	0.0315**	0.0234	0.0301	0.0335	0.0648***	0.0271	0.0627***	0.0562***	-0.0415***	-0.0300
	(0.0141)	(0.0190)	(0.0251)	(0.0354)	(0.0248)	(0.0266)	(0.0237)	(0.0210)	(0.00290)	(0.0252)
Avg birth spacing	0.00135	0.00257	0.00367**	0.00722***	0.00649***	0.00360***	0.00385***	0.00182	0.00152	0.00180
	(0.00137)	(0.00161)	(0.00159)	(0.00155)	(0.00156)	(0.00125)	(0.00147)	(0.00122)	(0.00116)	(0.00119)
M's education	0.0227***	0.0217***	0.0273***	0.0402***	0.0433***	0.0223***	0.0115	0.00666	0.00515	0.00103
	(0.00719)	(0.00727)	(0.00763)	(0.00778)	(0.00755)	(0.00650)	(0.00702)	(0.00571)	(0.00557)	(0.00553)
Mother height (cm)	0.0357***	0.0567***	0.0652***	0.0655***	0.0595***	0.0605***	0.0592***	0.0663***	0.0711***	0.0718***
	(0.00468)	(0.00510)	(0.00508)	(0.00539)	(0.00494)	(0.00396)	(0.00409)	(0.00345)	(0.00343)	(0.00347)
Interview date	-0.316***	-0.427***	-0.423***	-0.352***	-0.120	-0.0810	-0.112	-0.000912	-0.0163	-0.0223
	(0.0856)	(0.0856)	(0.0902)	(0.0894)	(0.0876)	(0.0897)	(0.0993)	(0.108)	(0.114)	(0.148)
Land/house value	0.0161	0.00974	0.0437	0.00986	0.00602	0.0678**	0.00721	0.00688	-0.00249	0.0116
	(0.0185)	(0.0284)	(0.0378)	(0.0374)	(0.0303)	(0.0305)	(0.0217)	(0.0181)	(0.0216)	(0.0202)
Adult equiv units	-0.0138	-0.0433***	-0.0235**	0.00512	0.00226	-0.0471***	-0.0613***	-0.0255**	-0.0176*	-0.00551
	(0.0106)	(0.0112)	(0.0120)	(0.0124)	(0.0124)	(0.0134)	(0.0136)	(0.0129)	(0.00948)	(0.00842)
Constant	20.44***	27.27***	25.99***	19.33**	-20.26**	4.109	0.784	-11.48	-11.65	-11.80
	(7.262)	(7.463)	(7.808)	(8.209)	(8.269)	(8.708)	(8.956)	(10.14)	(11.10)	(15.13)
R-Square	0.0626	0.148	0.183	0.201	0.223	0.321	0.244	0.251	0.245	0.245
Obs	1826	1738	1710	1703	1697	1812	1820	1789	1798	1823

This sidewaystable is a replication of the second row of Table 5.4 in Section 5.4.1 with all controls shown. Robust standard errors $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 5.18: Grade and birth order results, full controls

	Rounds: 1991-2005				
	8.5	11.5	15.6	18.7	21.4
Final relative bo	-0.111***	-0.180***	-0.126	-0.00397	0.0170
	(0.0329)	(0.0576)	(0.108)	(0.143)	(0.170)
Family size (final)	-0.0266***	-0.0640***	-0.0984***	-0.164***	-0.158***
	(0.00554)	(0.0131)	(0.0247)	(0.0281)	(0.0307)
Male	-0.100***	-0.287***	0.0248	-1.290***	-1.439***
	(0.0229)	(0.0434)	(0.235)	(0.103)	(0.121)
Age (months)	-0.00230	0.0804***	0.0617***	0.00457	-0.00671
	(0.00329)	(0.00658)	(0.0128)	(0.0162)	(0.0187)
Assets (std)	0.0851***	0.149***	0.591***	0.694***	1.102***
	(0.0154)	(0.0274)	(0.0565)	(0.0671)	(0.0857)
Monthly income (std)	0.00664	-0.00572	-0.0928*	0.0247**	-0.0630
	(0.00945)	(0.0193)	(0.0475)	(0.0105)	(0.0587)
Avg birth spacing	-0.000774	0.00132	0.00352	0.00502	0.00393
	(0.000700)	(0.00109)	(0.00227)	(0.00314)	(0.00361)
M's education	0.0215***	0.0530***	0.108***	0.130***	0.192***
	(0.00379)	(0.00685)	(0.0127)	(0.0164)	(0.0191)
Mother height (cm)	0.00167	0.00769*	0.00199	0.0139	0.0194
	(0.00224)	(0.00445)	(0.00846)	(0.0109)	(0.0129)
Interview date	0.122**	-1.113***	-0.898***	-0.298	0.344
	(0.0520)	(0.0992)	(0.266)	(0.371)	(0.511)
Land/house value	-0.0239**	-0.0306	-0.105***	-0.0696	-0.243***
	(0.0111)	(0.0225)	(0.0346)	(0.0469)	(0.0722)
Adult equiv units	-0.00121	0.0105	-0.00575	0.0821**	0.0127
	(0.00820)	(0.0167)	(0.0333)	(0.0331)	(0.0343)
Constant	-10.44**	98.10***	85.10***	37.19	-27.00
	(5.088)	(8.982)	(25.03)	(36.19)	(52.18)
R-Square	0.133	0.251	0.241	0.281	0.333
Obs	1755	1755	1755	1755	1755

This table is a replication of the second row of Table 5.5 in Section 5.4.1 with all controls shown. Robust standard errors $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

Table 5.19: Nonlinear effects of birth order on height-for-age z-score across childhood

	Coefficient	0	.5	1	1.5	2	8.5	11.5	15.6	18.7	21.4
(1) Basic controls	*RBO*	0.787***	0.143	-0.331	-0.515*	-0.768***	-0.806***	-0.938***	-0.336*	-0.182	-0.219
		(0.257)	(0.274)	(0.282)	(0.288)	(0.274)	(0.225)	(0.249)	(0.198)	(0.195)	(0.195)
	RBO^2	-0.608**	-0.248	0.0647	0.317	0.567**	0.713***	0.939***	0.391**	0.214	0.239
		(0.245)	(0.260)	(0.270)	(0.272)	(0.259)	(0.216)	(0.235)	(0.191)	(0.188)	(0.188)
(2) Full controls	*RBO*	0.862***	0.168	-0.234	-0.325	-0.606**	-0.624***	-0.773***	-0.238	-0.0623	-0.147
		(0.255)	(0.264)	(0.264)	(0.267)	(0.256)	(0.206)	(0.227)	(0.178)	(0.173)	(0.172)
	RBO^2	-0.652***	-0.207	0.0382	0.168	0.448*	0.569***	0.768***	0.242	0.0563	0.123
		(0.248)	(0.253)	(0.255)	(0.255)	(0.243)	(0.195)	(0.215)	(0.172)	(0.168)	(0.166)
(3) Full controls + birth length	*RBO*	-0.00759	-0.258	-0.595**	-0.674***	-0.927***	-0.847***	-0.975***	-0.422**	-0.268	-0.345**
		(0.0119)	(0.235)	(0.245)	(0.249)	(0.241)	(0.201)	(0.222)	(0.172)	(0.166)	(0.164)
	RBO^2	0.00555	0.111	0.310	0.434*	0.703***	0.740***	0.923***	0.382**	0.213	0.268*
		(0.00950)	(0.225)	(0.236)	(0.237)	(0.229)	(0.189)	(0.208)	(0.165)	(0.161)	(0.159)
Observations		1698	1617	1592	1587	1586	1685	1693	1664	1671	1695

Notes: Coefficients are taken from an OLS regression of child's height-for-age on final relative birth order (RBO) and RBO^2, first with a set of basic controls [age, sibling size and gender], then including a full set of controls [mother's education and height, household size, monthly income, land value, asset index and date of interview. Each column shows the results for a given round. Robust standard errors $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 5.20: The effect of relative birth order on BMI-for-age z-score

	Rounds: 1983-1985, average age listed below					Rounds: 1991-2005				
	0	.5	1	1.5	2	8.5	11.5	15.6	18.7	21.4
(1) Basic controls	0.329***	-0.175**	-0.00650	0.0197	-0.00837	0.0985	0.114	-0.00292	-0.00939	-0.0507
	(0.0647)	(0.0713)	(0.0683)	(0.0631)	(0.0647)	(0.0612)	(0.0766)	(0.0717)	(0.0682)	(0.0716)
(2) Full controls	0.344***	-0.149**	0.0511	0.0718	0.0512	0.131**	0.160**	-0.0217	-0.0151	-0.103
	(0.0658)	(0.0727)	(0.0691)	(0.0647)	(0.0684)	(0.0624)	(0.0773)	(0.0738)	(0.0713)	(0.0732)
(3) Full controls + birth length	0.309***	-0.169**	0.0320	0.0526	0.0301	0.113*	0.141*	-0.0398	-0.0229	-0.113
	(0.0642)	(0.0726)	(0.0692)	(0.0649)	(0.0685)	(0.0627)	(0.0772)	(0.0738)	(0.0711)	(0.0731)
Observations	1825	1738	1710	1707	1700	1812	1820	1789	1798	1821

Notes: Coefficients shown are taken from an OLS regression of child's BMI-for-age z-score on relative birth order. Each column is a separate CHNLS round (with column titles showing the mean age for the sample in that round). Each row is a separate specification: **Basic controls** = [Age, sibling size and gender]. **Full controls** = [Basic controls + mother's education and height, household size, monthly income, land value, asset index and date of interview. Robust standard errors $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 5.21: Birth order effects on probability of breastfeeding (2-24 months)

		2	6	10	12	14	18	22	24
(1) Full Controls	RBO	-0.0329	-0.0210	0.0284	0.105***	0.127***	0.224***	0.194***	0.188***
		(0.0246)	(0.0301)	(0.0332)	(0.0348)	(0.0365)	(0.0342)	(0.0291)	(0.0274)
(2) Full controls	Firstborn	-0.0269	-0.0241	-0.0712**	-0.143***	-0.137***	-0.153***	-0.0674***	-0.0458**
		(0.0240)	(0.0297)	(0.0330)	(0.0349)	(0.0353)	(0.0306)	(0.0243)	(0.0206)
	Lastborn	-0.0427*	-0.0209	-0.00563	0.0116	0.0401	0.101***	0.121***	0.145***
		(0.0224)	(0.0265)	(0.0288)	(0.0303)	(0.0329)	(0.0330)	(0.0301)	(0.0284)
Observations		1489	1489	1489	1489	1489	1489	1489	1489

Notes: Coefficients are taken from an linear probability model, regressing a child's probability of being breastfed in the past day on relative birth order (row 1) and firstborn/lastborn dummies (row 2). Both specifications use the full set of controls. Each column indicates a different round. Sample is restricted to index children observed in each of the first 12 rounds. Robust standard errors reported $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

Table 5.22: Firstborn/lastborn height-for-age effects, robustness checks

	Dummy	0	.5	1	1.5	2	8.5	11.5	15.6	18.7	21.4
(1) OLS - full controls	Firstborn	-0.207***	-0.0580	0.166***	0.175***	0.235***	0.187***	0.202***	0.0766*	0.0549	0.0587
		(0.0580)	(0.0622)	(0.0611)	(0.0631)	(0.0592)	(0.0477)	(0.0547)	(0.0431)	(0.0428)	(0.0445)
	Lastborn	-0.0120	-0.0602	-0.0976*	-0.0692	0.0320	0.102**	0.158***	0.0735*	0.0383	0.00623
		(0.0497)	(0.0531)	(0.0550)	(0.0558)	(0.0546)	(0.0470)	(0.0520)	(0.0433)	(0.0439)	(0.0458)
(2) IPW	Firstborn	-0.259***	-0.0642	0.146**	0.143*	0.190***	0.161***	0.208***	0.0576	0.0275	0.0465
		(0.0724)	(0.0742)	(0.0726)	(0.0752)	(0.0719)	(0.0562)	(0.0636)	(0.0484)	(0.0468)	(0.0468)
	Lastborn	-0.00982	-0.0102	-0.0781	-0.0472	0.0492	0.0717	0.122**	0.0218	-0.00119	-0.00453
		(0.0710)	(0.0699)	(0.0747)	(0.0739)	(0.0692)	(0.0565)	(0.0605)	(0.0497)	(0.0489)	(0.0492)
(3) Balanced panel	Firstborn	-0.251***	0.0206	0.211***	0.175**	0.239***	0.197***	0.212***	0.0839	0.0596	0.0612
		(0.0807)	(0.0792)	(0.0788)	(0.0795)	(0.0779)	(0.0648)	(0.0732)	(0.0551)	(0.0528)	(0.0535)
	Lastborn	-0.0290	0.00642	-0.0373	-0.0228	0.0910	0.0789	0.122*	0.0242	-0.0183	-0.0110
		(0.0762)	(0.0745)	(0.0757)	(0.0766)	(0.0724)	(0.0617)	(0.0656)	(0.0534)	(0.0522)	(0.0532)
Observations (1)-(2)	Firstborn	1698	1617	1592	1587	1586	1685	1693	1664	1671	1695
Observations (3)	Lastborn	1331	1331	1331	1331	1331	1331	1331	1331	1331	1331

Notes: Coefficients are taken from an OLS regression of child's height-for-age on dummies = 1 if the child is a firstborn or lastborn. Each column is a separate CLHNS round (with column titles showing the mean age for the sample in that round). Each row is a separate specification: **(1)** displays OLS results from row (2) of Table 5.6. **(2)** shows results using inverse-probability weighting. **(3)** shows OLS results when sample is restricted to observations that never drop out. Robust standard errors $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 5.23: Firstborn/lastborn effects on grade attainment, robustness checks

	Dummy	8.5	11.5	15.6	18.7	21.4
(1) OLS - full controls	Firstborn	0.0508	0.168***	0.257**	0.226*	0.163
		(0.0311)	(0.0494)	(0.100)	(0.127)	(0.155)
	Lastborn	-0.0620**	-0.0344	0.0616	0.0857	0.0472
		(0.0287)	(0.0558)	(0.107)	(0.141)	(0.162)
(2) IPW	Firstborn	0.0623*	0.188***	0.281***	0.213	0.141
		(0.0330)	(0.0512)	(0.106)	(0.133)	(0.165)
	Lastborn	-0.0498*	-0.00776	0.134	0.125	0.0685
		(0.0298)	(0.0598)	(0.112)	(0.150)	(0.172)
(3) Balanced panel	Firstborn	0.0883**	0.193***	0.282**	0.261*	0.162
		(0.0374)	(0.0568)	(0.123)	(0.157)	(0.184)
	Lastborn	-0.0631*	-0.0279	0.127	0.171	0.161
		(0.0336)	(0.0648)	(0.123)	(0.161)	(0.188)
Observations (1)-(2)	Firstborn	1686	1694	1665	1672	1697
Observations (3)	Lastborn	1331	1331	1331	1331	1331

Notes: Coefficients are taken from an OLS regression of child's grade attainment on dummies = 1 if the child is a firstborn or lastborn. Each column is a separate CLHNS round (with column titles showing the mean age for the sample in that round). Each row is a separate specification: **(1)** displays OLS results from row (2) of Table 5.7 shows results using inverse-probability weighting. **(4)** shows OLS results when sample is restricted to observations that never drop out. Robust standard errors $^{*}p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$

Chapter 6

Conclusion

In this thesis, I have investigated the determinants and potential impacts of the adoption of formal property rights in urban Tanzania, as well as shed some light on the persistence of birth order effects in the Philippines. In this chapter, I will revisit the main findings of eachs chapter, highlighting the key lessons and implications for further research.

In Chapter 2, I uncovered the existence of high levels of clustering in the adoption of formal land titles in the slums of Dar es Salaam. Using randomly-determined differences in the propensity to adopt driven by a field experiment, I identified large and significant endogenous peer effects in land title adoption. In my preferred specification, for every neighbour of the five closest who adopted a land title, the predicted probability that a household did the same increased by 15 percentage points, equivalent to a 50,0000 Tsh subsidy on the price of a land title. I then show that these peer effects are stronger for households with are physically proximate, and for households with a pre-existing fear of expropriation, suggesting that complementarities in the reduction of expropriation risk could be one explanation for the results. Overall, these results indicate that social multiplier effects will increase the effectiveness of interventions aimed at enticing informal settlers to adopt property titles, making large-scale interventions substantially more attractive.

In Chapter 3, I revealed that landowners in the urban slums are less likely to buy a limited land title, the residential license, if their parcel is surrounded by households of the same ethnic background. While this simple correlation is likely to be subject to a number of biases and endogenous sorting, I show that the effect survives the inclusion of a large set of household and parcel characteristics, geographic characteristics, ethnic fixed effects, spatial fixed effects and a control for the ethnic similarity of the household's neighbours when it arrived in the slum. Again, these effects seem to be channeled primarily through fear of expropriation, possibly from other landowners in the slum. Given that the unplanned settlements in Dar es Salaam are markedly unsegregated, this makes these areas substantially more appropriate for private, individualised tenure as opposed to customary methods of recording property rights. Yet, it appears that many households still cling to informal tenure, especially those surrounded by coethnics. While coethnicity may not be the driving force behind the decline in the residential license programme's popularity, these results do highlight the need of governments to consider the *de facto* tenure security of urban land owners when designing and marketing new tenure products.

Chapter 4 briefly steps inside the black box of household decision-making to consider the potential impacts of land titling on gender outcomes. Using a field experiment in Dar es Salaam which randomly allocated vouchers discounting the price of a land title, some of them conditional on including a woman as owner, this chapter shows that it is relatively easy to entice households to co-title with no adverse effects on demand. This suggests that the price of "empowerment" is close to zero: only a small nudge is required to get women included as *de jure* owners of urban land. Yet, the reasons the intervention was so successful are precisely the reasons we should be cautious about expecting substantial bargaining power impacts. Households which we might expect to have very low levels of female bargaining power, and thus where men might stand to 'lose' if the property is co-titled, are most responsive to the nudge of conditionality. While this improves the targeting of the intervention, it still indicates that more work must be done to verify that the intervention will be effective.

Taken together, Chapters 2-4 shed some light on the three observations I highlighted in Chapter 1. Through these results, we have learned that formal property rights regimes become more popular as more people opt in, but that the success of a given regime will depend highly on the types of benefits it offers, above and beyond what an informal or customary system provides. Furthermore, while it is relatively easy to get households to tweak the distribution of intrahousehold ownership once they have opted to join a formal system, this does not guarantee that these households plan to fully embrace all facets of the new regime. While households may be happy to protect themselves from local or government expropriation through purchasing a residential license or a CRO, this does not guarantee they value or will honour the other benefits or requirements that these titles bring. When introduced exogenously by governments, new tenure regimes are, by default, untested and uncertain enterprises. The literature on formal property rights would benefit from more research into how to correctly design and introduce new regimes, to complement the rather passive approach of impact evaluation.

Finally, Chapter 5 reveals birth order effects to be slightly more complex than has been currently presented in the literature. Differences between early and late-born children do not appear to be constant over time, and might also vary quite strongly over the birth order distribution. We seem to be quite far from an all-encompassing theory of birth order, as these effects appear to be highly context-dependent. Within the context of the Philippines, ending up as a later-born child does not appear to result in the same permanent deprivations which are shown or purported in other studies. However, these differences are sizable, indicating that future work to further disentangle the channels of these inequities will still have important implications for improving child welfare.

Bibliography

Afridi, F. (2005). Intra-household bargaining, birth order and the gender gap in schooling in india. Working paper, Univeristy of Michigan, Ann Arbor.

Alesina, A. and E. L. Ferrara (2000). Participation in heterogenous communities. *Quarterly Journal of Economics 115*(3), 847–904.

Alesina, A. and E. La Ferrara (2002). Who trusts others? *Journal of Public Economics 85*(2), 207–234.

Alesina, A. and E. Zhuravskaya (2011). Segregation and the quality of government in a cross section of countries. *The American Economic Review 101*(5), 1872–1911.

Algan, Y., C. Hémet, and D. Laitin (2011). Diversity and public goods: A natural experiment with exogenous residential allocation. IZA Discussion Papers 6053, Institute for the Study of Labor (IZA).

Allendorf, K. (2007). Do women's land rights promote empowerment and child health in Nepal? *World Development 35*(11), 1975–1988.

Angelucci, M. and G. De Giorgi (2009). Indirect effects of an aid program: how do cash transfers affect ineligibles' consumption? *The American Economic Review 99*(1), 486–508.

Ayalew Ali, D. and M. Goldstein (2011). Environmental and gender impacts of land tenure regularization in Africa. Working Paper 2011/74, World Institute for Development Economic Research (UNU-WIDER).

Bandiera, O. and I. Rasul (2006). Social networks and technology adoption in northern Mozambique. *The Economic Journal 116*(514), 869–902.

Basu, K. and P. Van (1998). The economics of child labor. *American Economic Review 88*(3), 412–427.

Behrman, J. (1988). Nutrition, health, birth order and seasonality: intrahousehold allocation among children in rural India. *Journal of Development Economics 28*(1), 43–62.

Behrman, J. R., R. A. Pollak, and P. Taubman (1982). Parental preferences and provision for progeny. *Journal of Political Economy 90*(1), 52–73.

Behrman, J. R. and P. Taubman (1986). Birth order, schooling, and earnings. *Journal of Labor Economics 4*(3), S121–S145.

Besley, T. (1995). Property rights and investment incentives: theory and evidence from Ghana. *Journal of Political Economy 103*(5), 903–937.

Besley, T. and M. Ghatak (2010). Property rights and economic development. In D. Rodrick and M. Rosenzweig (Eds.), *Handbook of Development Economics*. Amsterdam: North-Holland.

Biggs, T., M. Raturi, and P. Srivastava (2002). Ethnic networks and access to credit: evidence from the manufacturing sector in Kenya. *Journal of Economic Behavior and Organisation 49*, 473–486.

Birdsall, N. (1991). Birth order effects and time allocation. *Research in Population Economics 7*, 191–213.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005). The more the merrier? The effect of family size and birth order on children's education. *Quarterly Journal of Economics 120*(2), 669–700.

Blackburn, S. (2007). Postpartum period and lactation physiology. In *Maternal, fetal, and neonatal physiology: a clinical perspective*. Elsevier Health Sciences.

Blattman, C. (2011). The trials of randomization *Chris Blattman* (blog), April 23, 2011. `http://chrisblattman.com/2011/04/23/the-trials-of-randomization/`.

Bobonis, G. and F. Finan (2009). Neighborhood peer effects in secondary school enrollment decisions. *The Review of Economics and Statistics 91*(4), 695–716.

Boersma, B. and J. M. Wit (1997). Catch-up growth. *Endocrine reviews 18*(5), 646–661.

Booth, A. and H. Kee (2009). Birth order matters: the effect of family size and birth order on educational attainment. *Journal of Population Economics 22*(2), 367–397.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of Econometrics 150*(1), 41–55.

Brennan, J. (2007). Between segregation and gentrification: Africans, Indians, and the struggle for housing in Dar es Salaam, 1920-1950. In J. Brennan, A. Burton, and Y. Lawi (Eds.), *Dar es Salaam: histories from an emerging African metropolis*, pp. 223–233. Mkuki Na Nyota Publishers.

Brock, W. and S. Durlauf (2001). Discrete choice with social interactions. *The Review of Economic Studies 68*(2), 235–260.

Bruhn, M. and D. McKenzie (2009). In pursuit of balance: randomization in practice in development field experiments. *American Economic Journal: Applied Economics 1*(4), 200–232.

Bulow, J. I., J. D. Geanakoplos, and P. D. Klemperer (1985). Multimarket oligopoly: Strategic substitutes and complements. *The Journal of Political Economy 93*(3), 488–511.

Bursztyn, L., F. Ederer, B. Ferman, and N. Yuchtman (2012, July). Understanding peer effects in financial decisions: Evidence from a field experiment. NBER Working Papers 18241, National Bureau of Economic Research, Inc.

Caeyers, B. (2013). The exclusion bias in social interaction models: cause, consequences and solution. Working paper.

Cameron, A., J. Gelbach, and D. Miller (2011). Robust inference with multiway clustering. *Journal of Business ands Economic Statistics 29*(2), 238–249.

Chesnokova, T. and R. Vaithianathan (2008). Lucky last? Intra-sibling allocation of child labor. *The BE Journal of Economic Analysis and Policy 8*(1).

Collin, M., S. Dercon, H. Nielson, J. Sandefur, and A. Zeitlin (2012). The practical and institutional hurdles to obtaining land titles in urban Tanzania. IGC report, International Growth Centre.

Conley, D. and R. Glauber (2006). Parental educational investment and children's academic risk: estimates of the impact of sibship size and birth order from exogenous variation in fertility. *Journal of Human Resources 41*(4), 722–737.

Conley, T. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics 92*(1), 1–45.

Conley, T. and C. Udry (2010). Learning about a new technology: pineapple in Ghana. *The American Economic Review 100*(1), 35–69.

Court, D. (1984). The education system as response to inequality. In J. D. Barkan (Ed.), *Politics and public policy in Kenya and Tanzania.* New York: Praeger.

Crookston, B. T., M. E. Penny, S. C. Alder, T. T. Dickerson, R. M. Merrill, J. B. Stanford, C. A. Porucznik, and K. A. Dearden (2010). Children who recover from early stunting and children who are not stunted demonstrate similar levels of cognition. *The Journal of nutrition 140*(11), 1996–2001.

Dahl, G., K. Løken, and M. Mogstad (2012). Peer effects in program participation. Working Papers in Economics 12/12, University of Bergen, Department of Economics.

Damm, A. P. (2009). Ethnic enclaves and immigrant labor market outcomes: quasi-experimental evidence. *Journal of Labor Economics 27*(2), 281–314.

De Giorgi, G., M. Pellizzari, and S. Redaelli (2010). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics 2*(2), 241–275.

De Haan, M. (2010). Birth order, family size and educational attainment. *Economics of Education Review 29*(4), 576–588.

De Haan, M., E. Plug, and J. Rosero (2012, July). Birth order and human capital development: evidence from Ecuador. IZA Discussion Papers 6706, Institute for the Study of Labor (IZA).

De Meza, D. and J. R. Gould (1992). The social efficiency of private decisions to enforce property rights. *Journal of Political Economy*, 561–580.

De Soto, H. (2000). *The mystery of capital: why capitalism succeeds in the West and fails everywhere else.* New York: Basic Books.

De Soto, H. and F. Cheneval (2006). *Realizing property rights.* Zurich: Ruffer and Rub Pub.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature 48*(2), 424–455.

Deere, C. and M. León (2001). Who owns the land? Gender and land-titling programmes in Latin America. *Journal of Agrarian Change 1*(3), 440–467.

Deininger, K., C. Augustinus, S. Enemark, and P. Munro-Faure (2010). *Innovations in land rights recognition, administration and governance.* Washington D.C.: World Bank.

Deininger, K., A. Goyal, and H. Nagarajan (2010). Inheritance law reform and women's access to capital: evidence from India's Hindu Succession Act. Policy Research Working Paper Series 5338, The World Bank.

Dercon, S. and D. A. Ali (2007). Land rights, power and trees in rural Ethiopia. CSAE Working Paper Series 2007-07, Centre for the Study of African Economies, University of Oxford.

Dercon, S. and P. Krishnan (1998). Changes in poverty in rural Ethiopia 1989-1995: measurement, robustness tests and decomposition. CSAE Working Paper Series 1998-07, Centre for the Study of African Economies, University of Oxford.

Dercon, S., A. Park, and A. Singh (2012, June). School meals as a safety net: An evaluation of the midday meal scheme in india. CEPR Discussion Papers 9031, C.E.P.R. Discussion Papers.

Desmet, K., I. Ortuño-Ortín, and R. Wacziarg (2012). The political economy of linguistic cleavages. *Journal of Development Economics 97*(2), 322–338.

Doss, C. (2005). The effects of intrahousehold property ownership on expenditure patterns in Ghana. *Journal of African Economies 15*(1), 149–180.

Dower, P. and E. Potamites (2012, September). Signaling credit-worthiness: Land titles, banking practices and formal credit in Indonesia. Working Papers w0186, Center for Economic and Financial Research (CEFIR).

Duflo, E. and E. Saez (2003). The role of information and social interactions in retirement plan decisions: evidence from a randomized experiment. *The Quarterly Journal of Economics 118*(3), 815–842.

Durand-Lasserve, A. (2003). Land for housing the poor in african cities: are neo-customary processes an effective alternative to formal systems? In *Urban Research Symposium*, pp. 15–17.

Edin, P., P. Fredriksson, and O. Åslund (2003). Ethnic enclaves and the economic success of immigrants: evidence from a natural experiment. *The Quarterly Journal of Economics 118*(1), 329–357.

Edmonds, E. (2006). Understanding sibling differences in child labor. *Journal of Population Economics 19*(4), 795–821.

Ejrnæs, M. and C. C. Pörtner (2004). Birth order and the intrahousehold allocation of time and education. *Review of Economics and Statistics 86*(4), 1008–1019.

Emerson, P. M. and A. P. Souza (2008). Birth order, child labor, and school attendance in Brazil. *World Development 36*(9), 1647–1664.

Fafchamps, M. (2000). Ethnicity and credit in African manufacturing. *Journal of Development economics 61*(1), 205–235.

Fafchamps, M. and F. Gubert (2007). The formation of risk sharing networks. *Journal of Development Economics 83*(2), 326–350.

Fearon, J. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth 8*(2), 195–222.

Field, E. (2003). Fertility responses to urban land titling programs: the roles of ownership security and the distribution of household assets. Working paper, Harvard

University.

Field, E. (2005). Property rights and investment in urban slums. *Journal of the European Economic Association 3*(2-3), 279–290.

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

Fisman, R. J. (2003). Ethnic ties and the provision of credit: relationship-level evidence from African firms. *Advances in Economic Analysis and Policy 3*(1), 1–18.

Fitzgerald, J., P. Gottschalk, and R. Moffitt (1998). An analysis of sample attrition in panel data: the Michigan Panel Study of Income Dynamics. *The Journal of Human Resources 33*(2), 251–299.

Flegal, K., R. Wei, and C. Ogden (2002). Weight-for-stature compared with body mass index-for-age growth charts for the United States from the Centers for Disease Control and Prevention. *The American Journal of Clinical Nutrition 75*(4), 761–766.

Fort, R., R. Ruben, and J. Escobal (2006). Spillover and externality effects of titling on investments: evidence from Peru. In *11th Annual Meeting of the Latin American and Caribbean Economic Association.*

Galiani, S. and E. Schargrodsky (2010). Property rights for the poor: effects of land titling. *Journal of Public Economics 94*(9), 700–729.

Garg, A. and J. Morduch (1998). Sibling rivalry and the gender gap: evidence from child health outcomes in Ghana. *Journal of Population Economics 11*, 471–493.

Gerdes, C. (2011). The impact of immigration on the size of government: empirical evidence from Danish municipalities. *The Scandinavian Journal of Economics 113*(1), 74–92.

Glaeser, E., J. Scheinkman, and B. Sacerdote (2003). The social multiplier. *Journal of the European Economic Association*, 345–353.

Glennerster, R., E. Miguel, and A. Rothenberg (2010). Collective action in diverse

Sierra Leone communities. NBER Working Papers 16196, National Bureau of Economic Research, Inc.

Godlonton, S. and R. Thornton (2012). Peer effects in learning HIV results. *Journal of Development Economics 97*(1), 118–129.

Goldstein, M. and C. Udry (2008). The profits of power: land rights and agricultural investment in Ghana. *Journal of Political Economy 116*(6), 981–1022.

Grimard, F. (1997). Household consumption smoothing through ethnic ties: evidence from Cote d'Ivoire. *Journal of Development Economics 53*(2), 391–422.

Guryan, J., K. Kroft, and M. Notowidigdo (2009). Peer effects in the workplace: evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics 1*(4), 34.

Guthrie, G. M., A. Tayag, and J. Jimenez (1977). The Philippine Non-Verbal Intelligence Test. *Journal of Social Psychology 102*, 3–11.

Habyarimana, J., M. Humphreys, D. N. Posner, and J. M. Weinstein (2007). Why does ethnic diversity undermine public goods provision? *American Political Science Review 101*(4), 709–725.

Haughton, D. and J. Haughton (1997). Explaining child nutrition in Vietnam. *Economic Development and Cultural Change 45*(3), 541–556.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153–161.

Heckman, J. (2012). Promoting social mobility. *Boston Review* (September/October).

Hirvonen, K. (2013, February). Measuring catch-up growth in malnourished populations. Working Paper Series 5913, Department of Economics, University of Sussex.

Hoff, K. and A. Sen (2006). The kin system as a poverty trap? In S. Bowles, S. N. Durlauf, and K. Hoff (Eds.), *Poverty traps*, Chapter 5, pp. 95–115. Princeton: Princeton University Press.

Hooper, M. and L. Ortolano (2012). Confronting urban displacement social movement participation and post-eviction resettlement success in Dar es Salaam, Tanzania.

Journal of Planning Education and Research 32(3), 278–288.

Horrace, W. and R. Oaxaca (2006). Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters 90*(3), 321–327.

Horton, S. (1988). Birth order and child nutritional status: evidence from the Philippines. *Economic Development and Cultural Change 36*(2), 341–354.

Hotz, V. and J. Pantano (2011). Strategic parenting, birth order and school performance. Working paper, Washington University.

Imbens, G. and J. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Jacoby, H. and B. Minten (2007). Is land titling in sub-Saharan Africa cost-effective? Evidence from Madagascar. *The World Bank Economic Review 21*(3), 461–485.

Jakiela, P. and O. Ozier (2012). Does Africa need a rotten kin theorem? Experimental evidence from village economies. Policy Research Working Paper Series 6085, The World Bank.

Jayachandran, S. and I. Kuziemko (2011). Why do mothers breastfeed girls less than boys? Evidence and implications for child health in India. *The Quarterly Journal of Economics 126*(3), 1485–1538.

Kantarevic, J. and S. Mechoulan (2006). Birth order, educational attainment, and earnings an investigation using the PSID. *Journal of Human Resources 41*(4), 755–777.

Kebede, B. (2005). Genetic endowments, parental and child health in rural Ethiopia. *Scottish Journal of Political Economy 52*(2), 194–221.

Kironde, J. (1995). Access to land by the urban poor in Tanzania: some findings from Dar es Salaam. *Environment and Urbanization 7*(1), 77–96.

Kironde, J. L. (2006). Issuing of residential licences to landowners in unplanned settlements in Dar es Salaam, Tanzania. Technical report, UN-Habitat, Shelter Branch, Land and Tenure Section.

Kironde, J. L. (2009). Improving land sector governance in Tanzania: implementation of the land governance assessment framework. Working paper, Ardhi University.

Kironde, L. (2011). Is the residential licensing project still on the national development agenda? *Daily News, Online Edition*, October 29. `http://www.dailynews.co.tz/columnist/?n=24984`.

Kombe, W. (2005). Land use dynamics in peri-urban areas and their implications on the urban growth and form: the case of Dar es Salaam, Tanzania. *Habitat International 29*(1), 113–135.

Kombe, W. (2010). Land conflicts in Dar es Salaam: who gains? who loses? Cities and Fragile States Working Paper 82, Crisis States Research Centre, London School of Economics and Polical Science.

Kombe, W. and V. Kreibich (2002). Informal land management in Tanzania and the misconception about its illegality. In V. Kreibich and W. Olima (Eds.), *Urban Land Management in Africa. Dortmund (SPRING Research Series 40)*, pp. 267–283.

Lalive, R. and M. Cattaneo (2009). Social interactions and schooling decisions. *The Review of Economics and Statistics 91*(3), 457–477.

Lassen, D. D. (2007). Ethnic divisions, trust, and the size of the informal sector. *Journal of Economic Behavior and Organisation 63*, 423–438.

Lewis, M. (2009). *Ethnologue: languages of the world*, Volume 9. SIL international Dallas, TX.

Libecap, G. and D. Lueck (2011). The demarcation of land and the role of coordinating property institutions. *The Journal of Political Economy 119*(3), 426–467.

Lilleør, H. B. and D. D. Lassen (2008). Informal institutions and intergenerational contracts: Evidence from schooling and remittances in rural Tanzania. CAM Working Papers 2008-03, University of Copenhagen. Department of Economics. Centre for Applied Microeconometrics.

Lindert, P. H. (1977). Sibling position and achievement. *Journal of Human Resources 12*(2), 198–219.

Liu, H., T. Mroz, and L. Adair (2009). Parental compensatory behaviors and early child health outcomes in cebu, philippines. *Journal of Development Economics 90*(2), 209–230.

Macours, K. (2007). Ethnic divisions, contract choice, and search costs in the Guatemalan land rental market. Working paper, Johns Hopkins University.

Magigi, W. and B. Majani (2006). Community involvement in land regularization for informal settlements in Tanzania: a strategy for enhancing security of tenure in residential neighborhoods. *Habitat international 30*(4), 1066–1081.

Magruder, J. (2012a). Can minimum wages cause a big push? Evidence from indonesia. *Journal of Development Economics 100*(1), 48–62.

Magruder, J. (2012b). High unemployment yet few small firms: the role of centralized bargaining in south africa. *American Economic Journal: Applied Economics 4*(3), 138–166.

Manski, C. (1993). Identification of endogenous social effects: the reflection problem. *The Review of Economic Studies 60*(3), 531–542.

Medina, B. T. G. (1991). *The Filipino family: a test with selected readings*. Quezon City: University of the Philippines Press.

Mendez, P. P. and F. L. Jocano (1974). *The Filipino family in its rural and urban orientation: two case studies*. Manila: Centro Escolar University Research and Development Center.

Midheme, E. P. O. (2007). *State vs. community-led land tenure regularization in Tanzania: the case of Dar es Salaam City*. Ph. D. thesis.

Miguel, E. (2004). Tribe or nation? Nation building and public goods in Kenya versus Tanzania. *World Politics 56*(3), 327–62.

Miguel, E. and M. K. Gugerty (2005). Ethnic diversity, social sanctions, and public goods in Kenya. *Journal of Public Economics 89*(11), 2325–2368.

Miller, J. et al. (1994). Birth order, interpregnancy interval and birth outcomes among Filipino infants. *Journal of Biosocial Science 26*, 243–243.

Moffitt, R. et al. (2001). Policy interventions, low-level equilibria, and social interactions. *Social dynamics*, 45–82.

Moulton, B. (1986). Random group effects and the precision of regression estimates. *Journal of econometrics 32*(3), 385–397.

Müller, H. P., C. K. Marti, E. S. Schiedt, and B. Arpagaus (2000). *Atlas vorkolonialer Gesellschaften*. Berlin: Reimer.

Ndezi, T. (2009). The limit of community initiatives in addressing resettlement in Kurasini ward, Tanzania. *Environment and Urbanization 21*(77), 77–88.

Ngatia, M. (2011). Social interactions and individual reproductive decisions. Working paper, Yale University.

Oster, E. and R. Thornton (2009). Determinants of technology adoption: private value and peer effects in menstrual cup take-up. NBER Working Papers 14828, National Bureau of Economic Research.

Ota, M. and P. G. Moffatt (2007). The within-household schooling decision: a study of children in rural Andhra Pradesh. *Journal of Population Economics 20*(1), 223–239.

Outes, I. and C. Porter (2012). Catching up from early nutritional deficits? evidence from rural ethiopia. *Economics & Human Biology*.

Outes-Leon, I. and S. Dercon (2008). Survey attrition and attrition bias in young lives. Young Lives Technical Note No 5, Young Lives.

Payne, G., A. Durand-Lasserve, and C. Rakodi (2007). Social and economic impacts of land titling programmes in urban and peri-urban areas: a review of the literature. In *World Bank Urban Research Symposium*, Volume 14, pp. 16.

Peterman, A. (2011). Women's property rights and gendered policies: implications for women's long-term welfare in rural Tanzania. *Journal of Development Studies 47*(1), 1–30.

Platteau, J. P. (2000). *Institutions, social norms, and economic development*. Amsterdam: Harwood Academic Publishers.

Polome, E. C. (1980). *Tanzania: a socio-linguistic perspective*. Oxford: Oxford University Press.

Price, J. (2008). Parent-child quality time: does birth order matter? *Journal of Human Resources 43*(1), 240–265.

Qian, N. (2008). Missing women and the price of tea in china: the effect of sex-specific earnings on sex imbalance. *The Quarterly Journal of Economics 123*(3), 1251–1285.

Quisumbing, A. R. (1995). The extended family and intrahousehold allocation: inheritance and investments in children in the rural Philippines. FCND discussion papers 3, International Food Policy Research Institute (IFPRI).

Rammohan, A. and D. Dancer (2008). Gender differences in intrahousehold schooling outcomes: the role of sibling characteristics and birth-order effects. *Education Economics 16*(2), 111–126.

Ravallion, M. (2008). Evaluation in the practice of development. Policy Research Working Paper Series 4547, The World Bank.

Rosenzweig, M. R. and T. P. Schultz (1983). Market opportunities, genetic endowments, and intrafamily resource distribution: child survival in rural India. *American Economic Review 72*(4), 803–815.

Rous, J. (2001). Is breast-feeding a substitute for contraception in family planning? *Demography 38*(4), 497–512.

Roy, S. (2008). Female empowerment through inheritance rights: evidence from India. Working paper, London School of Economics.

Sacerdote, B. (2001). Peer effects with random assignment: results for Dartmouth roommates. *The Quarterly Journal of Economics 116*(2), 681–704.

Schelling, T. C. (1978). *Micromotives and macrobehavior*. WW Norton.

Senauer, B., M. Garcia, and E. Jancito (1988). Determinants of the intrahousehold allocation of food in the rural Philippines. *American Journal of Agricultural Economics 70*(1), 170–180.

Slater, E. (1962). Birth order and maternal age of homosexuals. *Lancet 1*(7220), 69.

Sulloway, F. J., R. Hertwig, and J. N. Davis (2002). Parental investment: how an equity motive can produce inequality. *Psychological Bulletin 128*(5), 728–745.

Sundet, G. (2005). The 1999 Land Act and Village Land Act: a technical analysis of the practical implications of the Acts. Technical report.

Telalagic, S. (2012). Domestic production as a source of marital power: theory and evidence from Malawi. Cambridge Working Papers in Economics 1243, Faculty of Economics, University of Cambridge.

Tenikue, M. and B. Verheyden (2010). Birth order and schooling: Theory and evidence from twelve sub-saharan countries. *Journal of African Economies 19*(4), 459–495.

Udry, C. (2012). Land tenure. In E. Aryeetey, S. Devarajan, R. Kanbur, and L. Kasekende (Eds.), *The Oxford Companion to the Economics of Africa*, pp. 411–415. Oxford: Oxford University Press.

UN-HABITAT (2010). *State of the world's cities 2010/2011: bridging the urban divide.* Earthscan/James & James.

Van Tassel, E. (2004). Credit access and transferable land rights. *Oxford economic papers 56*(1), 151–166.

Wang, S. (2011). Property rights and intra-household bargaining. Working paper, University of Pennsylvania.

Woodruff, C. (2001). Review of De Soto's The Mystery of Capital. *Journal of Economic Literature 39*(4), 1215–1223.

Wooldridge, J. (2002). Econometric analysis of cross section and panel data.

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*, Volume 1. The MIT Press.

Zajonc, R. B. and G. B. Markus (1975). Birth order and intellectual development. *Psychological Review 93*(2), 74–88.

Zeitlin, A. (2012). Identification and estimation of peer effects on endogenous affiliation networks: an application to Ghanaian agriculture. Working paper.

Zerfu, D., P. Zikhali, and I. Kabenga (2008). Does ethnicity matter for trust? Evidence from Africa. *Journal of African Economies 18*(1), 153–175.