Essays in the economics of networks

Arun Naresh Advani

Thesis submitted to the Department of Economics
in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

University College London

London
March 2017

Declaration

I, Arun Naresh Advani, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. Chapters 1, 3, and 6 are sole-authored work. Chapter 2 is joint work with Bansi Malde. Chapter 4 is joint work with Bryony Reich. Chapter 5 is joint work with Jonathan Shaw and William Elming.

Signed .... (Arun Naresh Advani)

Date:

Abstract

This thesis comprises four papers on the economics of networks.

The first provides a survey of methods for econometric work with networks, summarising what is known in the economics literature, and incorporating methods from other fields including mathematics, computer science, and sociology.

The second explains why informal insurance networks are able to provide consumption smoothing, but not credit for investment. It develops a dynamic contracting model to explain the difference in these motivations, generates from this a number of testable predictions, and then provides evidence in support of the model using data from Bangladesh.

The third examines the trade-off between economic and cultural incentives for migrants. It develops a model of network formation combined with a game-on-a-network, to capture this trade-off. The model's predictions are then tested using data from US censuses in the early 1900s.

The fourth studies the impact of tax audits by HMRC, the UK tax authority, on the future tax reporting behaviour of self-assessment taxpayers. It shows that income sources which are not also reported by a third party are more likely to be initially underreported, and that being audited increases future reports for a number of years. Future work in this area will study the spillover effects of the audits onto taxpayers connected to those who are audited, where networks of connections can be measured by sharing a place of residence, place of work, being partners in a common business, and sharing a common accountant.

Acknowledgments

I am very grateful to my advisors Richard Blundell and Imran Rasul for their unrelenting support and encouragement throughout the PhD process. I learned a great deal from them both about how to do good research and how to communicate it clearly. Their generosity with their time, their advice, and their patience as I found my way, have made me a better economist and shaped my thinking more generally. I am also grateful to Orazio Attanasio and Áureo de Paula for their insightful suggestions and guidance, and for helping me see my own work from different perspectives.

Thanks also to my coauthors on work included in this thesis: Bansi Malde (Chapter 2), Bryony Reich (Chapter 4), and William Elming and Jonathan Shaw (Chapter 5) — not least for permission to include our joint work in this thesis! Working with each of them has taught me new skills, as well as being a lot of fun.

During my PhD I had the good fortune to find many fantastic peers and mentors at UCL, too numerous to mention individually, who were an invaluable source of advice, computer tips, and coffee breaks. As well as this, I benefitted from being based at the Institute for Fiscal Studies, which provided both funding and a stimulating research environment.

Having thanked so many sources of encouragement in my professional life, it would be remiss of me to omit some more personal sources. Without the steadfast support of my parents I would never have had either the freedom or confidence to make the choices I did. Their advice, encouragement, humour, and occasional proof-reading, have all been an immense help during the PhD process. Most of all I am grateful to Rebecca for being my unwavering support and partner through all the ups and downs that come from a life in research. Rebecca, this thesis is dedicated to you.

Funding

This thesis was made possible by financial support from the Economic and Social Research Council (ESRC) through the Centre for the Microeconomic Analysis of Public Policy (CPP) at the Institute for Fiscal Studies (grant numbers ES/H021221/1 and ES/M010147/1), as part of the IFS PhD Scholars programme. Additional financial support for the work in Chapter 2 was received from the ESRC-NCRM Node 'Programme Evaluation for Policy Analysis' (grant number ES/I02574X/1). Additional financial support for the work in Chapter 3 was received from the Royal Economics Society (RES), through the RES Junior Fellowship Scheme. Additional financial support for the work in Chapter 4 was received from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. 313234. Finally, additional financial support for the work in Chapter 5 was received from the ESRC through the Tax Administration Research Centre (grant number ES/K005944/1). This funding is all gratefully acknowledged.

Data Provision

The US Census data used in Chapter 4 come from the Integrated Public Use Microdata Series (IPUMS-USA), and are housed at the Minnesota Population Center. Neither the owners nor distributors bear any responsibility for the interpretation of the data.

Analysis in Chapter 5 is based on statistical data from Her Majesty's Revenue and Customs (HMRC), which is Crown Copyright. The research datasets used may not exactly reproduce HMRC aggregates. The use of HMRC statistical data in this work does not imply the endorsement of HMRC in relation to the interpretation or analysis of the information.

Contents

List of Tables

5 The Dynamic Effects of Tax Audits

List of Figures

4 Melting Pot or Salad Bowl: The Formation of Heterogeneous Communities

5 The Dynamic Effects of Tax Audits

Chapter 1

Introduction

This thesis comprises four papers on the economics of networks. The first provides a comprehensive review of the literature studying the econometrics of networks, addressing questions of social effect identification, estimation of models of network formation, and techniques to account for measurement error. The following two chapters investigate the role of networks in different economic environments: informal risk sharing in Bangladesh, and among migrants to the United States in the early 1900s, respectively. To do this they first contain theoretical models which capture the key trade-offs in the environment studied, followed by empirical evidence to show the relevance of the proposed mechanisms. The fourth paper studies the dynamic impact – the impact on future income and tax declarations – from auditing taxpayers, as the first step towards studying the spillover effects of such audits on to other taxpayers linked to those who are audited.

Each chapter is a self-contained paper, with Chapters 3 to 5 having their own literature reviews (and Chapters 2 being a review). However, since each investigates a different area of networks in economics, I use this introduction to provide a brief overview of each chapter in turn, and then to pick out some overarching themes and commonalities.

Chapter 2 ("Empirical Methods for Networks Data", with Bansi Malde) reviews the current, burgeoning literature on econometric methods for analysing network data. Work on this topic aims to tackle three questions: (i) how does the action of one economic agent affect another; (ii) how do agents choose *which* other agents to interact with; and (iii) how can one collect data to answer these questions empirically. After setting out some notation, the chapter provides a survey of approaches to answering each of these questions in turn.

To study the effect of one agent on another, we lay out a common framework in which to think about different economic models of *social effects*. We explain why the general model cannot (point) identify the social effect parameter, and discuss important special cases where identification is possible. In each case we explain first the model(s) of behaviour which would generate the econometric specification. We then set out the conditions under which the social effect parameter can be identified, and some of the important practical issues that arise when trying to estimate the parameter in practice.

We next discuss the estimation of models of network formation. The key challenge is that the number of potential networks increases very rapidly in the number of agents, so as with models of social effects, some restrictions must be imposed on how the decision over one link can depend on the other possible link combinations. We first describe methods used in the statistics literature for estimating such models, using *only* information on what links are present, slowly relaxing the restrictions imposed to allow more general forms of correlation in link structure. We then discuss how information on covariates for the nodes can be used to learn more about the formation process. Work on estimating such models in a way that allows for interdependency in linking decisions is at the forefront of current research in this area, including several recent job market papers (Sheng, 2012; Mele, 2013; Leung, 2015; and Gualdani, 2017).

Finally, we examine the issues of sampling and measurement error. Both sampling and other sources of measurement error can lead to badly biased estimates of many network statistics, because nodes and links are interrelated in a way that means random sampling over one will not (generally) give a random sample over the other. We review the different approaches that have been developed to tackle this problem, depending on the network statistic of interest.

Chapter 3 ("Informal Insurance and Endogenous Poverty Traps") investigates why households in developing countries rarely use informal networks to borrow for investment, even though it is well-understood that they use these networks to borrow for consumption-smoothing. Recent work has shown high rates of return on investments in livestock, across a range of developing countries (Banerjee et al., 2015b; Bandiera et al., forthcoming). Since these are a well-understood technology, this raises a puzzle: why was more investment not already taking place? Existing work has shown the difficulties households face in saving for investment (Dupas and Robinson, 2013; Jakiela and Ozier, 2016), and in borrowing formally (Galor and Zeira, 1993; Banerjee and Newman, 1993; and Aghion and Bolton, 1997), but informal borrowing for consumption-smoothing is prevalent (e.g. Udry, 1994). The explanation I propose is that borowing for investment, unlike borrowing for consumption, changes the *future* value of the borrowing relationship, by changing the future need to engage in risk-sharing. Since investments will typically reduce the need for – and hence value of – risk-sharing, the lending households are worse off after making such a loan, as the borrower's incentives to repay and to remain in the relationship decline.

To formalise this argument, I build a simple model, in the spirit of Kocherlakota (1996) and Ligon et al. (2002). The model has four ingredients: (i) households have idiosyncratically variable incomes; (ii) households are risk-averse; (iii) households are able to walk away from any relationship that is not in their interest; and (iv) households are able to engage in lumpy investments. The first three are the elements of a standard model of risk-sharing with limited commitment. On adding this fourth ingredient, I can capture the mechanism described above. The model also generates additional predictions, including the existence of a *poverty trap* at the level of the risk-sharing network.

Using data from 1400 villages in Bangladesh, I provide empirical evidence of such a poverty trap, as well as for additional predictions of the model. Exploiting an asset transfer program that handed out cows to certain households in (randomly assigned) treated villages, I show that on average when more than 14% of the network receives cows, this is sufficient to push the network out of the trap, and to allow borrowing for the purpose of investment to take place. This highlights the importance of taking in to account spillovers when designing policies. In particular, the impact of such asset transfer programs could be increased dramatically if threshold effects were taken into account: a budget neutral reallocation of the same cows could have increased the additional investment by 44%.

Chapter 4 ("Melting Pot or Salad Bowl: The Formation of Heterogeneous Communities", with Bryony Reich) examines the trade-off between economic and social incentives for migrants. On the one hand, interaction with natives is likely to be economically valuable, since there may be gains from trade: gathering information, working for (or employing) natives, finding future customers for items produced. On the other hand, interacting with natives is likely to be more costly than interacting with migrants (of the same nationality), since natives will have differences in language and in other cultural and religious practices. We model the choices faced by combining a network formation game with a game-on-the-network. Individuals are of migrant or native type, and choose what links to form, as well as what actions to play in terms of cultural and non-cultural activities. Individuals have *ex ante* preferences over cultural activities, depending on their migrant status, but have no *ex ante* preference over the non-cultural ones. The gain from forming a link is increasing in the number of shared activities.

This relatively parsimonious model generates a prediction that is simultaneously obvious and surprising. It is obvious that larger migrant groups will be less likely to form links with natives and more likely to keep the migrant culture. It is perhaps not so immediately clear that, rather than a smooth tranisition, there will be a sharp threshold between starkly different equilibria. Migrants move from necessarily assimilating – interacting with and behaving like natives – when they are only a small group, to possibly being able to segregate – interact only with other migrants and maintain their own culture – when the migrant group size exceeds some threshold.[1] The presence of this threshold comes about because of strategic complementarity in migrant choices: if some migrants choose to switch action and adopt native practices, this reduces the gain to the remaining migrants of keeping their own culture and interacting with their own type, and also increases the gain from interacting with people who choose the native cultural action.

We test the prediction that such a threshold exists, as well as testing additional comparative static predictions by looking at how the location of the threshold varies, using data from the United States in the early 20th Century, at the end of the Age of Mass Migration. At this time there was a lot of variation in the share of the local community that were of various migrant types. We show that there does appear to be a sharp theshold in household actions as the migrant share of the community increases above 1/3, and verify the additional predictions. These results have important implications for migrant settlement policies, particularly in light of the current political climate: not only is it important to avoid 'ghettoising' communities, but the move from integration to segregated outcomes can be sudden and sharp. Additionally, for a given number of migrants, there can be gains from mixing distinct migrant types together, to reduce the attractiveness of segregation.

Chapter 5 ("The Dynamic Effects of Tax Audits", with William Elming and Jonathan Shaw) switches focus slightly, to study the role of tax audits on future tax reporting behaviour. In the search for revenue, governments have traditionally made choices about tax bases and tax rates. However, recent work by Kleven et al. (2011), Slemrod and Gillitzer (2013), and others, highlights the importance of a third policy choice: tax administration and enforcement. With a 'tax gap' of £36bn in 2014-15 (HMRC, 2016), worth around 7% of revenue collected, this has been an increasing focus in the UK. Tax compliance is typically enforced by auditing a proportion of returns, and punishing those who are found to have reported incorrectly. The effect of such audits is not only the direct revenue raised from fines and adjustments to tax paid, but also the indirect revenue raised by changes in reporting behaviour.

We study the impact of audits on future tax reporting – the *dynamic effect* – by exploiting the random audit program carried out by the UK tax authority, HMRC. This program is used to gather information on what characteristics are correlated with misreporting, to calibrate HMRC's targeted audit program. We compare randomly audited taxpayers with taxpayers who could have been, but were not, audited, and study how the difference in their tax declarations evolves over time. We show that audits have a significant dynamic component: that is, they raise additional revenue in the years after the audit has been completed. Unusually, we are able to follow

[1] In fact it is the share of the community made up of the migrant type, rather than their absolute number, that matters.

taxpayers for up to a decade after audit. This allows us to see that the additional revenue appears to last for (only) four tax years after the audited year, before becoming indistinguishable from the non-audited. We examine also how these impacts vary across income sources.[2] Consistent with Kleven et al. (2011), we see that income sources which are subject to third party reporting are much less likely to be misreported. These results suggest that audits are more valuable than would be understood simply from studying their direct effects. If this is not taken into account by policymakers, there is likely to be an undersupply of auditors and audits.

The study of networks is the common feature that links these four chapters. Chapter 2 reviews how to collect and analyse *network* data to answer different economic questions. Chapter 3 models and empirically studies the conditions under which a *network* of households engaged in informal risk-sharing can allow borrowing within the network for productive lumpy investments. Chapter 4 studies the *network* formation decisions, and other behaviours, of migrants, beginning with a theoretical model and then testing the predictions generated. Chapter 5 examines the dynamic effects of tax audits, as a prelude to the study of *network* spillovers: how does auditing an individual affect the declarations of family members, business partners, and co-workers.

Chapters 3 and 4 are also linked by a use of theory to discipline the empirical work. In particular, both models generate group-level predictions, and these are of non-linear relationships and threshold effects. In Chapter 3 this takes the form of a kink in the relationship between assets transferred and additional investment by the network. In Chapter 4 the prediction is a shift in the share of the migrant group that takes a particular action as migrant group size (relative to native size) increases. Both thresholds are at unknown locations, so similar econometric techniques are used to provide evidence for the mechanisms proposed.

In terms of data usage, each chapter uses data from entirely different sources. Despite this, there are important synergies. Chapters 4 and 5 both use administrative data: the US census, and data from the UK tax authority respectively. Such data are large, containing many millions of observations, but are relatively more limited in the availability of covariates. It is crucial then, to think carefully about how to handle issues of unobserved heterogeneity, since many characteristics important to any research question are likely not to be available. Chapters 3 and 5 also share some similarities in making use of a third-party's experiment: transfers of cows and random audits of taxpayers respectively. Although in both cases the experiments were not designed for the research questions under consideration, the exogenous variation from the experiment can usefully be exploited to provide a credible answer.

The next four chapters contain the work described above. The final chapter, Chapter 6, concludes with some directions for future research, following a similar structure to Chapter 2.

[2]Recent work in the US by DeBacker et al. (2015), which was developed concurrently with our work, also investigates this.

Chapter 2

Empirical Methods for Networks Data[1]

2.1 Introduction

Whilst anonymous markets have long been central to economic analysis, the role of networks as an alternative mode of interaction is increasingly being recognised. Networks might act as a substitute for markets, for example providing access to credit in the absence of a formal financial sector, or as a complement, for example transmitting information about the value of a product. Analysis that neglects the potential for such *social effects* when they are present is likely to mismeasure any effects of interest.

In this paper we provide an overview of econometric methods for working with network data – data on agents ('nodes') and the links between them – taking into account the peculiarities of the dependence structures present in this context. We draw on both the growing economic literature studying networks, and on research in other fields, including maths, computer science, and sociology. The discussion proceeds in three parts: (i) estimating social effects given a (conditionally) exogenous observed network; (ii) estimating the underlying network formation process, given only a single cross-section of data; and (iii) data issues, with a particular focus on accounting for measurement error, since in a network-context this can have particularly serious consequences.

The identification and estimation of social effects – direct spillovers from the characteristics or outcome of one agent to the outcome of others – are of central interest in empirical research

[1]This chapter is co-authored with Bansi Malde. We are grateful to Imran Rasul for his support and guidance. We also thank Richard Blundell, Andreas Dzemski, Toru Kitagawa, Aureo de Paula, and Yves Zenou for their useful comments and suggestions. Financial support from the ESRC-NCRM Node 'Programme Evaluation for Policy Analysis', Grant reference RES-576-25-0042 is gratefully acknowledged.

on networks in economics. Whilst researchers have tended to focus on the effects from the average characteristics and outcomes of network 'neighbours', different theoretical models will imply different specifications for social effects. In Section 2.3 we begin by setting out a common framework for social effects, which has as a special case the common 'linear-in-means' specification, as well as a number of other commonly used specifications. Since the general model is not identified, we then go through some important special cases, first outlining the theoretical model which generates the specification, before discussing issues related to identification of parameters.[2] For most of our discussion we focus on identification of the parameters using only observational data, since this is typically what researchers have available to them. We then go on to consider the conditions under which experimental variation can help weaken the assumptions needed to identify the parameters of interest.

The key challenge for credible estimation of social effects comes from the likely endogeneity of the network. Thus far, most of the empirical literature has simply noted this issue without tackling it head on, but more recently researchers have tried to tackle it directly. The main approach to doing this has been to search for instruments which change the probability of a link existing without directly affecting the outcome. Alternatively, where panel data are available, shocks to network structure – such as node death – have been used to provide exogenous variation. These approaches naturally have all the usual limitations: a convincing story must be provided to motivate the exclusion restriction, and where there is heterogeneity they identify only a local effect. Additionally, they rely on the underlying network formation model having a unique equilibrium. Without uniqueness we do not have a complete model, as we have not specified how an equilibrium is chosen. Hence a particular realisation of the instrument for some group of nodes is consistent with multiple resulting network structures, and a standard IV approach cannot be used.

This provides one natural motivation for the study of network formation models: being able to characterise and estimate a model of network formation would, in the presence of exclusion restrictions (or functional form assumptions motivated by theory), allow us to identify social effects using the predicted network. Formation models can also be useful for tackling measurement error, by imputing unobserved links. Finally, in some circumstances we might be interested in these models *per se*, for example to understand how we can influence network structure and hence indirectly the distribution of outcomes.

In Section 2.4 we consider a range of network formation models, drawing from literatures outside economics as well as recent work by economists, and show how these methods relate

[2] A different presentation of some of the material in this part of Section 2.3 can be found in Topa and Zenou (2015). Of the models we discuss, their focus is on two of the more common specifications used. Topa and Zenou (2015) compare these models to each other, and also to neighbourhood effect models, and discuss the relationship between neighbourhood and network models.

to each other. We first consider purely descriptive models that make use of only data on the observed links, and can be used to make in-sample predictions about unobserved links given the observed network structure. Next we turn to reduced form economic models, which make use of node characteristics in predicting links, but which do not allow for dependencies in linking decisions. Lastly we discuss the growing body of work estimating games of strategic network formation, which allow for such dependencies and so at least, in principle, can have multiple equilibria.[3]

The methods discussed until now have all assumed access to data on a population of nodes and all the relevant interconnections between them. However, defining and measuring the appropriate network is often not straightforward. In Section 2.5 we begin by discussing issues in network definition and measurement. We then discuss different sampling approaches: these are important because networks are comprised of interrelated nodes and links, meaning that a sampling strategy over one of these objects will define a non-random sampling process over the other. For example if we sample edges randomly, and compute the mean number of neighbours for the nodes to whom those edges belong, this estimated average will be higher than if the average were computed across all nodes, since nodes with many edges are more likely to have been included in the sample by construction. Next we discuss different sources of measurement error, and their implications for the estimation of network statistics and regression parameters. We end with an explanation of the various methods available to correct for these problems, and the conditions under which they can be applied.

Given the breadth of research in these areas alone, we naturally have to make some restrictions to narrow the scope of what we cover. In the context of social effects estimation, we omit entirely any discussion of *peer effects* where all that is known about agents' links are the groups to which they belong. A recent survey by Blume et al. (2010) more than amply covers this ground, and we direct the interested reader to their work. We also restrict our focus to linear models, which are appropriate for continuous outcomes but may be less suited to discrete choice settings such as those considered by Brock and Durlauf (2001) and Brock and Durlauf (2007). Similarly in our discussion of network formation, we do not consider in any detail the literature on the estimation of games. Although strategic models of network formation can be considered in this framework, the high dimension of these models typically makes it difficult to employ the same methods as are used in the game context. For readers who wish to know more about these methods, the survey paper by de Paula (2013) is a natural starting point. Finally, for a survey of applied work on networks in developing countries, see the review by Chuang and Schechter (2014).

[3] Another review of the material on strategic network formation is provided by Graham (2015).

We round off the paper with some concluding remarks, drawing together the various areas discussed, noting the limits of what we currently know about the econometrics of networks, and considering the potential directions for future research. Subsection 2.7.1 then provides detailed definitions of the various network measures and topologies that are mentioned in the text below.

2.2 Notation

Before we proceed, we first outline the notation we use throughout the paper.[4] We define a *network* or *graph* $g = (\mathscr{N}_g, \mathscr{E}_g)$ as a set of nodes, $\mathscr{N}_g$, and edges or links, $\mathscr{E}_g$.[5] The nodes represent individual agents, and the edges represent the links between pairs of nodes. In economic applications, nodes are usually individuals, households, firms or countries. Edges could be social ties such as friendship, kinship, or co-working, or economic ties such as purchases, loans, or employment relationships. The number of nodes present in g is $N_g = |\mathscr{N}_g|$, and the number of edges is $E_g = |\mathscr{E}_g|$. We define $\mathcal{G}_N = \{g : |\mathscr{N}_g| = N\}$ as the set of all possible networks on N nodes.

In the simplest case – the *binary network* – any (ordered) pair of nodes $i, j \in \mathscr{N}_g$ is either linked, $ij \in \mathscr{E}_g$, or not linked, $ij \notin \mathscr{E}_g$. If $ij \in \mathscr{E}_g$ then j is often described as being a *neighbour* of i. We denote by $nei_{i,g} = \{j : ij \in \mathscr{E}_g\}$ the *neighbourhood* of node i, which contains all nodes with whom i is linked. Nodes that are neighbours of neighbours will often be referred to as *'second degree neighbour'*. Typically it is convenient to assume that $ii \notin \mathscr{E}_g \ \forall i \in \mathscr{N}_g$. Edges may be directed, so that a link from node i to node j is not the same as a link from node j to node i; in this case the network is a *directed graph* (or *digraph*). In Section 2.4 we will at times find it useful to explicitly enumerate the edges; we denote by Λ this set of enumerated edges, with typical element l. Unlike $\mathscr{E}_g$, Λ is an ordered set, with order 12, 13, ...$N(N-1)$, so that we may use $(l-1)$ to denote the element in the set one position before l.

A more general case than the binary graph is that of a *weighted graph*, in which the edge set contains all possible combinations of nodes, other than to the node itself. That is, $\mathscr{E}_g = \{ij : \forall i, j \in \mathscr{N}_g, i \neq j\}$. Moreover, edges have *edge weights* $wei(i, j)$ which measure some metric of distance or link strength. Care is needed in interpreting the value of weights, as these differ by context. 'Distance' weighted graphs, which arise for example when weights represent transaction costs between two nodes, would typically have $wei^d(i, j) \in [0, \infty)$, with $wei^d(i, j) = \infty$ being equivalent to i and j being unconnected in the binary graph case. Conversely, 'strength' weighted graphs, where weights capture for example the frequency of interaction between agents, typically

[4] In Subsection 2.7 we provide further useful definitions.

[5] In a slight abuse of notation, we will also use g to index individual networks when data from multiple networks is available.

have $wei^s(i,j) \in [0, \bar{w}]$, with $wei^s(i,j) = 0$ being equivalent to i and j being unconnected in the binary graph case and $\bar{w} < \infty$.[6] Which definition is used depends on the context and application, but similar methods can be used for analysis in either case.[7]

Network graphs, whether directed or not, can also be represented by an *adjacency matrix*, $\boldsymbol{G}_g$, with typical element $G_{ij,g}$. This is an $N_g \times N_g$ matrix with the leading diagonal normalised to 0. When the network is binary, $G_{ij,g}= 1$ if $ij \in \mathscr{E}_g$, and 0 otherwise, while for weighted graphs, $G_{ij,g} = wei(i,j)$. We will use the notation $\boldsymbol{G}_{i,g}$ to denote the i^{th} row of the adjacency matrix $\boldsymbol{G}_g$, and $\boldsymbol{G}'_{i,g}$ to denote its i^{th} column.[8] Many models defined for binary networks make use of the row-stochastic adjacency matrix or *influence matrix,* $\tilde{\boldsymbol{G}}_g$.[9] Elements of this matrix are generally defined as $\tilde{G}_{ij,g} = {G_{ij,g}}/{\sum_j G_{ij,g}}$ if two agents are linked and 0 otherwise.

When we describe empirical methods for identifying and estimating social effects, we will frequently work with data from a number of network graphs. Graphs for different networks will be indexed, in a slight abuse of notation, by $g = 1, ..., M$, where M is the total number of networks in the data. Node-level variables will be indexed with $i = 1, ..., N_g$, where N_g is the number of nodes in graph g. Node-level outcomes will be denoted by $y_{i,g}$, while exogenous covariates will be denoted by the $1 \times K$ vector $\boldsymbol{x}_{i,g}$ and common network-level variables will be collected in the $1 \times Q$ vector, $\boldsymbol{z}_g$.

The node-level outcomes, covariates and network-level variables can be stacked for each node in a network. In this case, we will denote the stacked $N_g \times 1$ outcome vector as $\boldsymbol{y}_g$ and the $N_g \times K$ matrix stacking node-level vectors of covariates for graph g as $\boldsymbol{X}_g$. Common network-level variables for graph g will be gathered in the matrix $\boldsymbol{Z}_g = \boldsymbol{\iota}_g \boldsymbol{z}_g$ where $\boldsymbol{\iota}_g$ denotes an $N_g \times 1$ vector of ones. The adjacency and influence matrices for network g will be denoted by $\boldsymbol{G}_g$ and $\tilde{\boldsymbol{G}}_g$. At times we will also make use of the $N_g \times N_g$ identity matrix, $\boldsymbol{I}_g$, consisting of ones on the leading diagonal, and zeros elsewhere.

Finally, we introduce notation for vectors and matrices stacking together the network-level outcome vectors, covariate matrices and adjacency matrices for all networks in the data. $\boldsymbol{Y} = (\boldsymbol{y}_1^{'}, ..., \boldsymbol{y}_M^{'})^{'}$ is an $\sum_{g=1}^{M} N_g \times 1$ vector that stacks together the outcome vectors; $\boldsymbol{G} = diag\{\boldsymbol{G}_g\}_{g=1}^{g=M}$ denotes the $\sum_{g=1}^{M} N_g \times \sum_{g=1}^{M} N_g$ block-diagonal matrix with network-level adjacency matrices along the leading diagonal and zeros off the diagonal, and analogously $\tilde{\boldsymbol{G}} = diag\{\tilde{\boldsymbol{G}}_g\}_{g=1}^{g=M}$ (with similar dimensions as $\boldsymbol{G}$) for the influence matrices; and $\boldsymbol{X} = (\boldsymbol{X}_1^{'}, ..., \boldsymbol{X}_M^{'})^{'}$ and $\boldsymbol{Z} = (\boldsymbol{Z}_1^{'}, ..., \boldsymbol{Z}_M^{'})^{'}$ are

[6] In both of these examples, $wei(i,j) = wei(j,i)$. More generally this need not be true. For example, in some settings one might use 'flow weights' where $wei^f(i,j)$ represents the net flow of, say, resources from i to j. Then by definition $wei^f(i,j) = -wei^f(j,i)$, and the weighted adjacency matrix, defined shortly, is skew-symmetric.

[7] With distance weighted graphs, one must be careful in dealing with edges where $wei^d(i,j) = \infty$. A good approximation can usually be made by replacing infinity with an arbitrarily high finite value.

[8] $\boldsymbol{G}'_{i,g}$ is the i^{th} row of $\boldsymbol{G}'_g$, which is the i^{th} column of $\boldsymbol{G}_g$.

[9] A row stochastic (also called 'right stochastic' matrix) is one whose rows are normalised so they each sum to one.

respectively, $\sum_{g=1}^{M} N_g \times K$ and $\sum_{g=1}^{M} N_g \times Q$ matrices, that stack together the covariate matrices across networks. Finally, we define the vector $\boldsymbol{\iota}$ as a $\sum_{g=1}^{M} N_g \times 1$ vector of ones and the matrix $\boldsymbol{L} = diag\{\boldsymbol{\iota}_g\}_{g=1}^{g=M}$, as an $\sum_{g=1}^{M} N_g \times M$ matrix with each column being an indicator for being in a particular network.

2.3 Social Effects

Researchers are typically interested in understanding how the behaviour, choices and outcomes of agents are influenced by the agents that they interact with, *i.e.* by their neighbours. This section reviews methods that have been used to identify and estimate these social effects.[10] We consider a number of restrictions that would allow parameters of interest to be recovered, and place them into a broader framework. We focus on linear estimation models, which cover the bulk of methods used in practice.

We begin by providing a common organisational framework for the different empirical specifications that have been applied in the literature. Thereafter, we discuss in turn a series of commonly used specifications, the underlying theoretical models that generate them, and outline conditions for the causal identification of parameters with observational cross-sectional data. We then briefly discuss how experimental and quasi-experimental variation could be used to uncover social effects. Finally, we discuss some methods that can be applied to overcome confounding due to endogenous formation of edges, and discuss their limitations. A comprehensive overview of models of network formation is provided in Section 2.4.

We will use a specific example throughout this section to better illustrate the restrictions imposed by each of the different models and empirical specifications. Specifically, we will consider how we can use these methods to answer the following question: How is a teenager's schooling performance influenced by his friends? This is a widely studied question in the education and labour economics literatures, and is of great policy interest.[11]

We take as given throughout this section that the researcher knows the network(s) for which he is trying to estimate social effects and that he observes the entirety of this network without error. In Section 2.5 we will discuss how these data might be collected, and the consequences of having only a partial sample of the network and/or imperfectly measured networks.

[10] We leave aside the important issues of inference, in order to keep the scope of this survey manageable.
[11] See Sacerdote (2011) for an overview of this literature.

2.3.1 Organising Framework

Almost all (linear) economic models of social effects can be written as a special case of the following equation (written in matrix terms using the notation specified in Section 2.2):

$$\boldsymbol{Y} = \alpha\boldsymbol{\iota} + \boldsymbol{w_y}(\boldsymbol{G},\, \boldsymbol{Y})\boldsymbol{\beta} + \boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{w_x}(\boldsymbol{G},\, \boldsymbol{X})\boldsymbol{\delta} + \mathbf{Z}\boldsymbol{\eta} + \boldsymbol{L}\boldsymbol{\nu} + \boldsymbol{\varepsilon} \tag{2.1}$$

$\boldsymbol{Y}$ is a vector stacking individual outcomes of nodes across all networks.[12] $\boldsymbol{X}$ is a matrix of observable background characteristics that influence a node's own outcome and potentially that of others in the network. $\boldsymbol{G}$ is a block-diagonal matrix with the adjacency matrices of each network along its leading diagonal, and zeros on the off-diagonal. $\boldsymbol{w_y}(\boldsymbol{G},\, \boldsymbol{Y})$ and $\boldsymbol{w_x}(\boldsymbol{G},\, \boldsymbol{X})$ are functions of the adjacency matrix, and the outcome and observed characteristics respectively. These functions indicate how network features, interacted with outcomes and exogenous characteristics of (possibly all) nodes in the network, influence the outcome, $\boldsymbol{Y}$. The block-diagonal nature of $\boldsymbol{G}$ means that only the characteristics and outcomes of nodes in the same network are allowed to influence a node's outcome. $\boldsymbol{Z}$ is a matrix of observed network-specific variables; $\boldsymbol{\nu} = \{\nu_g\}_{g=1}^{g=M}$ is the associated vector of network-specific mean effects, unobserved by the econometrician but known to agents; and $\boldsymbol{\varepsilon}$ is a vector stacking the (unobservable) error terms for all nodes across all networks.

We make the following assumptions on the $\boldsymbol{\varepsilon}$ term:

$$\mathbb{E}[\varepsilon_{i,g}|\, \boldsymbol{X}_g, \boldsymbol{Z}_g, \boldsymbol{G}_g] = 0 \quad \forall\, i \in g;\; g \in \{1, ..., M\} \tag{2.2}$$

$$Cov[\varepsilon_{i,g}\varepsilon_{k,h}|\, \boldsymbol{X}_g, \boldsymbol{X}_h, \boldsymbol{Z}_g, \boldsymbol{Z}_h, \boldsymbol{G}_g, \boldsymbol{G}_h] = 0 \quad \forall\, i \in g;\; k \in h;\; g, h \in \{1, ..., M\};\; g \neq h \tag{2.3}$$

Equation 2.2 says that the error term for individual nodes in a network is mean independent of observed node-level characteristics of all network members, of network-level characteristics and of the network structure, as embodied in the adjacency matrix $\boldsymbol{G}_g$. The network, is in this sense assumed to be exogenous, conditional on individual-level observable characteristics and network-level observable characteristics. Later in Subsection 2.3.7 below, we will review some approaches taken to relax this assumption. In addition, Equation 2.3 implies that the error terms of all nodes, i and k in different networks, g and h, are uncorrelated conditional on observable characteristics of the nodes, the observable characteristics of the networks, and the structure of the network. Finally, note that no assumptions are imposed on the covariance of node-level error terms within the same network.

[12] We allow $\boldsymbol{Y}$ to be univariate, so individuals have only a single outcome. A recent paper by Cohen-Cole et al. (forthcoming) discusses how to relax this assumption, and provides some initial evidence that restricting outcomes to only a single dimension might be important in empirical settings.

In some cases, the following assumption is made on $\boldsymbol{\nu}$:

$$\mathbb{E}[\nu_g | \boldsymbol{X}_g, \boldsymbol{Z}_g, \boldsymbol{G}_g] = 0 \quad \forall g \in \{1, ..., M\} \tag{2.4}$$

That is, the network-level unobservable is mean independent of observable node- and network-level characteristics, and of the network. Many of the models that we consider below relax this assumption and allow for correlation between $\boldsymbol{\nu}$ and the other right hand side variables in Equation 2.1.

The social effect parameter that is most often of interest to researchers is $\boldsymbol{\beta}$ - the effect of a function of a node's neighbours' outcomes (*e.g.* an individual's friends' schooling performance) and the network. This is also known as the *endogenous effect*, to use the term coined by Manski (1993). This parameter is often of policy interest, since in many linear models, the presence of endogenous effects implies the presence of a social multiplier: the aggregate effects of changes in $\boldsymbol{X}$, $\boldsymbol{w_x}(\boldsymbol{G}, \boldsymbol{X})$, and $\boldsymbol{Z}$ are amplified beyond their direct effects, captured by $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, and $\boldsymbol{\eta}$. The parameters $\boldsymbol{\delta}$ and $\boldsymbol{\eta}$ are known as the *exogenous or contextual effect* while $\boldsymbol{\nu}$ captures a *correlated effect.*

This representation nests a range of models estimated in the economics literature:

1. *Local average models*: This model corresponds with $\boldsymbol{w_y}(\boldsymbol{G}, \boldsymbol{Y}) = \tilde{\boldsymbol{G}}\boldsymbol{Y}$ and $\boldsymbol{w_x}(\boldsymbol{G}, \boldsymbol{X}) = \tilde{\boldsymbol{G}}\boldsymbol{X}$, which arises when node outcomes are influenced by the average behaviour and characteristics of his direct neighbours. In our schooling example, this model implies that an individual's schooling performance is a function of the average schooling performance of his friends, his own characteristics, the average characteristics of his friends and some background network characteristics. This can apply, for example, when social effects operate through a desire for a node to conform to the behaviour of its neighbours. The identifiability of the parameters β, $\boldsymbol{\gamma}$, and $\boldsymbol{\delta}$ from the data available to a researcher depends on the structure of the network and the level of detail available about the network:[13]

   (a) With data containing information only on the broad peer group that a node belongs to and where a node can belong to a single group only (*e.g.* a classroom), it is common to assume that the node is directly linked with all other nodes in the same group and that there are no links between nodes in different groups. In this case, the peer group corresponds to the network. All elements of the influence matrix of a network g, $\tilde{\boldsymbol{G}}_g$, (including the diagonal) are set to $\frac{1}{N_g}$ where N_g is the number of agents within the network.[14] This generates the linear-in-means peer group model studied by Manski

[13]The parameter $\boldsymbol{\eta}$ can also be identified under the assumption that $\mathbb{E}[\boldsymbol{\nu}| \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{G}] = 0$.

[14]Note that in this case, since all nodes are linked to all others (including themselves), the total number of i's edges (or *degree*), $d_{i,g} = \sum_j G_{ij,g} = N_g \,\forall\, i \in g$. Hence by definition, all elements of $\tilde{\boldsymbol{G}}_g$ are set to $\frac{1}{N_g}$.

(1993) among others. Manski (1993) shows that identification of the parameter β is hampered by a simultaneity problem that he labels the *reflection problem*: it is not possible to differentiate whether the choices of a node i in the network influence the choices of node j, or vice versa. An alternative definition for $\tilde{\boldsymbol{G}}$ sets all diagonal terms of the network-level influence matrices, $\tilde{\boldsymbol{G}}_g$, to 0 and off-diagonal terms to $\frac{1}{N_g-1}$, which implies using the leave-self-out mean outcome as the regressor generating social effects. With this definition, identification of the parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\delta}$ is possible in some circumstances as shown by Lee (2007).[15] Identification issues related to this model with single peer groups have been surveyed in detail elsewhere, and thus will not be considered here. The interested reader should consult the comprehensive review by Blume et al. (2010).

(b) If instead detailed network data (*i.e.* information on nodes and the edges between them) are available, or if nodes belong to multiple partially overlapping peer groups, it may be possible to separately identify the parameters β, $\boldsymbol{\gamma}$, and $\boldsymbol{\delta}$ from a single cross-section of data. In this case, elements of the network-level influence matrices, $\tilde{\boldsymbol{G}}_g$ are defined as $\tilde{G}_{ij,g} = \frac{1}{d_{i,g}}$ when a link between i and j exists, where $d_{i,g}$ is the total number of i's links (or degree); and 0 otherwise. Identification results for observational network data have been obtained by Bramoullé et al. (2009). These are explored in more detail in Subsection 2.3.2 below.

2. *Local aggregate models*: When there are strategic complementarities or substitutabilities between a node's outcomes and the outcomes of its neighbours one can obtain the local aggregate model. In our schooling example, it may be more productive for an individual to put in more effort in studying if his friends also put in more effort, consequently leading to better schooling outcomes. In this case a node's outcome depends on the aggregate outcome of its neighbours. In the context of Equation 2.1, this implies that $\boldsymbol{w_y}(\boldsymbol{G}, \boldsymbol{Y}) = \boldsymbol{GY}$ and $\boldsymbol{w_x}(\boldsymbol{G}, \boldsymbol{X})$ is typically defined to be $\tilde{\boldsymbol{G}}\boldsymbol{X}$, implying that the outcome of interest is influenced by the *average* exogenous characteristics of a node's neighbours.[16] Identification and estimation of this model in observational networks data has been studied by Calvó-Armengol et al. (2009), Lee and Liu (2010) and Liu et al. (2014b). More details are provided in Subsection 2.3.3 below.

[15] Other solutions to the reflection problem have also been proposed, such as those by Glaeser et al. (1996), Moffitt (2001), and Graham (2008). Kwok (2013) provides a general study of the conditions under which identification of parameters can be achieved. He finds that network *diameter* – the length of the longest geodesic – is the key parameter in determining identification.

[16] This choice of definition for $\boldsymbol{w_x}(\boldsymbol{G}, \boldsymbol{X})$ is, to our understanding, not based on any explicit theoretical justification. It does, however, ease identification as $\boldsymbol{w_x}(.)$ and $\boldsymbol{w_y}(.)$ are now different functions of $\boldsymbol{G}$.

3. *Hybrid local models:* This class of models nests both the local average and local aggregate models. This allows the social effect to operate through both a desire for conformism and through strategic complementarities/substitutabilities. In the schooling example, the model implies that individuals may want to 'fit-in' and thus put in similar amounts of effort in studying as their friends, but their studying efforts may also be more productive if their friends also put in effort. Both of these channels then influence their schooling performance. In the notation of Equation 2.1, it implies that $\boldsymbol{w_y}(\boldsymbol{G}, \boldsymbol{Y}) = \mathbf{G}\boldsymbol{Y} + \tilde{\boldsymbol{G}}\boldsymbol{Y}$. As in the local average and aggregate models above, $\boldsymbol{w_x}(\boldsymbol{G},\, \boldsymbol{X})$ is typically defined to be $\tilde{\boldsymbol{G}}\boldsymbol{X}$. Identification and estimation of this model with observational data is studied by Liu et al. (2014a). See Subsection 2.3.4 for more details.

4. Networks may influence node outcomes (and consequently aggregate network outcomes) through more general features or functionals of the network. For instance, the DeGroot (1974) model of social learning implies that an individual's eigenvector centrality, which measures a node's importance in the network by how important its neighbours are, determines how influential it is in affecting the behaviour of other nodes.[17] In the schooling context, if an individual's friends are also friends of each other (a phenomenon captured by clustering), he may have to spend less time maintaining these friendships due to scale economies, allowing him more time for school work thereby leading to better schooling performance.

   Denoting a specific network statistic (such as eigenvector centrality in the social learning model above) by ω^r, where r indexes the statistic, we can specialise the term $\boldsymbol{w_y}(\boldsymbol{G},\, \boldsymbol{Y})\boldsymbol{\beta}$ in Equation 2.1 for node i in network g in a model with node-level outcomes as:

   - $\sum_{r=1}^{R} \omega^r_{i,g}\beta_r$: R different network statistics; or
   - $\sum_{r=1}^{R}\sum_{j\neq i} G_{ij,g}y_{j,g}\omega^r_{j,g}\beta_r$: the sum of neighbours' outcomes weighted by R different network statistics; or
   - $\sum_{r=1}^{R}\sum_{j\neq i} \tilde{G}_{ij,g}y_{j,g}\omega^r_{j,g}\beta_r$: the average of neighbours' outcomes weighted by R different network statistics.

   Analogous definitions are used for $\boldsymbol{w_x}(\boldsymbol{G},\, \boldsymbol{X})\boldsymbol{\delta}$. Models of this type have been estimated by Jackson et al. (2012) and Alatas et al. (2014).

   When researchers are interested in *aggregate* network outcomes, rather than node level

[17] Eigenvector centrality is a more general function of the network than those considered above, since it relies on the whole structure of the network.

outcomes, the following specification is typically estimated:

$$\bar{\boldsymbol{y}} = \phi_0 + \bar{\boldsymbol{w}}_{\bar{\boldsymbol{y}}}(\boldsymbol{G})\boldsymbol{\phi}_1 + \bar{\boldsymbol{X}}\boldsymbol{\phi}_2 + \bar{\boldsymbol{w}}_{\bar{\boldsymbol{X}}}(\boldsymbol{G},\, \bar{\boldsymbol{X}})\boldsymbol{\phi}_3 + \boldsymbol{u} \tag{2.5}$$

where $\bar{\boldsymbol{y}}$ is an $(M \times 1)$ vector stacking the aggregate outcome of the M networks, $\bar{\boldsymbol{w}}_{\bar{\boldsymbol{y}}}(\boldsymbol{G})$ is a matrix of $\bar{R}$ network statistics (*e.g.* average degree) that directly influence the outcome, $\bar{\boldsymbol{X}}$ is an $(M \times K)$ matrix of network-level characteristics (which could include network-averages of node characteristics) and $\bar{\boldsymbol{w}}_{\bar{\boldsymbol{X}}}(\boldsymbol{G},\, \bar{\boldsymbol{X}})$ is a term interacting the network-level characteristics with the network statistics. $\boldsymbol{\phi}_1$ captures how the network-level aggregate outcome varies with specific network features while $\boldsymbol{\phi}_2$ and $\boldsymbol{\phi}_3$ capture, respectively, the effects of the network-level characteristics and these characteristics interacted with the network statistic on the outcome. Models of this type have been estimated by among others, Banerjee et al. (2013), and are discussed further in Subsection 2.3.5.

In Subsections 2.3.2 to 2.3.5 below, we review methods relating to identification of the parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ and $\boldsymbol{\phi}_3$ in these models, under the assumption that the network is exogenous conditional on observable individual and network-level variables.[18] For each case discussed, we start by outlining a theoretical model that generates underlying the resulting empirical specification, and outline identification conditions using observational data.

Thereafter, in Subsection 2.3.6, we outline how experimental and quasi-experimental variation has been used to uncover social effects, and highlight some of the challenges faced in using such variation to uncover parameters of the structural models outlined in Subsections 2.3.2 to 2.3.4 below.

Subsection 2.3.7 outlines methods used by researchers to relax the assumption made in Equation 2.2: that the individual error term is mean independent of the network and observed individual and network-level characteristics. Dealing with endogenous formation of social links is quite challenging, and so most of the methods outlined in this section fail to satisfactorily deal with the identification challenges posed by endogenous network formation. Moreover, none of these methods deal with the issue of measurement error in the network. These issues are considered in Sections 2.4 and 2.5 respectively.

2.3.2 Local Average Models

In local average models, a node's outcome (or choice) is influenced by the average outcome of its neighbours. Thus, an individual's schooling performance is influenced by the average schooling performance of his friends. The outcome for node i in network g, $y_{i,g}$, is typically modelled as

[18] $\boldsymbol{\eta}$ can also be identified in some cases, particularly when the assumption $\mathbb{E}[\boldsymbol{\nu}|\, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{G}] = 0$ is imposed.

being influenced by its own observed characteristics, $\boldsymbol{x}_{i,g}$, scalar unobserved heterogeneity $\varepsilon_{i,g}$, observed network characteristics $\boldsymbol{z}_g$, unobserved network characteristic ν_g, and also the average outcomes and characteristics of neighbours. Below, we consider identification conditions when data are available from multiple networks, though some results apply to data from a single network.[19]

Stacking together data from multiple networks yields the following empirical specification, expressed in matrix terms:

$$\boldsymbol{Y} = \alpha\boldsymbol{\iota} + \beta\tilde{\boldsymbol{G}}\boldsymbol{Y} + \boldsymbol{X}\boldsymbol{\gamma} + \tilde{\boldsymbol{G}}\boldsymbol{X}\boldsymbol{\delta} + \boldsymbol{Z}\boldsymbol{\eta} + \boldsymbol{L}\boldsymbol{\nu} + \boldsymbol{\varepsilon} \tag{2.6}$$

where $\boldsymbol{Y}$, $\boldsymbol{\iota}$, $\boldsymbol{X}$, $\boldsymbol{Z}$, $\boldsymbol{L}$ and $\boldsymbol{\nu}$ are as defined previously; and $\tilde{\boldsymbol{G}}$ is a block diagonal matrix stacking network-level influence matrices along its leading diagonal, with all off-diagonal terms set to 0. The social effect, β, is a scalar in this model.

Given the simple empirical form of this model, it has been widely applied in the economics literature. Examples include:

- Understanding how the average schooling performance of an individual's peers influences the individual's own performance in a setting where students share a number of different classes (*e.g.* De Giorgi et al., 2010), or where students have some (but not all) common friends (*e.g.* Bramoullé et al., 2009).
- Understanding how non-market links between firms arising from company directors being members of multiple company boards influence firm choices on investment and executive pay (*e.g.* Patnam, 2013).

Although this specification is widely used in the empirical literature, few studies consider or acknowledge the form of its underlying economic model, even though parameter estimates are subsequently used to evaluate alternative policies and to make policy recommendations. Indeed, parameters are typically interpreted as in the econometric model of Manski (1993), whose parameters do not map back to 'deep' structural (*i.e.* policy invariant) parameters without an economic model.

An economic model that leads to this specification is one where nodes have a desire to conform to the average behaviour and characteristics of their neighbours (Patacchini and Zenou, 2012b). In our schooling example, conformism implies that individuals would want to exert as much effort in their school work as their friends so as to 'fit in'. Thus, if one's friends may want

[19] When data on only a single network are available, the empirical specification is as follows: $\boldsymbol{y}_g = \boldsymbol{a} + \beta\tilde{\boldsymbol{G}}_g\boldsymbol{y}_g + \boldsymbol{X}_g\boldsymbol{\gamma} + \tilde{\boldsymbol{G}}_g\boldsymbol{X}_g\boldsymbol{\delta} + \boldsymbol{\varepsilon}_g$, where $\boldsymbol{a} = \alpha\boldsymbol{\iota}_g + \boldsymbol{Z}_g\boldsymbol{\eta} + \boldsymbol{\iota}_g\nu_g$ in our earlier notation, capturing all of the network-level characteristics.

to exert no effort in their school work, the individual would also not want to exert any effort in his school work.

Below we show how this model leads to Equation 2.6. However, this is not the only economic model that leads to an empirical specification of this form: a similar specification arises from, for example, models of perfect risk sharing, where a well-known result is that under homogeneous preferences, when risk is perfectly shared, the consumption of risk-averse households will move with average household consumption in the risk sharing group or network (Townsend, 1994).

Conformism is commonly modelled by node payoffs that are decreasing in the distance between own outcome and network neighbours' average outcomes. Payoffs are also allowed to vary with an individual heterogeneity parameter, $\pi_{i,g}$, which captures the individual's ability or productivity associated with the outcome:[20]

$$U_i(y_{i,g}; \boldsymbol{y}_{-i,g}, \boldsymbol{X}_g, \tilde{\boldsymbol{G}}_{i,g}) = \left(\pi_{i,g} - \frac{1}{2} \left(y_{i,g} - 2\beta \sum_{j=1}^{N_g} \tilde{G}_{ij,g} y_{j,g} \right) \right) y_{i,g} \tag{2.7}$$

β in Equation 2.7 can be thought of as a taste for conformism. Although we write this model as though nodes are perfectly able to observe each others' actions, this assumption can be relaxed. In particular, an econometric specification similar to Equation 2.6 can be obtained from a static model with imperfect information (see Blume et al., 2013).

The best response function derived from the first order condition with respect to $y_{i,g}$ is thus:

$$y_{i,g} = \pi_{i,g} + \beta \sum_{j=1}^{N_g} \tilde{G}_{ij,g} y_{j,g} \tag{2.8}$$

Patacchini and Zenou (2012b) derive the conditions under which a Nash equilibrium exists, and characterise properties of this equilibrium.

The individual heterogeneity parameter, $\pi_{i,g}$, can be decomposed into a linear function of individual and network characteristics (both observed and unobserved):

$$\pi_{i,g} = \boldsymbol{x}_{i,g}\boldsymbol{\gamma} + \sum_{j=1}^{N_g} \tilde{G}_{ij,g} \boldsymbol{x}_{j,g} \boldsymbol{\delta} + \boldsymbol{z}_g \boldsymbol{\eta} + \nu_g + \varepsilon_{i,g} \tag{2.9}$$

Substituting for this in Equation 2.8, we obtain the following best response function for individual outcomes:

$$y_{i,g} = \beta \sum_{j=1}^{N_g} \tilde{G}_{ij,g} y_{j,g} + \boldsymbol{x}_{i,g}\boldsymbol{\gamma} + \sum_{j=1}^{N_g} \tilde{G}_{ij,g} \boldsymbol{x}_{j,g} \boldsymbol{\delta} + \boldsymbol{z}_g \boldsymbol{\eta} + \nu_g + \varepsilon_{i,g} \tag{2.10}$$

[20] Notice that in Equation 2.7, $\sum_{j=1}^{N_g} \tilde{G}_{ij,g} y_{j,g}$ is identical to the i^{th} row of $\tilde{\boldsymbol{G}}_g \boldsymbol{y}_g$, which appears in Equation 2.6.

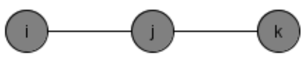
Then, stacking observations for all nodes in multiple networks, we obtain Equation 2.6, which can be taken to the data.

Bramoullé et al. (2009) study the identification and estimation of Equation 2.6 in observational data with detailed network information or data from partially overlapping peer groups.[21] To proceed further, one needs to make some assumptions on the relationship between the unobserved variables – $\boldsymbol{\nu}$ and $\boldsymbol{\varepsilon}$ – and the other right hand side variables in Equation 2.6.

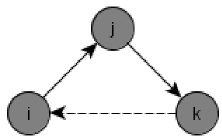
One specific assumption is that $\mathbb{E}[\boldsymbol{\varepsilon}|\boldsymbol{X}, \boldsymbol{Z}, \tilde{\boldsymbol{G}}] = 0$, *i.e.* the individual level error term, $\boldsymbol{\varepsilon}$, is assumed to be mean independent of the observed individual and network-level characteristics and of the network. The network level unobservable is also initially assumed to be mean independent of the right hand side variables, *i.e.* $\mathbb{E}[\boldsymbol{\nu}|\ \boldsymbol{X}, \boldsymbol{Z}, \tilde{\boldsymbol{G}}] = 0$; though this assumption will be relaxed further on.

Under these assumptions, the parameters $\{\alpha, \beta, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\eta}\}$ are identified if $\{\boldsymbol{I}, \tilde{\boldsymbol{G}}, \tilde{\boldsymbol{G}}^2\}$ are linearly independent. Identification thus relies on the network structure. In particular, the condition would not hold in networks composed only of cliques – subnetworks comprising of completely connected components – of the same size, and where the diagonal terms in the influence matrix, $\tilde{\boldsymbol{G}}$ are not set to 0. In this case, $\tilde{\boldsymbol{G}}^2$ can be expressed as a linear function of $\boldsymbol{I}$ and $\tilde{\boldsymbol{G}}$. Moreover, the model is then similar to the single peer group case of Manski (1993), and the methods outlined in Blume et al. (2010) apply.

In an undirected network (such as the in the left panel in Figure 2.1 below), this identification condition holds when there exists a triple of nodes (i, j, k) such that i is connected to j but not k, and j is connected to k. The exogenous characteristics of k, $\boldsymbol{x}_{k,g}$, directly affect j's outcome, but not (directly) that of i, hence forming valid instruments for the outcome of i's neighbours (*i.e.* j's outcome) in the equation for node i. Intuitively this method uses the characteristics of second-degree neighbours who are not direct neighbours as instruments for outcomes of direct neighbours.



(a) Intransitive triad in undirected network



(b) Intransitive triad in directed network

Figure 2.1: *Intransitive triad in an undirected network (left panel) and a directed network (right panel)*

[21] Similar identification results have been independently described by De Giorgi et al. (2010), who have data with overlapping peer groups of students who share a number of classes.

It is thus immediately apparent why identification fails in networks composed only of cliques: in such networks, there is no triple of nodes $(i,\ j,\ k)$ such that i is connected to j, and j is connected to k, but i is not connected to k.

In the directed network case, the condition is somewhat weaker, requiring only the presence of an intransitive triad: that is, a triple such that $ij \in \mathscr{E}$, $jk \in \mathscr{E}$ and $ik \notin \mathscr{E}$ (as in the right panel of Figure 1 above).[22] This is weaker than in undirected networks, which would also require that $ki \notin \mathscr{E}$.

As an example, consider using this method to identify the influence of the average schooling performance of an individual's friends on the individual, controlling for the individual's age and gender, the average age and gender of his friends, and some observed school characteristics (such as expenditure per pupil). Assume first that the underlying friendship network in this school is undirected as in the left panel of Figure 2.1, so that if i considers j to be his friend, j also considers i to be his friend. j also has a friend k who is not friends with i. We could then use the age and gender of k as instruments for the schooling performance of j in the equation for i. If instead, the network were directed as in the right panel of Figure 2.1, where the arrows indicate who is affected by whom (*i.e.* i is affected by j in the Figure, and so on), we can still use the age and gender of k as instruments for the school performance of j in the equation for i even though k is connected with i. This is possible since the direction of the relationship is such that k's school performance is affected by i's performance, but the converse is not true.

The identification result above requires that the network-level unobservable term be mean independent of the observed covariates, $\boldsymbol{X}$ and $\boldsymbol{Z}$, and of the network, $\tilde{\boldsymbol{G}}$. However, in many circumstances one might be concerned that unobservable characteristics of the network might be correlated with $\boldsymbol{X}$, so that $\mathbb{E}[\boldsymbol{\nu}|\boldsymbol{X}, \boldsymbol{Z}, \tilde{\boldsymbol{G}}] \neq 0$. For example, in our schooling context, when we take the network of interest to be constrained to be within the school, it is plausible that children with higher parental income will be in schools with teachers who have better unobserved teaching abilities, since wealthier parents may choose to live in areas with schools with good teachers. In this case, a natural solution when data on more than one network is available, is to include network fixed effects, $\boldsymbol{L}\tilde{\boldsymbol{\nu}}$ in place of the network-level observables, $\boldsymbol{Z}$, and the network-level unobservable, $\boldsymbol{L}\boldsymbol{\nu}$; where $\tilde{\boldsymbol{\nu}}$ is an $M \times 1$ vector that captures the network fixed effects.

Since the fixed effects themselves are generally not of interest, to ease estimation they are removed using a *within transformation*. This is done by pre-multiplying Equation 2.6 by $\boldsymbol{J}^{glob}$, a block diagonal matrix that stacks the network-level transformation matrices $\boldsymbol{J}_g^{glob} = \boldsymbol{I}_g -$

[22] Equivalently, a triple such $ji \in \mathscr{E}$, $kj \in \mathscr{E}$ and $ki \notin \mathscr{E}$ forms an intransitive triad.

$\frac{1}{N_g}(\boldsymbol{\iota}_g\boldsymbol{\iota}_g')$ along the leading diagonal, and off-diagonal terms are set to 0.[23] The resulting model, suppressing the superscript on $\boldsymbol{J}^{glob}$ for legibility, is of the following form:

$$\boldsymbol{JY} = \beta\boldsymbol{J\tilde{G}Y} + \boldsymbol{JX}\boldsymbol{\gamma} + \boldsymbol{J\tilde{G}X}\boldsymbol{\delta} + \boldsymbol{J}\boldsymbol{\varepsilon} \tag{2.11}$$

In this case, the identification condition imposes a stronger requirement on network structure. In particular, the matrices $\{\boldsymbol{I}, \boldsymbol{\tilde{G}}, \boldsymbol{\tilde{G}}^2, \boldsymbol{\tilde{G}}^3\}$ should be linearly independent. This requires that there exists a pair of agents (i, j) such that the shortest path between them is of length 3, that is, i would need to go through at least two other nodes to get to j (as in Figure 2.2 below). The presence of at least two intermediate agents allows researchers to use the characteristics of third-degree neighbours (neighbours-of-neighbours-of-neighbours who are not direct neighbours or neighbours-of-neighbours) as an additional instrument to account for the network fixed effect.

Figure 2.2: *Identification with network fixed effects*

Notes: *The picture on the left panel shows an undirected network with an agent l who is at least 3 steps away from i, while the picture on the right panel shows the same for a directed network.*

A concern that arises when applying this method is that of instrument strength. Bramoullé et al. (2009) find that this varies with graph *density*, *i.e.*, the proportion of node pairs that are linked; and the level of *clustering*, *i.e.* the proportion of node triples such that precisely two of the possible three edges are connected.[24] Instrument strength is declining in density, since the number of intransitive triads tends to zero. The results for clustering are non-monotone, and depend on density.

The discussion thus far has assumed that the network through which the endogenous social effect operates is the same as the network through which the contextual effect operates. It is possible to allow for these two networks to be distinct. This could be useful in a school setting, for instance, where contextual effects could be driven by the average characteristics of all students in the school, while endogenous effects by the outcomes of a subset of students who are friends. This might occur if the contextual effect operates through the level of resources the school has, which depends on the parental income of all students, whilst the peer learning might come only from friends.

[23] This is a *global* within transformation, which subtracts the average across the entire network from the individual's value. Alternatively, a *local* within transformation, $\boldsymbol{J}_g^{loc} = \boldsymbol{I}_g - \boldsymbol{\tilde{G}}_g$, can be used, which would subtract only the average of the individual's peers rather than the average for the whole network. The latter transformation has slightly stricter identification conditions than the former, since it does not make use of the fact that the network fixed effect is common across all network members, and not just among directly linked nodes.

[24] This definition is also referred to as the clustering coefficient.

Let $\boldsymbol{G}_{\boldsymbol{X},g}$ and $\boldsymbol{G}_{\boldsymbol{y},g}$ denote the network-level adjacency matrices through which, respectively, the contextual and endogenous effects operate. As before we define the block diagonal matrices $\boldsymbol{G}_{\boldsymbol{X}} = diag\{\boldsymbol{G}_{\boldsymbol{X},g}\}_{g=1}^{g=M}$ and $\boldsymbol{G}_{\boldsymbol{y}} = diag\{\boldsymbol{G}_{\boldsymbol{y},g}\}_{g=1}^{g=M}$. Blume et al. (2013) study identification of this model assuming that the two networks are (conditionally) exogenous and show that when the matrices $\boldsymbol{G}_{\boldsymbol{y}}$ and $\boldsymbol{G}_{\boldsymbol{X}}$ are observed by the econometrician, and at least one of δ and γ is non-zero, then the necessary and sufficient conditions for the parameters of Equation 2.6 to be identified are that the matrices $\boldsymbol{I}$, $\boldsymbol{G}_{\boldsymbol{y}}$, $\boldsymbol{G}_{\boldsymbol{X}}$ and $\boldsymbol{G}_{\boldsymbol{y}}\boldsymbol{G}_{\boldsymbol{X}}$ are linearly independent.

Although all parameters of interest can be identified by this method, the assumption that the network structure is conditionally exogenous is highly problematic. Though endogeneity caused by selection into a network can be overcome by allowing for group fixed effects which can be differenced out, endogenous formation of links within the network remains problematic and is substantially more difficult to overcome. Formally, the problem arises from the fact that agents' choices of with whom to link are correlated with unobservable (at least to the researcher) characteristics of both agents, so $\Pr(G_{ij,g} = 1|\varepsilon_{i,g}) \neq \Pr(G_{ij,g})$.

This means that the absence of a link between two nodes i and k may be correlated with $\varepsilon_{i,g}$ and $\varepsilon_{k,g}$, meaning that $\mathbb{E}[\varepsilon_{i,g}|\boldsymbol{X}_g, \boldsymbol{Z}_g, \boldsymbol{G}_g] \neq 0$.[25] Consequently the condition in Equation 2.2 no longer holds. This is problematic for the method of Bramoullé et al. (2009), where the absence of a link is used to identify the social effect, and this absence could be for reasons related to the outcome of interest, thereby invalidating the exclusion restriction. For instance, more motivated pupils in a school may choose to link with other motivated pupils; or individuals may choose to become friends with other individuals who share a common interest (such as an interest in reading, or mathematics) that is unobserved in the data available to the researcher. In such examples, the absence of a link is due to the unobserved terms of the two agents being correlated in a specific way rather than the absence of correlation between these terms. Solutions to this problem are considered in Subsection 2.3.7.

2.3.3 Local Aggregate Model

The local aggregate class of models considers settings where agents' utilities are a function of the aggregate outcomes (or choices) of their neighbours. Such a model applies to situations where there are strategic complementarities or strategic substitutabilities. For example:

- An individual's costs of engaging in crime may be lower when his neighbours also engage in crime (*e.g.* Bramoullé et al., 2014).[26].

[25] Similarly, $\mathbb{E}[\varepsilon_{k,g}|\boldsymbol{G}_g] \neq 0$.

[26] The games considered in both Bramoullé and Kranton (2007) and Bramoullé et al. (2014) are not strictly linear models, since there are corner solutions at zero.

- An agent is more likely to learn about a new product and how it works if more of his neighbours know about it and have used it.

The local aggregate model corresponds empirically to Equation 2.1 with $\boldsymbol{w_y}(\boldsymbol{G}, \boldsymbol{Y})$=$\boldsymbol{GY}$ and $\boldsymbol{w_x}(\boldsymbol{G}, \boldsymbol{X})$=$\tilde{\boldsymbol{G}}\boldsymbol{X}$, and a scalar social effect parameter, β. This specification can be motivated by the best responses of a game in which nodes have linear-quadratic utility and there are strategic complementarities or substitutabilities between the actions of a node and those of its neighbours. A model of this type has studied by Ballester et al. (2006).[27] In particular, the utility function for node i in network g takes the following form:

$$U_i(y_{i,g}; \boldsymbol{y}_{-i,g}, \boldsymbol{X}_g, \boldsymbol{G}_g) = \left(\pi_{i,g} - \frac{1}{2} y_{i,g} + \beta \sum_{j=1}^{N_g} G_{ij,g} y_{j,g} \right) y_{i,g} \tag{2.12}$$

where $y_{i,g}$ is i's action or choice, and $\pi_{i,g}$ is, as before, an individual heterogeneity parameter.[28] $\pi_{i,g}$ is parameterised as

$$\pi_{i,g} = \boldsymbol{x}_{i,g}\boldsymbol{\delta} + \sum_{j=1}^{n} \tilde{G}_{ij,g} \boldsymbol{x}_{j,g} \boldsymbol{\gamma} + \boldsymbol{z}_g \boldsymbol{\eta} + \nu_g + \varepsilon_{i,g}$$

so that individual heterogeneity is a function of a node's own characteristics, the *average* characteristics of its neighbours, network-level observed characteristics, and some unobserved network- and individual-level terms.

The quadratic cost of own actions means that in the absence of any network, there would be a unique optimal amount of effort the node would exert. $\beta > 0$ implies that neighbours' actions are complementary to a node's own actions, so that the node increases his actions in response to those of his neighbours. If $\beta < 0$, then nodes' actions are substitutes, and the reverse is true. Nodes choose $y_{i,g}$ so as to maximise their utility.

The best response function is:

$$y^*_{i,g}(\boldsymbol{G}_g) = \beta \sum_{j=1}^{n} G_{ij,g} y_{j,g} + \boldsymbol{x}_{i,g}\boldsymbol{\delta} + \sum_{j=1}^{n} \tilde{G}_{ij,g} \boldsymbol{x}_{j,g} \boldsymbol{\gamma} + \boldsymbol{z}_g \boldsymbol{\eta} + \nu_g + \varepsilon_{i,g} \tag{2.13}$$

Ballester et al. (2006) solve for the Nash equilibrium of this game when $\beta > 0$ and show that when $|\beta\omega_{max}(\boldsymbol{G}_g)| < 1$, where $\omega_{max}(\boldsymbol{G}_g)$ is the largest eigenvalue of the matrix $\boldsymbol{G}_g$, the equilibrium is unique and the equilibrium outcome relates to a node's Katz-Bonacich centrality, which is defined as $\boldsymbol{b}(\boldsymbol{G}_g, \beta) = (\boldsymbol{I}_g - \beta \boldsymbol{G}_g)^{-1}(\boldsymbol{\iota}_g)$.[29]

[27] Ballester et al. (2006) focus on the case where there are strategic complementarities. Bramoullé et al. (2014) study the case where there are strategic substitutabilities and characterise all equilibria of this game.

[28] Notice that $\sum_{j=1}^{N_g} G_{ij,g} y_{j,g} = \boldsymbol{G}_{i,g} \boldsymbol{y}_g$.

[29] A more general definition for Katz-Bonacich centrality is $\boldsymbol{b}(\boldsymbol{G}_g, \beta, a) = (\boldsymbol{I}_g - \beta \boldsymbol{G}_g)^{-1}(a \boldsymbol{G}_g \boldsymbol{\iota}_g)$, where $a > 0$ is a constant (Jackson, 2008).

Bramoullé et al. (2014) study the game with strategic substitutabilities between the action of a node and those of its neighbours. They characterise the set of Nash equilibria of the game and show that, in general, multiple equilibria will arise. A unique equilibrium exists only when $\beta|\omega_{min}(\boldsymbol{G}_g)| < 1$, where $\omega_{min}(\boldsymbol{G}_g)$ is the lowest eigenvalue of the matrix $\boldsymbol{G}_g$. When there are multiple equilibria possible, they must be accounted for in any empirical analysis. Methods developed in the literature on the econometrics of games may be applied here (Bisin et al., 2011a). See de Paula (2013) for an overview.

When a unique equilibrium exists, this theoretical set-up implies the following empirical model (stacking data from multiple networks):

$$\boldsymbol{Y} = \alpha\boldsymbol{\iota} + \beta\boldsymbol{G}\boldsymbol{Y} + \boldsymbol{X}\boldsymbol{\gamma} + \tilde{\boldsymbol{G}}\boldsymbol{X}\boldsymbol{\delta} + \boldsymbol{Z}\boldsymbol{\eta} + \boldsymbol{L}\boldsymbol{\nu} + \boldsymbol{\varepsilon} \tag{2.14}$$

which corresponds to Equation 2.1 with $\boldsymbol{w_y}(\boldsymbol{G}, \boldsymbol{Y})=\boldsymbol{G}\boldsymbol{Y}$ and $\boldsymbol{w_x}(\boldsymbol{G}, \boldsymbol{X})=\tilde{\boldsymbol{G}}\boldsymbol{X}$, and where all other variables and parameters are as defined above in Subsection 2.3.1.

Identification of Equation 2.14 using observational data has been studied by Calvó-Armengol et al. (2009), Lee and Liu (2010) and Liu et al. (2014b). They proceed under the assumption that $\mathbb{E}[\boldsymbol{\varepsilon}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{G}, \tilde{\boldsymbol{G}}] = 0$ and $\mathbb{E}[\boldsymbol{\nu}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{G}, \tilde{\boldsymbol{G}}] \neq 0$. That is, the node-varying error component is conditionally independent of node- and network-level observables and of the network, while the network-level unobservable could be correlated with node- and network-level characteristics and/or the network itself.

These assumptions imply a two-stage network formation process. First agents select into a network based on a set of observed individual- and network-level characteristics and some common network-level unobservables. Then in a second stage they form links with other nodes. There are no network-level unobservable factors that determine link formation once the network has been selected by the node. Moreover, there are no node-level unobservable factors that determine the choice of network or link formation within the chosen network.

To proceed, we assume that data is available for multiple networks. Then, as in Subsection 2.3.2, we replace the network-level observables, $\boldsymbol{Z}$, and the network-level unobservable, $\boldsymbol{L}\boldsymbol{\nu}$ in Equation 2.14 with network fixed effects, $\boldsymbol{L}\tilde{\boldsymbol{\nu}}$, where $\tilde{\boldsymbol{\nu}}$ is a $M \times 1$ vector that captures the network fixed effects.

To account for the fixed effect, a global within-transformation is applied, as in Subsection 2.3.2. This transformation is represented by the block diagonal matrix $\boldsymbol{J}^{glob}$ that stacks the following network-level transformation matrices – $\boldsymbol{J}_g^{glob} = \boldsymbol{I}_g - \frac{1}{N_g}(\boldsymbol{\iota}_g\boldsymbol{\iota}_g')$ – along the leading diagonal, with off-diagonal terms set to 0. Again we suppress the superscript on $\boldsymbol{J}^{glob}$ in the rest of this subsection. The resulting model, analogous to Equation 2.11, is:

$$\boldsymbol{JY} = \beta \boldsymbol{JGY} + \boldsymbol{JX}\boldsymbol{\gamma} + \boldsymbol{J}\tilde{\boldsymbol{G}}\boldsymbol{X}\boldsymbol{\delta} + \boldsymbol{J}\boldsymbol{\varepsilon} \tag{2.15}$$

The model above suffers from the reflection problem, since $\boldsymbol{Y}$ appears on both sides of the equation. However, the parameters of Equation 2.15 can be identified using linear IV if the deterministic part of the right hand side, $[\mathbb{E}(\boldsymbol{JGY}), \boldsymbol{JX}, \boldsymbol{J}\tilde{\boldsymbol{G}}\boldsymbol{X}]$, has full column rank. To see the conditions under which this is satisfied, we examine the term with the endogenous variable, $\mathbb{E}(\boldsymbol{JGY})$. Under the assumption that $|\beta\omega_{max}(\boldsymbol{G}_g)| < 1$, we obtain the following from the reduced form equation of Equation 2.14:

$$\begin{aligned}\mathbb{E}(\boldsymbol{JGY}) = \boldsymbol{J}(\boldsymbol{GX} + \beta\boldsymbol{G}^2\boldsymbol{X} + ...)\gamma + \boldsymbol{J}(\boldsymbol{G}\tilde{\boldsymbol{G}}\boldsymbol{X} + \beta\boldsymbol{G}^2\tilde{\boldsymbol{G}}\boldsymbol{X} + ...)\boldsymbol{\delta} \\ + \boldsymbol{J}(\boldsymbol{GL} + \beta\boldsymbol{G}^2\boldsymbol{L} + ...)\tilde{\boldsymbol{\nu}}\end{aligned} \tag{2.16}$$

We can thus see that if there is variation in node degree within at least one network g (which means that $\boldsymbol{G}_g$ and $\tilde{\boldsymbol{G}}_g$ are linearly independent), and the matrices $\{\boldsymbol{I}, \mathbf{G}, \tilde{\mathbf{G}}, \mathbf{G}\tilde{\mathbf{G}}\}$ are linearly independent with $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, and $\tilde{\boldsymbol{\nu}}$ each having non-zero terms, the parameters of Equation 2.14 are identified.[30] This is a special case of the Blume et al. (2013) result discussed earlier. Node degree ($\boldsymbol{GL}$), along with the total and average exogenous characteristics of the node's direct neighbours (*i.e.* $\boldsymbol{GX}$ and $\tilde{\boldsymbol{G}}\boldsymbol{X}$) and sum of the average exogenous characteristics of its second-degree neighbours (*i.e.* $\boldsymbol{G}\tilde{\boldsymbol{G}}\boldsymbol{X}$) can be used as instruments for the total outcome of the node's neighbours (i.e. $\boldsymbol{GY}$). The availability of node degree as an instrument can allow one to identify parameters without using the exogenous characteristics, $\boldsymbol{X}$, of second- or higher-degree network neighbours, which could be advantageous in some situations as we will see in Section 2.5 below.

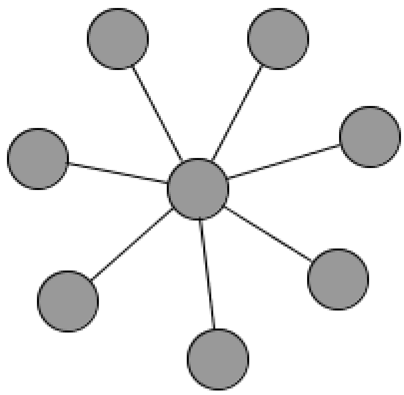
In terms of practical application, consider using this method to identify whether there are complementarities between the schooling performance of an individual and that of his friends, conditional on how own characteristics (age and gender), the composition of his friends (average age and gender), and some school characteristics. Then, if there are individuals in the same network with different numbers of friends, and the matrices $\{\boldsymbol{I}, \mathbf{G}, \tilde{\mathbf{G}}, \mathbf{G}\tilde{\mathbf{G}}\}$ are linearly independent, the individual's degree, along with the total and average characteristics of his friends (*i.e.* total and average age and gender) and the sum of the average age and gender of the individual's friends of friends can be used as instruments for the sum of the individual's friends' schooling performance.

[30] See Liu et al. (2014b) for a different identification condition that allows for some linear dependence among these matrices under additional restrictions.

Parameters can still be identified if there no variation in node degree within a network for all networks in the data, but there is variation in degree across networks. In this case, $\boldsymbol{G}_g = \bar{d}_g \tilde{\boldsymbol{G}}_g$ and $[E(\boldsymbol{JGY}), \boldsymbol{JX}, \boldsymbol{J\tilde{G}X}]$ has full column rank if the matrices $\{\boldsymbol{I}, \boldsymbol{G}, \tilde{\boldsymbol{G}}, \boldsymbol{G\tilde{G}}, \tilde{\boldsymbol{G}}^2, \boldsymbol{G\tilde{G}}^2\}$ are linearly independent and $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ each have non-zero terms.[31] Finally, when there is no variation in node degree within and across all networks in the data, parameters can be identified using a similar condition as encountered in Subsection 2.3.3 above: the matrices $\{\boldsymbol{I}, \tilde{\boldsymbol{G}}, \tilde{\boldsymbol{G}}^2, \tilde{\boldsymbol{G}}^3\}$ should be linearly independent.

It is possible to identify model parameters in the local aggregate model in networks where the local average model parameters cannot be identified. For example, in a star network (see Figure 2.3) there is no pair of agents that has a geodesic distance (*i.e.* shortest path) of 3 or more, so this fails the identification condition for the local average model (see Subsection 2.3.2 above). However, there is variation in node degree within the network and the matrices $\boldsymbol{I}_g, \mathbf{G}_g, \tilde{\mathbf{G}}_g, \mathbf{G}_g\tilde{\mathbf{G}}_g$ can be shown to be linearly independent, thus satisfying the identification conditions for the local aggregate model.

Figure 2.3: *Star Network*



2.3.4 Hybrid Local Models

The local average and local aggregate models embody distinct mechanisms through which social effects arise. One may be interested in jointly testing these mechanisms, and empirically identifying the most relevant one for a particular context. Liu et al. (2014a) present a framework nesting both the local aggregate and local average models, allowing for this.

The utility function for node i in network g that nests both the (linear) local aggregate and local average models has the following form:

[31] See Liu et al. (2014b) for a different identification condition that allows for some linear dependence among these matrices under additional restrictions.

$$U_i(y_{i,g}; \boldsymbol{y}_{-i,g}, \boldsymbol{X}_g, \tilde{\boldsymbol{G}}_{i,g}, \boldsymbol{G}_{i,g}) = \left(\pi_{i,g} + \beta_1 \sum_{j=1}^{N_g} G_{ij,g} y_{j,g} - \frac{1}{2} \left(y_{i,g} - 2\beta_2 \sum_{j=1}^{N_g} \tilde{G}_{ij,g} y_{j,g} \right) \right) y_{i,g} \tag{2.17}$$

where $\pi_{i,g}$ is node-specific observed heterogeneity, which affects the node's marginal return from the chosen outcome level $y_{i,g}$. A node's utility is thus affected by the choices of its neighbours through changing the marginal returns of its own choice (*e.g.* in a schooling context, an individual's studying effort is more productive if his friends also study), as in the local aggregate model, and by a cost of deviating from the average choice of its neighbours (*i.e.* individuals face a utility cost if they study when their friends don't study), as in the local average model.

The best reply function for a node i nests both the local average and local aggregate terms. Liu et al. (2014a) prove that under the condition that $\beta_1 \geq 0$, $\beta_2 \geq 0$ and $d_g^{max}\beta_1 + \beta_2 < 1$, where d_g^{max} is the largest degree in network g, the simultaneous move game has a unique interior Nash equilibrium in pure strategies.

The econometric model, assuming that the node-specific observed heterogeneity parameter takes the form $\pi_{i,g} = \boldsymbol{x}_{i,g}\boldsymbol{\gamma} + \sum_{j=1}^{N_g} \tilde{G}_{ij,g}\boldsymbol{x}_{j,g}\boldsymbol{\delta} + \boldsymbol{z}_g\eta_g + \nu_g + \varepsilon_{i,g}$, is as follows:

$$\boldsymbol{Y} = \alpha\boldsymbol{\iota} + \beta_1\boldsymbol{G}\boldsymbol{Y} + \beta_2\tilde{\boldsymbol{G}}\boldsymbol{Y} + \boldsymbol{X}\boldsymbol{\gamma} + \tilde{\boldsymbol{G}}\boldsymbol{X}\boldsymbol{\delta} + \boldsymbol{Z}\boldsymbol{\eta} + \boldsymbol{L}\boldsymbol{\nu} + \boldsymbol{\varepsilon} \tag{2.18}$$

using the same notation as before (see *e.g.* Subsection 2.3.1).

With data from only a single network it is not possible to separately identify β_1 and β_2 and hence test between the local aggregate and local average models (or indeed find that the truth is a hybrid of the two effects). Identification of parameters is considered when data from multiple networks are available under the assumption that $\mathbb{E}[\varepsilon_{i,g}|\boldsymbol{X}_g, \boldsymbol{Z}_g, \boldsymbol{G}_g, \tilde{\boldsymbol{G}}_g] = 0$ and $\mathbb{E}[\boldsymbol{\nu}_g|\boldsymbol{X}_g, \boldsymbol{Z}_g, \boldsymbol{G}_g, \tilde{\boldsymbol{G}}_g] \neq 0$. Thus, as in Subsections 2.3.2 and 2.3.3 above, the individual error term, $\varepsilon_{i,g}$ is assumed to be mean independent of node- and network-level observable characteristics and the network. The network-level unobservable, $\boldsymbol{\nu}_g$, by contrast is allowed to be correlated with node- and network-level characteristics and/or the network.

To proceed, as in the local average and local aggregate model, $\boldsymbol{Z}\boldsymbol{\eta}$ and $\boldsymbol{L}\boldsymbol{\nu}$ are replaced by a network-level fixed effect, $\boldsymbol{L}\tilde{\boldsymbol{\nu}}$, which is then removed using the global within-transformation, $\boldsymbol{J}^{glob}$. Again, we suppress the superscript on $\boldsymbol{J}^{glob}$. The resulting transformed network model is:

$$\boldsymbol{J}\boldsymbol{Y} = \beta_1\boldsymbol{J}\boldsymbol{G}\boldsymbol{Y} + \beta_2\boldsymbol{J}\tilde{\boldsymbol{G}}\boldsymbol{Y} + \boldsymbol{J}\boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{J}\tilde{\boldsymbol{G}}\boldsymbol{X}\boldsymbol{\delta} + \boldsymbol{J}\boldsymbol{\varepsilon} \tag{2.19}$$

When there is variation in the degree within a network g, then the reduced form equation of Equation 2.19 implies that $\boldsymbol{J}\boldsymbol{G}(\boldsymbol{I} - \beta_1\boldsymbol{G} - \beta_2\tilde{\boldsymbol{G}})^{-1}\boldsymbol{L}$ can be used as an instrument for the local aggregate term $\boldsymbol{JGY}$ and $\boldsymbol{J}\tilde{\boldsymbol{G}}(\boldsymbol{I} - \beta_1\boldsymbol{G} - \beta_2\tilde{\boldsymbol{G}})^{-1}\boldsymbol{L}$ can be used as an instrument for the local average term $\boldsymbol{J}\tilde{\boldsymbol{G}}\boldsymbol{Y}$. The model parameters may thus be identified even if there are no node-level exogenous characteristics, $\boldsymbol{X}$, in the model. Caution must be taken though when the model contains no exogenous characteristics, $\boldsymbol{X}$, since, in this case, the model may be only tautologically identified if $\beta_1 = 0$ (Angrist, 2013). The availability of such characteristics offers more possible IVs: in particular, the total and average exogenous characteristics of direct and indirect neighbours can be used as instruments. These are necessary for identification when all nodes within a network have the same degree, though average degree may vary across networks. In this case, parameters can be identified if the matrices $\{\boldsymbol{I}, \mathbf{G}, \tilde{\mathbf{G}}, \mathbf{G}\tilde{\mathbf{G}}, \tilde{\boldsymbol{G}}^2, \boldsymbol{G}\tilde{\boldsymbol{G}}^2, \tilde{\boldsymbol{G}}^3\}$ are linearly independent. If, however, all nodes in all networks have the same degree, it is not possible to identify separately the parameters β_1 and β_2.

This specification nests both the local average and local aggregate models, so a J-test for non-nested regression models can be applied to uncover the relevance of each mechanism. The intuition underlying the J-test is as follows: if a model is correctly specified (in terms of the set of regressors), then the fitted value of an alternative model should have no additional explanatory power in the original model, *i.e.* its coefficient should not be significantly different from zero. Thus, to identify which of the local average or local aggregate mechanisms is more relevant for a specific outcome, one could first estimate one of the models (*e.g.* the local average model), and obtain the predicted outcome value under this mechanism. In a second step, estimate the other model (in our example, the local aggregate model), and include as a regressor the predicted value from the other (*i.e.* local average) model. If the mechanism underlying the local average model is also relevant for the outcome, the coefficient on the predicted value will be statistically different from 0. The converse can also be done to test the relevance of the second model (the local aggregate model in our case). See Liu et al. (2014a) for more details.

2.3.5 Models with Network Characteristics

The models considered thus far allow for a node's outcomes to be influenced only by outcomes of its neighbours. However, the broader network structure may affect node- and aggregate network- outcomes through more general functionals or features of the network. Depending on the theoretical model used, there are different predictions on which network features relate to different outcomes of interest. For example, the DeGroot (1974) model of social learning implies that a node's eigenvector centrality, which measures its 'importance' in the network by how important its neighbours are, determines how influential it is in affecting the beliefs of other nodes.

Empirical testing and verification of the predictions of these theoretical models has greatly lagged the theoretical literature due to a lack of datasets with both information on network structure and socio-economic outcomes of interest. The recent availability of detailed network data from many contexts has begun to relax this constraint.

The following types of specification are typically estimated when assessing how outcomes vary with network structure, for node-level outcomes:

$$\boldsymbol{Y} = \boldsymbol{f_y}(\boldsymbol{w_y}(\boldsymbol{G},\, \boldsymbol{Y}), \boldsymbol{X}, \boldsymbol{w_x}(\boldsymbol{G},\, \boldsymbol{X}), \boldsymbol{Z}) + \boldsymbol{\varepsilon} \tag{2.20}$$

and network-level outcomes:

$$\bar{\boldsymbol{y}} = \boldsymbol{f_{\bar{y}}}(\bar{\boldsymbol{w}}_{\bar{\boldsymbol{y}}}(\boldsymbol{G}), \bar{\boldsymbol{X}}, \bar{\boldsymbol{w}}_{\bar{\boldsymbol{x}}}(\boldsymbol{G},\, \bar{\boldsymbol{X}})) + \boldsymbol{u} \tag{2.21}$$

$\boldsymbol{f_y}(.)$ and $\boldsymbol{f_{\bar{y}}}(.)$ are functions that specify the shape of the relationship between the network statistics and the node- and network-level outcomes. When $\boldsymbol{f_y}(.)$ is simply a linear index in its argument, Equation 2.22 remains nested in Equation 2.1. Though, in principle, the shape of $\boldsymbol{f_y}(.)$ should be guided by theory (where possible), through the rest of this subsection, we take $\boldsymbol{f_y}(.)$ to be a linear index in its argument. $\boldsymbol{w_y}(\boldsymbol{G},\, \boldsymbol{Y})$ includes R network statistics that vary at the node- or network-level and that may be interacted with $\boldsymbol{Y}$ while $\bar{\boldsymbol{w}}_{\bar{\boldsymbol{y}}}(\boldsymbol{G})$ contains the $\bar{R}$ network statistics in the network-level regression.[32] $\boldsymbol{X}$ is a matrix of observable characteristics of nodes, $\boldsymbol{w_x}(\boldsymbol{G},\, \boldsymbol{X})$ interacts network statistics with exogenous characteristics of nodes, and $\boldsymbol{Z}$ and $\bar{\boldsymbol{X}}$ are network-level observable characteristics. $\bar{\boldsymbol{w}}_{\bar{\boldsymbol{X}}}(\boldsymbol{G},\, \bar{\boldsymbol{X}})$ interacts network statistics with network-level observable characteristics.

The complexity of networks poses an important challenge in understanding how outcomes vary with network structure. In particular, there are no sufficient statistics that fully describe the structure of a network. For example, networks with the same average degree may vary greatly on dimensions such as density, clustering and average path length among others. Moreover, the adjacency matrix, $\boldsymbol{G}$, which describes fully the structure of a network, is too high-dimensional an object to include directly in tests of the influence of broader features of network structure. Theory can provide guidance on which statistics are likely to be relevant, and also on the shape of the relationship between the network statistic and the outcome of interest. A limitation though is that theoretical results may not be available (given currently known techniques) for outcomes one is interested in studying. This is a challenge faced by, for instance Alatas et al. (2014) who study how network structure affects information aggregation.

[32] The term $\boldsymbol{w_y}(\boldsymbol{G},\, \boldsymbol{Y})$ will be endogenous when network statistics are interacted with $\boldsymbol{Y}$.

Below we outline methods that have been applied to analyse the effects of features of network structure on socio-economic outcomes. We do so separately for node-level specifications and network-level specifications. This literature is very much in its infancy and few methods have been developed to allow for identification of causal parameters.

Node-Level Specifications

Many theoretical models predict how node-level outcomes vary with the 'position' of a node in the network, captured by node varying network statistics such as centrality; or with features of the node's local neighbourhood such as node clustering; or with the 'connectivity' of the network, represented by statistics that vary at the network-level such as network density.

A common type of empirical specification used in the literature correlates network statistics with some relevant socio-economic outcome of interest. This approach is taken by, for example, Jackson et al. (2012) who test whether informal favours take place across edges that are supported (*i.e.* that nodes exchanging a favour have a common neighbour), which is the prediction of their theoretical model.

This corresponds with $\boldsymbol{w_y}(\boldsymbol{G}, \boldsymbol{Y})$ in Equation 2.20 above being defined as $\boldsymbol{w_y}(\boldsymbol{G}, \boldsymbol{Y}) = \boldsymbol{\omega}$, where $\boldsymbol{\omega}$ is an $(\sum_{g=1}^{M} N_g \times R)$ matrix stacking $\boldsymbol{\omega}_{i,g}$, the $(1 \times R)$ node-level vector of network statistics of interest for all nodes in all networks, and $\boldsymbol{w_x}(.)$ being defined as $\boldsymbol{\iota}$. Here, $\boldsymbol{w_y}(\boldsymbol{G}, \boldsymbol{Y})$ is defined to be a function of the network only.

When $\boldsymbol{f_y}(.)$ is linear, the specification is as follows:

$$\boldsymbol{Y} = \alpha\boldsymbol{\iota} + \boldsymbol{\omega}\boldsymbol{\beta} + \boldsymbol{X}\boldsymbol{\gamma} + \mathbf{Z}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{2.22}$$

where the variables and parameters are as defined above and the parameter of interest is $\boldsymbol{\beta}$. Defining $\boldsymbol{W} = (\boldsymbol{\omega}, \boldsymbol{X}, \boldsymbol{Z})$, the key identification assumption is that $\mathbb{E}[\boldsymbol{\varepsilon}'\boldsymbol{W}] = 0$, that is that the right hand side terms are uncorrelated with the error term. This may not be satisfied if there are unobserved factors that affect both the network statistic (through affecting network formation decisions) and the outcome, $\boldsymbol{Y}$ or if the network statistic is mismeasured. Both of these are important concerns that we cover in detail in Sections 2.4 and 2.5 below.

In some cases, one may also be interested in estimating a model where an agent's outcome is affected by the outcomes of his neighbours, weighted by a measure of their network position. For example, in the context of learning about a new product or technology, the DeGroot (1974) model of social learning implies that nodes' eigenvector centrality determines how influential they are in influencing others' behaviour. Thus, conditional on the node's eigenvector centrality, its choices may be influenced more by the choices of his neighbours with high eigenvector centrality.

Thus, one may want to weight the influence of neighbours' outcomes on own outcomes by their eigenvector centrality, conditional on own eigenvector centrality. This implies a model of the following form:

$$\boldsymbol{Y} = \alpha\boldsymbol{\iota} + \boldsymbol{w_y}(\boldsymbol{G},\, \boldsymbol{Y})\boldsymbol{\beta} + \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\gamma}} + \boldsymbol{w_x}(\boldsymbol{G},\, \tilde{\boldsymbol{X}})\tilde{\boldsymbol{\delta}} + \mathbf{Z}\boldsymbol{\eta} + \boldsymbol{L}\boldsymbol{\nu} + \boldsymbol{\varepsilon} \tag{2.23}$$

$\boldsymbol{w_y}(\boldsymbol{G},\, \boldsymbol{Y})$ is an $\sum_g N_g \times R$ matrix, with the $(i, r)^{th}$ element being the weighted sum of i's neighbours' outcomes, $\sum\limits_{j \neq i} G_{ij,g} y_{j,g} \omega^r_{j,g}$ or $\sum\limits_{j \neq i} \tilde{G}_{ij,g} y_{j,g} \omega^r_{j,g}$, with weights $\omega^r_{j,g}$ being the neighbour's r^{th} network statistic. $\tilde{\boldsymbol{X}} = (\tilde{\boldsymbol{X}}'_1, \tilde{\boldsymbol{X}}'_2, ..., \tilde{\boldsymbol{X}}'_M)'$, where $\tilde{\boldsymbol{X}}_g = (\boldsymbol{X}_g, \boldsymbol{\omega}_g)$ is a matrix stacking together the network-level matrices of exogenous explanatory variables and network statistics of interest. $\boldsymbol{w_x}(\boldsymbol{G},\, \tilde{\boldsymbol{X}})$ could be defined as $\boldsymbol{G}\tilde{\boldsymbol{X}}$ or $\tilde{\boldsymbol{G}}\tilde{\boldsymbol{X}}$. Identification of parameters in this case is complicated by the fact that $\boldsymbol{w_y}(\boldsymbol{G},\, \boldsymbol{Y})$ is a (possibly non-linear) function of $\boldsymbol{Y}$, and thus endogenous. It may be possible to achieve identification using network-based instrumental variables, as done in Subsections 2.3.2, 2.3.3, and 2.3.4 above, though it is not immediately obvious how such an IV could be constructed. Future research is needed to shed light on these issues.

Network-level Specifications

Aggregate network-level outcomes, such as the degree of risk sharing or the aggregate penetration of a new product, may also be affected by how 'connected' the network is, or the 'position' of nodes that experience a shock or who first hear about a new product.

Empirical tests of the relationship between aggregate network-level outcomes and network statistics involves estimating specifications such as Equation 2.21, where the shape of the function $\boldsymbol{f_{\bar{y}}}(.)$ and the choice of statistics in $\bar{\boldsymbol{w}}_{\bar{\boldsymbol{y}}}(\boldsymbol{G}) = \bar{\boldsymbol{\omega}}$, where $\bar{\boldsymbol{\omega}}$ is an $(M \times \bar{R})$ matrix of network statistics, are, ideally, motivated by theory. With linear $\boldsymbol{f_{\bar{y}}}(.)$, this implies the following equation:

$$\bar{\boldsymbol{y}} = \phi_0 + \bar{\boldsymbol{\omega}}\boldsymbol{\phi}_1 + \bar{\boldsymbol{X}}\boldsymbol{\phi}_2 + \bar{\boldsymbol{w}}_{\bar{\boldsymbol{X}}}(\boldsymbol{G},\, \bar{\boldsymbol{X}})\boldsymbol{\phi}_3 + \boldsymbol{u} \tag{2.24}$$

where the variables are as defined after Equation 2.21. The parameter of interest is typically $\boldsymbol{\phi}_1$. Defining $\bar{\boldsymbol{W}} = (\boldsymbol{\omega}, \bar{\boldsymbol{X}}, \bar{\boldsymbol{w}}_{\bar{\boldsymbol{X}}}(\boldsymbol{G},\, \bar{\boldsymbol{X}}))$, the key identification assumption is that $\mathbb{E}\left[\boldsymbol{u}\bar{\boldsymbol{W}}\right] = 0$, which will not hold if there are unobserved variables in $\boldsymbol{u}$ that affect both the formation of the network and the outcome $\bar{\boldsymbol{y}}$; or if the network statistics are mismeasured. Recent empirical work, such as that by Banerjee et al. (2013), has used quasi-experimental variation to try and alleviate some of the challenges posed by the former issue in identifying the parameter $\boldsymbol{\phi}_1$.

Since this specification uses data at the network-level, estimation will require a large sample of networks in order to recover precise estimates of the parameters, even in the absence of endogeneity from network formation and mismeasurement of the network. This is a problem in practice, since as we will see below in Subsection 2.5.3, the difficulties and costs involved in collecting network data often mean that in practice researchers have data for a small number of networks only.

2.3.6 Experimental Variation

Subsections 2.3.2 to 2.3.5 above considered the identification of the social effect parameters using observational data. In this subsection, we consider identification of these parameters using experimental data. We focus on the case where a policy is assigned randomly to a sub-set of nodes in a network. Throughout we assume that the network is pre-determined and unchanged by the exogenously assigned policy.[33]

We focus the discussion on identifying parameters of the local average model specified in Subsection 2.3.2 above. The issues related to using experimental variation to uncover the parameters of the local aggregate model are similar. As outlined above, this model implies that a node's outcome is affected by the average outcome of its network neighbours, its own and network-level exogenous characteristics (which may be subsumed into a network fixed effect), and the average characteristics of its network neighbours. We are typically interested in parameters β, $\boldsymbol{\gamma}$, and $\boldsymbol{\delta}$ in the following equation:

$$\boldsymbol{Y} = \alpha\boldsymbol{\iota} + \beta\tilde{\boldsymbol{G}}\boldsymbol{Y} + \boldsymbol{X}\boldsymbol{\gamma} + \tilde{\boldsymbol{G}}\boldsymbol{X}\boldsymbol{\delta} + \boldsymbol{L}\tilde{\boldsymbol{\nu}} + \boldsymbol{\varepsilon} \tag{2.25}$$

where the variables are as defined above.

Throughout this subsection, we assume that the policy shifts outcomes for the nodes that directly receive the policy.[34] To proceed further, we first assume that a node that does not receive the policy (*i.e.* is untreated, to use the terminology from the policy evaluation literature), is only affected by the policy through its effects on the outcomes of the node's network neighbours. This implies the following model for the outcome $\boldsymbol{Y}$:

$$\boldsymbol{Y} = \alpha\boldsymbol{\iota} + \beta\tilde{\boldsymbol{G}}\boldsymbol{Y} + \boldsymbol{X}\boldsymbol{\gamma} + \tilde{\boldsymbol{G}}\boldsymbol{X}\boldsymbol{\delta} + \rho\boldsymbol{t} + \boldsymbol{L}\tilde{\boldsymbol{\nu}} + \boldsymbol{\varepsilon} \tag{2.26}$$

where $\boldsymbol{t}$ is the treatment vector, and ρ is the direct effect of treatment. We assume that $\mathbb{E}[\boldsymbol{\varepsilon}|\boldsymbol{X},\boldsymbol{Z},\tilde{\boldsymbol{G}},\boldsymbol{t}] = 0$. Moreover, random allocation of the treatment implies that $\boldsymbol{t} \perp\!\!\!\perp \boldsymbol{X},\boldsymbol{Z},\tilde{\boldsymbol{G}},\boldsymbol{\varepsilon}$.

Applying the same within-transformation as in Subsection 2.3.2 above to account for the network-level fixed effect leads to the following specification:

[33] This assumption is not innocuous. Comola and Prina (2014) provide an example where the policy intervention does change the network.

[34] Below, we will consider identification conditions in the case where a node may be affected by the treatment status of his network neighbours even if their outcomes do not shift in response to the treatment.

$$\boldsymbol{JY} = \alpha \boldsymbol{J\iota} + \beta \boldsymbol{J\tilde{G}Y} + \boldsymbol{JX\gamma} + \boldsymbol{J\tilde{G}X\delta} + \rho \boldsymbol{Jt} + \boldsymbol{J\varepsilon} \tag{2.27}$$

We can use instrumental variables to identify β as long as the deterministic part of the right hand side of Equation 2.27, $[\mathbb{E}(\boldsymbol{J\tilde{G}Y}), \boldsymbol{JX}, \boldsymbol{J\tilde{G}X}]$ has full column rank. $\boldsymbol{JX}$ and $\boldsymbol{J\tilde{G}X}$ can be used as instruments for themselves. We thus need an instrument for $\mathbb{E}[\boldsymbol{J\tilde{G}Y}]$. We use the following expression for $\boldsymbol{J\tilde{G}Y}$, derived from the reduced form of Equation 2.26 under the assumption that $|\beta| < 1$, to construct instruments:

$$\begin{aligned}\mathbb{E}[\boldsymbol{J\tilde{G}Y}] = \boldsymbol{J\tilde{G}}\sum_{s=0}^{\infty} \beta^s \boldsymbol{\tilde{G}}^s \alpha \boldsymbol{\iota} + \boldsymbol{J}(\boldsymbol{\tilde{G}X\gamma} + \beta \boldsymbol{\tilde{G}}^2 \boldsymbol{X\gamma} + ...) + \boldsymbol{J}(\boldsymbol{\tilde{G}}^2 \boldsymbol{X\delta} + \beta \boldsymbol{\tilde{G}}^3 \boldsymbol{X\delta} + ...) \\ + \boldsymbol{J}(\rho \boldsymbol{\tilde{G}t} + \beta \rho \boldsymbol{\tilde{G}}^2 t + ...)\end{aligned} \tag{2.28}$$

From this equation, we can see that $\boldsymbol{\tilde{G}t}$, the average treatment status of a node's network neighbours, does not appear in Equation 2.26. It can thus be used as an instrument for $\boldsymbol{\tilde{G}Y}$, either in addition to, or as an alternative to $\boldsymbol{\tilde{G}}^2\boldsymbol{X}$ and $\boldsymbol{\tilde{G}}^3\boldsymbol{X}$, the average characteristics of the node's second- and third-degree neighbours. Thus, the policy could be used to identify the model parameters, albeit under a strong assumption on who it affects.[35]

In many cases, however, the assumption that the policy affects a node's outcome only if it is directly treated may be too strong. The treatment status of a node's neighbours could affect its outcome even when the neighbours' outcomes do not shift in response to receiving the policy. An example of such a case, studied by Banerjee et al. (2013), is when the treatment involves providing individuals with information on a new product, and the outcome of interest is the take-up of the product. Then neighbours' treatment status could affect the individual's own adoption decision by (1) shifting his neighbours' decision (endorsement effects), and also (2) through neighbours passing on information about the product and letting the individual know of its existence (diffusion effect).[36] In this case, a more appropriate model would be as follows:

$$\boldsymbol{Y} = \alpha \boldsymbol{\iota} + \beta \boldsymbol{\tilde{G}Y} + \boldsymbol{X\gamma} + \boldsymbol{\tilde{G}X\delta} + \rho \boldsymbol{t} + \boldsymbol{\tilde{G}t}\mu + \boldsymbol{\varepsilon} \tag{2.29}$$

where ρ captures the direct treatment effect, *i.e.* the effect of a node itself being treated, and μ is the direct effect of the average treatment status of social contacts. This highlights the limits to using exogenous variation from randomised experiments to identify social effect parameters. We might want to use the exogenous variation in the average treatment allocation of a node's neighbours, $\boldsymbol{\tilde{G}t}$, as an instrument for neighbours' outcomes, $\boldsymbol{\tilde{G}Y}$. However, this will identify β

[35]Similar results can be shown for the local aggregate model when $|\beta\omega_{max}(\boldsymbol{G})| < 1$. However, as shown above, node degree can also be used as an additional instrument in this model.

[36]The study of how to use these effects to maximise the number of people who adopt relates closely to study of the 'key player' in work by Ballester et al. (2006) and Liu et al. (2014b).

only under the assumption that $\mu = 0$, *i.e.* there is no direct effect of neighbours' treatment status. This rules out economic effects such as the diffusion effect.

We can still make use of the treatment effect for identification, by using the average treatment status of a node's second-degree (and higher-degree) neighbours, $\tilde{\boldsymbol{G}}^2\boldsymbol{t}$, as instruments for the average outcome of his neighbours ($\tilde{\boldsymbol{G}}\boldsymbol{Y}$). This is the same identification result as discussed earlier, from Bramoullé et al. (2009), and simply treats $\tilde{\boldsymbol{G}}^2\boldsymbol{t}$ in the same way the other covariates of second-degree neighbours, $\tilde{\boldsymbol{G}}^2\boldsymbol{X}$. Such instruments rely not only on variation in treatment status, but also on the network structure, with identification not possible for certain network structures as we saw in Subsection 2.3.2.[37]

Thus far, we have discussed how exogenous variation arising from the random assignment of a policy can be used to identify the social effect associated with a specific model – the local average model – which, as we saw, arises from an economic model where agents conform to their peers. In empirical work, though, it is common for researchers to directly include the average treatment status of network neighbours, rather than their average outcome, as a regressor in the model. In other words, the following type of specification is usually estimated:

$$\boldsymbol{Y} = b_1\boldsymbol{\iota} + b_2\tilde{\boldsymbol{G}}\boldsymbol{t} + \boldsymbol{X}\boldsymbol{b}_3 + \tilde{\boldsymbol{G}}\boldsymbol{X}\boldsymbol{b}_4 + b_5\boldsymbol{t} + \boldsymbol{u} \tag{2.30}$$

A non-zero value for b_2 is taken to indicate the presence of some social effect. However, without further modelling, it is not possible to shed light on the exact mechanism underlying this social effect, or the value of some 'deep' structural parameter.

2.3.7 Identification of Social Effects with Endogenous Links

In the previous subsections we focused on the identification of social effects under the assumption that the edges along which the effects are transmitted are exogenous. By exogenous we mean that the probability that agent i forms an edge with agent j is mean independent of any unobservables that might influence the outcome of interest for any individual in our social effects model. Formally, we assumed $\mathbb{E}[\boldsymbol{\varepsilon}|\boldsymbol{X}, \boldsymbol{Z}, \tilde{\boldsymbol{G}}] = 0$.[38]

However, in many contexts this may not be hold. Suppose we have observational data on farming practices amongst farmers in a village, and want to understand what features influence take-up of a new practice. We might see that more connected farmers are more likely to take up the practice. However, without further analysis we cannot necessarily interpret this as being *caused* by the network.

[37] Note that instruments based on random treatment allocation and network structure (*e.g.* $\tilde{\boldsymbol{G}}\boldsymbol{t}$ and $\tilde{\boldsymbol{G}}^2\boldsymbol{t}$) may be more plausible than those based on the exogenous characteristics, $\boldsymbol{X}$, and the network structure (*e.g.* $\tilde{\boldsymbol{G}}^2\boldsymbol{X}$), since $\boldsymbol{t}$ has been randomly allocated, whereas $\boldsymbol{X}$ need not be.

[38] Goldsmith-Pinkham and Imbens (2013) suggest a test for endogeneity.

One possibility is that there is some underlying correlation in the unobservables of the outcome and connection equations. More risk-loving people, who might be more likely to take up new farming practices, may also be more sociable, and thus have more connections. The endogeneity problem here comes from not being able to hold constant risk-preferences. Hence the coefficient on the network measures is not independent of this unobserved variable. This problem could be solved if we could find an instrument: something correlated with network connections that is unrelated to risk-preferences.

Another possibility is that connections were formed explicitly because of their relationship with the outcome. If agents care about their outcome $y_{i,g}$, and if the network has some impact on $y_{i,g}$, then they have incentives to be strategic in choosing the links in which they are involved. Suppose agents' utility (or profit) varies with $y_{i,g}$, but that some agents have a higher marginal utility from increases in $y_{i,g}$. Agents have incentives to manipulate the parts of the network they are involved in *i.e.* the elements of the i^{th} row and i^{th} columns of $\boldsymbol{G}_g$ – $\{\boldsymbol{G}_{i,g}, \boldsymbol{G}'_{i,g}\}$ – to try to maximise $y_{i,g}$. Moreover, if links are costly, but there is heterogeneity in the agents' valuations of $y_{i,g}$, then agents who value $y_{i,g}$ most should form more costly links, and have higher $y_{i,g}$, but the network is a consequence and not a cause of the individual value for $y_{i,g}$.

Returning to the farming example, some agents may have a greater preference for taking up new technologies. If talking to others is costly, but can help in understanding the new techniques, these farmers will form more connections. Now the unobservable factors which influence the outcome – preference for take up – will be be correlated with the number of connections. Unlike the previous case, this time we cannot find an 'instrumental' solution: it is the same unobservable driving both y_i and $\boldsymbol{G}_i$.

To overcome this issue experimentally one would need to be able to assign links in the network. However, with the exception of rare examples (including one below), this is difficult to achieve in practice. Additionally there can be external validity issues, as knowing the effect that randomly assigned networks have may not be informative about what effect non-randomly assigned networks have. Alternatively, one can randomly assign treatment status, as discussed in Subsection 2.3.6.[39]

Carrell et al. (2013) provide a cautionary example of the importance of considering network formation when using estimated social effects to inform policy reform. Carrell et al. (2009) use data from the US Air Force Academy, where students are randomly assigned to classrooms. They estimate a non-linear model of peer effects, implicitly assuming that conditional on classroom

[39] However, when the network is allowed to be endogenous, one needs to make (implicit) assumptions on the network formation process in order to obtain causal estimates. For example, if we assume that the network formation process is such that nodes with similar observed and unobserved characteristics hold similar positions in the resulting network, we can obtain causal estimates if we compare outcomes of nodes with similar network characteristics and different levels of indirect treatment exposure – i.e. exposure to the treatment through their neighbours. See Manski (2013) for more discussion on these issues.

assignment friendship formation is exogenous. They find large and significant peer effects in maths and English test scores, and some non-linearity in these effects. Carrell et al. (2013) use these estimated effects to 'optimally assign' a random sample of students to classrooms, with the intention of maximising the achievement of lower ability students. However, test performance in the 'optimally assigned' classrooms is worse than in the randomly assigned classrooms. They suggest that this finding comes from not taking into account the structure of the linkages between individuals within classrooms.[40]

Instrumental Variables

In the first example above, the outcome y was determined by an equation of the form of Equation 2.1, where the network $\boldsymbol{G}$ was determined potentially by some of the observables already in Equation 2.1 and also the unobservables $\boldsymbol{u}$, and $\mathbb{E}[\boldsymbol{\varepsilon}|\boldsymbol{X}, \boldsymbol{Z}, \tilde{\boldsymbol{G}}] \neq 0$. The failure of the mean independence assumption prevents us from identifying the parameters of Equation 2.1 in the ways suggested previously.

If our interest is in identifying only those parameters, one (potential) solution to the problem is to randomly assign the network structure. However, this is typically prohibitively difficult to enforce in real world settings. It is also unlikely to be representative of the edges people actually choose (see for example Carrell et al., 2013).[41]

Alternatively we can attempt to overcome the endogeneity of the network by taking an instrumental variables (IV) approach and finding an exclusion restriction. Here one needs to have a covariate that affects the structure of the network in a way relevant to the outcome equation – something which changes $\boldsymbol{w_y}(\boldsymbol{G}, \boldsymbol{Y})$ – but is excluded from the outcome equation itself. For example, if the outcome equation has only in-degree as a network covariate, then one needs to find a covariate that is correlated with in-degree but not the outcome. If instead the outcome equation included some other network covariate, for example Bonacich centrality, a different variable might be appropriate as an instrument.

Mihaly (2009) takes this approach. In trying to uncover the effect of popularity – as measured by various network statistics – on the educational outcomes of adolescents in the US, she uses an interaction between individual and school characteristics as an instrument for popularity.[42] This is a valid instrument if the composition of the school has no direct effect on educational

[40]Booij et al. (2015) have a different interpretation of this result. They suggest that the problem with the assignment based on the results of Carrell et al. (2009) is that the peer groups constructed fall far outside the support of the data used. Hence predictions about student performance come from extrapolation based on the functional form assumptions used, which should have been viewed with caution.

[41]In the models discussed this means we might observe outcomes that wouldn't be seen without manipulation, because we have changed the support of $\boldsymbol{G}$. In interpreting these results in the context of unmanipulated data we need to be cautious, since we are relying heavily on the functional form assumptions as extrapolate outside the support of what we observe.

[42]She uses four definitions of popularity: in-degree, network density (which only varies between networks), eigenvector centrality, and Bonacich centrality.

attainment (something which the education literature suggests is unlikely), but does affect all of the measures of popularity.

As ever with instrumental variables, the effectiveness of this approach relies on having a good instrument: something which has strong predictive power for the network covariate but does not enter the outcome equation directly. As noted earlier, if individuals care about the outcome of interest, they will have incentives to manipulate the network covariate. Hence such a variable will generally be easiest to find when there are some exogenous constraints that make particular edges much less likely to form than others, despite their strong potential benefits. For example Munshi and Myaux (2006) consider the role of strong social norms that prevent the formation of cross-religion edges even where these might otherwise be very profitable, when studying fertility in rural Bangladesh. The restrictions on cross-religion connections means that having different religions is a strong predictor that two women are not linked. Alternatively, secondary motivations for forming edges that are unrelated to the primary outcome could be used to provide an independent source of variation in edge formation probabilities.[43]

It is important to note that this type of solution can only be employed when the underlying network formation model has a unique equilibrium. Uniqueness requires that there is only one network structure consistent with the (observed and unobserved) characteristics of the agents and environment. However, when multiple equilibria are possible, which will generally be the case if the incentives for a pair of agents to link depend on the state of the other potential links, IV solutions cannot be used. We discuss further in Section 2.4 issues of uniqueness in network formation models, and how one might estimate the formation equation in these circumstances.

One should also be aware, when interpreting the results, that if there is heterogeneity in β then this approach delivers a local average treatment effect (LATE). This is a particular weighted average of the individual-specific β's, putting more weight on those for whom the instrument (in our example, school composition) creates most variation in the network characteristic. Hence if the people whose friendship decisions are most affected by school characteristics are also those who, perhaps, are most affected by their friends' outcomes, then the estimated social effect will be higher than the average social effect across all individuals.

Jointly model formation and social effects

In our second example at the beginning of Subsection 2.3.7 we considered the case where the outcome y was determined by an equation of the form of Equation 2.1, and the network $\boldsymbol{G}$ was strategically chosen to maximise the (unobserved) individual return from this outcome, subject to unobserved costs of forming links. Here the endogeneity comes from $\boldsymbol{G}$ being a function of u.

[43] An application of this idea is provided by Cohen-Cole et al. (forthcoming), who consider multiple outcomes of interest, but where agents can form only a single network which influences all of these.

If there is heterogeneity in the costs of forming links, these costs might be useful as instruments, if observed.[44] Without this we must take an alternative approach.

Rather than treating the endogeneity of the network as a problem, jointly modelling $\boldsymbol{G}$ and y uses the observed choices over links to provide additional information about the unobservables which enter the outcome equation. Rather than looking for a variable that can help explain the endogenous covariate but is excluded from the outcome, we now model an explicit economic relationship, and rely on the imposed model to provide identification. Such an approach is taken, for example, by Badev (2013), Blume et al. (2013), Hsieh and Lee (2014), and Goldsmith-Pinkham and Imbens (2013).

Typically the process is modelled as a two-stage game, where agents first form a network and then make outcome decisions.[45] Agents are foresighted enough to see the effect of their network decisions on their later outcome decisions. Consequently they solve the decision process by backward induction, first determining actions for each possible network, and then choosing network links with knowledge of what this implies for outcomes. For this approach to work one needs to be able to characterise the payoff of each possible network, so as to account for agents' network formation incentives in a tractable way.

There are two main limitations for this approach. First, by avoiding the use of exclusion restrictions, the role of functional form assumptions in providing identification becomes critical. Since theory rarely specifies precise functional forms, it is not unreasonable to worry about the robustness of results based on assumptions that are often due more to convenience than conviction.

Second, we typically need to impose limits on the form of the network formation model that mean the model is unable to generate many of the features of observed networks, such as the relatively high degree of clustering and low diameter. Particularly restrictive, and discussed further in Section 2.4, is the restriction that links are formed conditionally independently.

Changes in network structure

An alternative approach to those suggested above relies on *changes* in network structure to provide exogenous variation. In some circumstances one might believe that particular nodes or edges are removed from the network for exogenous reasons (this is sometimes described as 'node/edge failure'). For example, Patnam (2013) considers a network of interlocking company board memberships in India. A pair of firms is considered to be linked if the firms have a common board member. Occasionally edges between companies are severed due to the death of a board

[44] However, even this will depend on the timing of decisions. See Blume et al. (2013) for details on when such an argument might not hold.

[45] Of the papers mentioned above, Badev (2013) models the choice of friendships and actions simultaneously, whilst the others assume a two-stage process.

member, and to the extent that this is unpredictable, it provides plausibly exogenous variation in the network structure. One can then see how outcomes change as the network changes, and this gives a local estimate of the effect of the network on the outcome of interest. A similar idea is used by Waldinger (2010, 2012) using the Nazi expulsion of Jewish scientists to provide exogenous changes in academic department membership.

The difficulty with this approach in general is finding something that exogenously changes the network, but to which agents do not choose to respond.[46] Non-response includes both not adjusting edges in response to the changes that occur, and not *ex ante* choosing edges strategically to insure against the probabilistic exogenous edge destruction process. In the examples above these relate to not taking into account a board member's probability of death when hiring (*e.g.* not considering age when recruiting), and not hiring new scientists to replace those expelled.

2.4 Network Formation

Network formation is commonly defined as the process of edge formation between a fixed set of nodes. Although, in principle, one could also consider varying the nodes, in most applications the set of nodes will be well-defined and fixed. The empirical study and analysis of this process is important for three reasons.

First, the analysis in most of the previous section described how one might estimate social effects under the critical assumption that the networks of connections were themselves exogenous, or exogenous conditional on observed variables. In many circumstances, such as those described in Subsection 2.3.7, one might think that economic agents are able to make some choice over the connections they form, and that if their connections influence their outcomes they might be somewhat strategic in which edges they choose to form. In this case the social effects estimated earlier will be contaminated by correlations between an individual's observed covariates and the unobserved covariates of his friends. This is in addition to the problems of correlations in group-level unobservables that is well-known in the peer effects literature. For example, someone with a pre-disposition towards smoking is likely to choose to form friendships with others who might also enjoy smoking. An observed correlation in smoking decision, even once environmental characteristics are controlled for, might then come from the choice of friends, rather than any social influence. One solution to this problem, is to use a two-step procedure, in which a predicted network is estimated as a first stage. This predicted network is then used in place of the observed

[46]It is important to note that one also needs access to a panel of data for the network, which is not often available.

network in the second stage. This approach is taken by König et al. (2014).[47] Again the first stage will require estimation of a network formation process.

Second, an important issue when working with network data is that of measurement error. We return to this more fully in the next section, but where networks are incompletely observed, direct construction of network statistics using the sampled data typically introduces non-classical measurement error in these network statistics. If these statistics are used as covariates in models such as those in Section 2.3, we will obtain biased parameter estimates. One potential solution to this problem – proposed in different contexts by Goldberg and Roth (2003), Popescul and Ungar (2003), Hoff (2009), and Chandrasekhar and Lewis (2011) – is to use the available data and any knowledge of the sampling scheme to predict the missing data. This can be used to recover the (predicted) structure of the entire network, which can then be used for calculating any network covariates. Such procedures require estimation of network formation models on the available data.

Finally, we saw in Section 2.3 that social contacts can be important for a variety of outcomes, including education outcomes (Duflo et al., 2011; De Giorgi et al., 2010), risk-sharing (Ambrus et al., 2014; Angelucci et al., forthcoming; Jackson et al., 2012), and agricultural practices (Conley and Udry, 2010). Hence one might want to understand where social connections come from *per se* and how they can be influenced, in order to create more desirable outcomes. For example, there is substantial evidence of homophily (Currarini et al., 2010). Homophily might in some circumstances limit the benefits of connections, since there may be bigger potential gains from interaction by agents who are more different, *e.g. ceteris paribus* the benefits of mutual insurance are decreasing in the correlation of income. We might then want to consider what the barriers are to the creation of such links, and what interventions might support such potentially profitable edges.

The key challenge to dealing with network formation models is the size of the joint distribution for edges. For a directed binary network, this is a $N(N-1)$-dimensional simplex, which has $2^{N(N-1)}$ points of support (potential networks).[48] To give a sense of scale, for a network of more than 7 agents the support of this space is larger than the number of neurons in the human brain,[49] with 13 agents it is larger than the number of board configurations in chess,[50] and with 17 agents it is larger than the number of atoms in the observed universe.[51] Yet networks with so few agents are clearly much smaller than one would like to work with in practice. Hence simplifications will

[47] The same idea is used by Kelejian and Piras (2014) in the context of spatial regression.

[48] Through Section 2.4 we will be concerned with the identification and estimation of network formation models using data on a single network only. Throughout this section we therefore suppress the subscript g.

[49] Estimated to be around 8.5×10^{10} (Azevedo et al., 2009).

[50] Around $10^{46.25}$ (Chinchalkar, 1996).

[51] Around 10^{80} (Schutz, 2003).

typically need to be made to limit the complexity of the probability distribution defined on this space, in order to make work with these distributions computationally tractable.

We begin in Subsection 2.4.1 by considering methods which allow us to use data on a subset of observed nodes to predict the status of unsampled nodes. Here the focus is purely on in-sample prediction of link probabilities, not causal estimates of model parameters, so econometric concerns about endogeneity can be neglected. Such methods allow us to impute the missing network edges, providing one method for dealing with measurement error.

In Subsection 2.4.2, we then discuss conditions for estimating a network formation model, when the ultimate objective is controlling for network endogeneity in the estimation of a social effects model, as discussed in Subsection 2.3.7. Now we may have data on some or all of the edges of the network, and methods used for estimation will in many cases be similar to those for in-sample prediction. The key difference is that only exogenous predictors/covariates may be used. Additionally, in order to be useful as a first-stage for a social effects model, there must be at least one covariate which is a valid instrument *i.e.* it must have explanatory power for edge status, and not directly affect the outcome in the social effects model.

Next in Subsection 2.4.3, we consider economic models of network formation. Here we think about individual nodes as being economic agents, who make choices to maximise some objective *e.g.* students maximising their utility by choosing who to form friendships with. We first consider non-strategic models of formation, where the formation of one edge does not generate externalities, so that $\Pr(G_{ij} = 1|G_{kl}) = \Pr(G_{ij} = 1) \, \forall ij \neq kl$. Estimation of these models is relatively straightforward, and again relates closely to the discussion in the first two subsections.

Finally, we end with a discussion of more recent work on network formation, which has begun allowing for strategic interactions. Here the value to i of forming edges with j might depend on the status of other edges in the network. For example, when trying to gather information about jobs, individuals might find it more profitable to form edges with highly linked individuals who are more likely to obtain information, rather than those with few contacts. This dependence of edges on the status of other edges introduces important challenges, particularly when only a single cross-section of data are observed, as will typically be the case in applications. Since this work is at the frontier of research in network formation, we will focus on describing the assumptions and methods that have so far been used to estimate these models, without being able to provide any general guidance on how practitioners should use these methods.

2.4.1 In-sample prediction

Network formation models have long been studied in maths, computer science, statistical physics, and sociology. These models are characterised by a focus on the probability distribution $\Pr(\boldsymbol{G})$

as the direct object of interest.[52] For economists the main use for such models is likely to be for imputation/in-sample prediction when all nodes, and only a subset of edges in a network are observed.

The data available are typically a single realisation for a particular network, although occasionally multiple networks are observed and/or the network(s) is (are) observed over time. We focus on the case of one observation for a single network, since even when multiple networks are observed their total number is still small.[53] If multiple networks are available one could clearly at a minimum use the procedures described below, treating each separately, although one could also impose some restrictions on how parameters vary across networks if there is a good justification for doing so in a particular context. For example, suppose one observed edges between children in multiple classrooms in a school, with no cross-edges existing between children in different classes. If one believed that the parameters affecting edge formation were common across classrooms then one could improve the efficiency of estimation by combining the data. It could also provide additional identifying power, as network-level variables could also be incorporated into the model.

Identifying any non-trivial features of the probability distribution over the set of possible (directed) networks, $\Pr(\boldsymbol{G})$, is not possible from a single observation without making further restrictive assumptions. It is useful to note that $\Pr(\boldsymbol{G})$ is by definition equal to the joint distribution over all of the individual edges, $\Pr\left(G_{12}, ..., G_{N(N-1)}\right)$. Hence a single network containing N agents can be seen instead as $N(N-1)$, potentially dependent, observations of directed edge statuses.[54] This joint distribution can be decomposed into the product of a series of conditionals. For notational ease, let $l \in \Lambda$ index edges, so $\Lambda = \{12, 13, ..., 1N, 21, 23, ..., N(N-1)\}$. Then we can write $\Pr(\boldsymbol{G}) = \prod_{l \in \Lambda} \Pr(G_l | G_{l-1}, ..., G_1)$, so that each conditional distribution in the product is the distribution for a particular edge conditional on all previous edges. This conditioning encodes any dependencies which may exist between particular edges.

We begin with the simplest model of network formation, which assumes away both heterogeneity and dependence in edge propensities, and then reintroduce these features, describing the costs and benefits associated with doing so.

Independent edge formation

The *Bernoulli random graph* model is the simplest model of network formation. It imposes a common edge probability for each edge, and that probabilities are independent across edges.

[52]Economists, in contrast, are often interested in microfoundations, so the focus is typically instead on understanding the preferences, constraints, and/or beliefs of the agents involved in forming $\boldsymbol{G}$. We consider models of this form in Subsection 2.4.3.

[53]As noted in footnote 48, we therefore suppress the subscript g throughout this section to avoid unnecessarily cluttered notation.

[54]If the network is undirected there are only half that many edges.

Independence ensures that the joint distribution $\Pr\left(G_{12}, ..., G_{N(N-1)}\right)$ is just the product of the marginals, $\prod_{l \in \Lambda} \Pr(G_l)$. A common probability for each edge means that $\Pr(G_l) = p \; \forall \, l \in \Lambda$, so all information about the distribution $\Pr(\boldsymbol{G})$ is condensed into a single parameter, p, the probability an edge exists.[55] This can be straightforwardly estimated by maximum likelihood, with the resulting estimate of the edge probability $\hat{p} = \frac{|E|}{N(N-1)}$, equal to the proportion of potential edges that are present.[56]

A natural extension of this model allows the probability $\Pr(G_{ij} = 1)$ to depend on characteristics of the nodes involved, $(\boldsymbol{x}_i, \boldsymbol{x}_j)$, but conditional on these characteristics independence across edges is maintained. This type of model can be motivated either by pairs of individuals with particular characteristics $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ being more likely to meet each other and hence form edges, or by the benefits of forming an edge depending on these characteristics, or some combination of these. In general one cannot separate meeting probabilities from the utility of an edge without either parametric restrictions or an exclusion restriction, so additional assumptions will be needed if one wants to interpret the parameters structurally. We discuss this further in Subsection 2.4.3.

The key restriction here is the assumption of independence across edge decisions. In many cases this is unlikely to be reasonable. For example, in a model of directed network formation, there might well be correlation in edges G_{ij} and G_{il} driven by some unobservable node-specific fixed effect for node i *e.g.* i might be very friendly, so be relatively likely to form edges. Use of the estimated model to generate predicted networks will be problematic, as it will fail to generate some of the key features typically observed, such as the high degree of clustering.

Allowing for fixed effects

The simplest form of dependencies that one might want to allow for are individual-specific propensities to form edges with others, and to be linked to by others. Such models were developed by Holland and Leinhardt (1981) and are known as p_1-*models*. They parameterise the log probability an edge exists, $\log(p_{ij})$, as a linear index in a (network-specific) constant θ_0, a fixed effect for the edge 'sender' $\theta_{1,i}$, and a fixed effect for the edge 'receiver' $\theta_{2,j}$, so $\log(p_{ij}) = \theta_0 + \theta_{1,i} + \theta_{2,j}$. The fixed effects are interpreted as individual heterogeneity in propensity to make or receive edges. Additional restrictions $\sum_i \theta_{1,i} = \sum_j \theta_{2,j} = 0$ provide a normalisation that deals with the perfect collinearity that would otherwise be present.

The use of such fixed effects creates inferential problems, since increasing the size of the network also increases the number of parameters,[57] sometimes described as an *incidental pa-*

[55] Theoretical work on this type of model was done by Gilbert (1959), and it relates closely to the model of Erdős and Rényi (1959).

[56] Or twice that probability if edges are undirected, so that there are only $\frac{1}{2}N(N-1)$ potential edges.

[57] Every new node adds two new parameters to be estimated.

rameters problem. One natural solution to the latter problem is to impose homogeneity of the θ_1 and θ_2 parameters within certain groups, such as gender and race.[58] If there are C groups, then the number of parameters is now $2C+1$ and this remains fixed as N goes to infinity. This removes the inference problem and also allows agents' characteristics to be used in predicting edge formation.[59]

Alternatively, if node-specific effects are uncorrelated with node characteristics, then variations in edge formation propensity 'only' create a problem for inference. This comes from the unobserved node-specific effects inducing a correlation in the residuals, analogous to random effects. Fafchamps and Gubert (2007) show how clustering can be used to adjust standard errors appropriately.

However, in both cases the maintenance of the conditional independence assumption across edges continues to present a problem for the credibility of this method. In particular it rules out cases where the *status* of other edges, rather than just their probability of existence, affects the probability of a given edge being present. This would be inappropriate if for example i's decision on whether to form an edge with j depends on how many friends j actually has, not just on how friendly j is.

Allowing for more general dependencies

As discussed earlier in this section, identification of features of $\Pr(\boldsymbol{G})$ whilst allowing for completely general dependencies in edge probabilities is not possible. However, it is possible to allow the probability of an edge to depend on a subset of the network, where this subset is specified *ex ante* by the researcher. Such models are called p^*-models (Wasserman and Pattison, 1996) or *exponential random graph models* (ERGMs). These have already been used in economics by, for example, Mele (2013), who shows how such models can arise as the result of utility maximising decisions by individual agents, and Jackson et al. (2012) studying favour exchange among villagers in rural India.

Frank and Strauss (1986) showed how estimation could be performed in the absence of edge independence under the assumption that the structure of any dependence is known. For example, one might want to assume that edge ij depends not on all other edges, but only on the other edges that involve either i or j. This dependency structure, $\Pr_{\boldsymbol{\theta}}(G_{ij}|\boldsymbol{G}_{-ij}) = \Pr_{\boldsymbol{\theta}}(G_{ij}|G_{rs}\ \forall\ r \in \{i,j\}\text{or } s \in \{i,j\} \text{ but } rs \neq ij)$ where $\boldsymbol{\theta}$ is a vector of parameters and $\boldsymbol{G}_{-ij} = \boldsymbol{G}\backslash G_{ij}$, is called the *pairwise Markovian* structure.

[58] This is sometimes described as *block modelling*, since we allow the parameters, and hence edge probability, to vary across 'blocks'/groups.

[59] A related approach to solving this problem is suggested by Dzemski (2014).

Drawing from the spatial statistics literature, where this is a more natural assumption, Frank and Strauss show how an application of the Hammersley-Clifford theorem can be used to account for *any* arbitrary form of dependency.[60] The key result is that if the probability of the observed network is modelled as an exponential function of a linear index of network statistics, appropriately defined, any dependency can be allowed for.

To construct the appropriate network statistics, they first construct a *dependency graph,* g^{dep}. This graph contains $N(N-1)$ nodes, with each node here representing one of the $N(N-1)$ edges in the original graph.[61] Then an edge between a pair of nodes ij and rs in the dependency graph denotes that the conditional probability that edge ij exists is not independent of the status of edge rs *i.e.* $\Pr_{\boldsymbol{\theta}}(G_{ij} = 1|G_{rs}) \neq \Pr_{\boldsymbol{\theta}}(G_{ij} = 1)$. Further, conditional on the set of neighbours of node ij in the dependency graph, nei_{ij}^{dep}, $\Pr(G_{ij} = 1)$ is independent of all other edges in the original graph. So $\Pr_{\boldsymbol{\theta}}(G_{ij} = 1|\boldsymbol{G}_{-ij}) = \Pr_{\boldsymbol{\theta}}(G_{ij} = 1|G_{rs} \in nei_{ij}^{dep})$. For example, the p_1 graph, with independent edges, has a dependency graph containing no edges. By contrast, a 5-node graph with a pairwise Markovian dependency structure would have, for example, edge 12 dependent on edges $(13, 14, 15, 23, 24, 25, 31, 32, 41, 42, 51, 52)$, *i.e.* all edges which have one end at either 1 or 2.

We let $\mathcal{A}$ be the set of cliques of the dependency graph, where isolates are considered to be cliques of size one.[62] For example, if G_{ij} is independent of all other edges conditional on G_{ji} then $\mathcal{A} = \{(ij), (ij, ji)\}_{i \neq j}$.[63] Then we define A as representing the different architectures or *motifs* in $\mathcal{A}$. In the previous example these would be 'edges', (ij), and 'reciprocated edges' (ij, ji). This imposes a homogeneity assumption: that the probability a particular graph g is selected from $\mathcal{G}_N$ depends only on the number of edges and reciprocated edges, rather than to whom those edges belong, so all networks with the same overall architecture (called 'isomorphic networks') are equally likely.[64] If instead we allow dependence between any edges that share a common node, then $\mathcal{A}$ is the set of all edges (ij), reciprocated edges (ij, ji), triads (ij, ir, rj), and *k-stars* $(ij_1, ij_2, ..., ij_k)$. Now A represents 'edges', 'reciprocated edges', 'triads', and 'k-stars'.[65]

Invoking the Hammersley-Clifford theorem, Frank and Strauss (1986) note that the probability distribution over the set of graphs $\mathcal{G}_N$ allows for the imposed dependencies if it takes the

[60] Originally due to Hammersley and Clifford (1971) in an unpublished manuscript, and later proved independently by Grimmett (1973); Preston (1973); Sherman (1973); and Besag (1974).

[61] Nodes in this graph will be referred to by the name of the edge they represent in the original graph.

[62] A clique is any group of nodes such that every node in the group is connected to every other node in the group.

[63] (i, j) is always a member of $\mathcal{A}$, since we defined isolates as cliques of size one. Dependence of ij on ji means that we can also define (ij, ji) as a clique, since in the dependency graph these nodes are connected to each other.

[64] Formally, two networks are isomorphic iff we can move from one to the other only by permuting the node labels. For example, all six directed networks composed of three nodes and one edge are isomorphic. Isomorphism implies that all network statistics are also identical, since these statistics are measured at a network level so are not affected by node labels.

[65] This represents all triads in an undirected network, but in a directed network there are six possible edges between three nodes, since $ij \neq ji$, so we may define a number of different triads.

form

$$\Pr_{\boldsymbol{\theta}}(\boldsymbol{G}) = \frac{1}{\kappa(\boldsymbol{\theta})} \exp\left\{\sum_{A} \theta_A S_A(\boldsymbol{G})\right\} \tag{2.31}$$

where $S_A(\boldsymbol{G})$ is a summary statistic for motif A calculated from $\boldsymbol{G}$, θ_A is the parameter associated with that statistic, and $\kappa(\boldsymbol{\theta})$ is a normalising constant, sometimes described as the *partition function*, such that $\sum_{\boldsymbol{G}\in\mathcal{G}_N} \Pr_{\boldsymbol{\theta}}(\boldsymbol{G}) = 1$.[66] In particular, $S_A(\boldsymbol{G})$ must be a positive function of the number of occurrences of motif A in $\boldsymbol{G}$. Since we are working with binary edges, without loss of generality we can define $S_A(\boldsymbol{G})$ as simply a count of the number of occurrences of motif A in the graph represented by $\boldsymbol{G}$. For example, defining $\mathbf{S}(\boldsymbol{G})$ as the vector containing the $S_A(\boldsymbol{G})$, if $\mathcal{A} = \{(ij), (ij, ji)\}_{i\neq j}$ then $\mathbf{S}(\boldsymbol{G})$ is a 2×1 vector containing a count of the number of edges and a count of the number of reciprocated edges.

Estimation of the ERGM model is made difficult by the presence of the partition function, $\kappa(\boldsymbol{\theta})$. Since this function normalises the probability of each graph so that the probabilities across all potential graphs sum to unity, it is calculated as $\sum_{\boldsymbol{G}\in\mathcal{G}_N} \exp\{\sum_A \theta_A S_A(\boldsymbol{G})\}$. The outer summation is a sum over the $2^{N(N-1)}$ possible graphs. As noted earlier, even for moderate N this is a large number, so computing the sum analytically is rarely possible.

Three approaches to estimation have been taken to overcome this difficulty: (1) the *coding method*; (2) the *pseudolikelihood* approach; and (3) the *Markov Chain Monte Carlo* approach. The first two are based on the maximising the conditional likelihoods of edges, rather than the joint likelihood, thus obviating the need for calculating the normalising constant, whilst the third instead calculates an approximation to this constant.

Coding Method The coding method (Besag, 1974) writes the joint distribution of the edge probabilities as the product of conditional distributions

$\Pr_{\boldsymbol{\theta}}(\boldsymbol{G}) = \prod_{l\in\Lambda} \Pr_{\boldsymbol{\theta}}(G_l|G_{l-1}, ..., G_1)$, where as before Λ is the set of all $N(N-1)$ potential edges. Under the assumption that edge G_l depends only on a subset of other edges $G_{l'} \in nei_l^{dep}$ one could 'colour' each edge, such that each edge depends only on edges of a different colour.[67] All edges of the original graph that have the same colour are therefore independent of each other by construction. Let Λ_c be the set of all edges of a particular colour. One could then estimate the parameter vector of interest, $\boldsymbol{\theta}$, by maximum likelihood, using only $\Pr_{\boldsymbol{\theta}}(G_l|G_{l'} \in nei_l^{dep})\, \forall l \in \Lambda_c$, which treats only edges of the same colour as containing any independent information.

We define the 'change statistic' $D_A(\boldsymbol{G}; l) := S_A(G_l = 1, \boldsymbol{G}_{-l}) - S_A(G_l = 0, \boldsymbol{G}_{-l})$ as the change in statistic S_A from edge G_l being present, compared with it not being present, given all

[66] In a slight abuse of notation we write $\sum_{\boldsymbol{G}\in\mathcal{G}_N} \Pr_{\boldsymbol{\theta}}(\boldsymbol{G})$ to mean $\sum_{g\in\mathcal{G}_N} \Pr_{\boldsymbol{\theta}}(\boldsymbol{G}_g)$.

[67] This is equivalent to saying that no two adjacent (*i.e.* linked) nodes of the dependency graph should have the same colour. Note that this colouring will not be unique. For example, one could trivially always colour every edge a different colour. However, for estimation it is optimal to try to minimise the number of colours used, as this makes the most of any information available about independence.

the other edges $\boldsymbol{G}_{-l}$. Then, given the log-linear functional form assumption that we have made (see Equation 2.31), the conditional probability of an edge l can be estimated from the logit regression $\log\left\{\frac{\Pr(G_l=1|\boldsymbol{G}_{-l})}{\Pr(G_l=0|\boldsymbol{G}_{-l})}\right\} = \sum_A \theta_A D_A(\boldsymbol{G};\, l)$. This can be implemented in most standard statistical packages. Hence we can estimate $\boldsymbol{\theta}$ using maximum likelihood under the assumption that the edge probability takes a logit form and treating the edges $l \in \Lambda_c$ as independent, conditional on the edges not in Λ_c. Since all the conditioning edges which go into S_A are of different colours, they are not included in the maximisation, so $\hat{\boldsymbol{\theta}}_c$ will be consistent.

By performing this maximisation separately for each colour, a number of different estimates can be recovered. Researchers may choose to then report the range of estimates produced, or to create a single estimate from these many results, for example taking a mean or median.

The main disadvantage of this approach is that the resulting estimates will each be inefficient, since they treat the edges $l \notin \Lambda_c$ as if they contain no information about the parameters. In practice the proportion of edges in even the largest colour set Λ_c is likely to be small. For example, if any edges that share a node are allowed to be dependent, then the number of independent observations will only be $\frac{1}{2}N$.[68] Hence efficiency is far from a purely theoretical concern in this environment.

Pseudolikelihood approach The pseudolikelihood approach, introduced to the social networks literature by Strauss and Ikeda (1990), attempts to overcome the inefficiency problem, by finding $\boldsymbol{\theta}$ which jointly maximises *all* the conditional distributions, not just those of the same colour. We write the log likelihood based on edges of colour c as $L_c = \sum_{l\in\Lambda_c} \log \Pr_{\boldsymbol{\theta}}(G_l = 1|G_{l'} \in nei_l^{dep})$, with $\hat{\boldsymbol{\theta}}_c$ as the maximiser of this. Besag (1975) notes that the log (pseudo)likelihood $PL = \sum_c L_c = \sum_c \sum_{l\in\Lambda_c} \log \Pr_{\boldsymbol{\theta}}(G_l = 1|G_{l'} \in nei_l^{dep})$, constructed by simply combining all the data as if there were no dependencies, is equivalent to a particular weighting of the individual, 'coloured' log likelihoods. This likelihood is misspecified, since the correct log likelihood using all the data should be $L = \sum_l \log \Pr_{\boldsymbol{\theta}}(G_l = 1|G_{l-1}, ..., G_1)$, whilst here we have instead $L = \sum_l \log \Pr_{\boldsymbol{\theta}}(G_l = 1|\boldsymbol{G}_{-l}) = \sum_l \log \Pr_{\boldsymbol{\theta}}(G_l = 1|G_L, ...G_{l+1}, G_{l-1}, ..., G_1)$.[69] Nevertheless, under a particular form of asymptotics it may still yield consistent estimates.

We have already noted that for any given colour, the standard maximum likelihood consistency result applies, as the observations included are independent. If the number of colours are held fixed as the number of potential edges is increased,[70] then under some basic regularity

[68] Or $\frac{1}{2}(N-1)$ if N is odd.

[69] A likelihood based on $\Pr_{\boldsymbol{\theta}}(G_l|\boldsymbol{G}_{-l})$ without any correction suffers from simultaneity, since the probability of each edge is being estimated conditional on all others remaining unchanged. In a two node directed network, as a simple example, we effectively have two simultaneous equations, one for $\Pr_{\boldsymbol{\theta}}(G_{12}|G_{21})$ and $\Pr_{\boldsymbol{\theta}}(G_{21}|G_{12})$. It is well-known that such systems will not generally yield consistent parameter estimates if the dependence between the equations is not considered, and that strong restrictions will typically be needed even to achieve identification.

[70] In the language of spatial statistics, this is described as 'domain increasing asymptotics'.

conditions (Besag, 1975), maximising the log pseudolikelihood function $PL(\boldsymbol{\theta})$ as though there were no dependencies will also give a consistent estimate of $\boldsymbol{\theta}$.

Unfortunately, in practice this approach suffers from a number of problems. First, although it makes use of more information in the data, so is potentially more efficient, the standard errors that are produced by standard statistical packages such as Stata will clearly be incorrect as they will not take into account the dependence in the data. Little is known about how to provide correct standard errors, but in some cases inference can proceed using an alternative, non-parametric procedure: *multiple regression quadratic assignment procedure* (MRQAP). This method can provide a test as to whether particular edge characteristics or features of the local network, such as a common friend, are important for predicting the probability that a pair of individuals is linked. It is based on the quadratic assignment procedure (QAP): a type of permutation test for correlation between variables. For more details see Subsection 2.7.2.

A second issue is that in network applications we need to impose some structure on the way in which new nodes are added to the network when we do asymptotics (Boucher and Mourifié, 2013; Goldsmith-Pinkham and Imbens, 2013). If, as we increase the sample size, new nodes added could be linked to all the existing nodes, then there is no reduction in dependence between links. In the spatial context for which the theory was developed, the key idea is that increasing sample size creates new geographic locations that are added at the 'edge' of the data. If correlations reduce with distance, then as new, further away, locations are added, they will be essentially independent from most existing locations. Such asymptotics are called *domain-increasing* asymptotics. The analogy in a networks context, proposed by Boucher and Mourifié (2013) and Goldsmith-Pinkham and Imbens (2013), is that new nodes are further away in the support of the covariates. If there is homophily, so that nodes which are far apart in covariates never link, then the decisions of these nodes are almost independent. Asymptotics results from the spatial case can then be used.

Third, Kolaczyk (2009) suggests that in practice this method only works well when the extent of dependence in the data is small. In general there is no reason to assume dependence will be small in network data; indeed it is precisely because we did not wish to assume this that we considered ERGMs at all.

Markov Chain Monte Carlo Maximum Likelihood An alternative approach, not based on the *ad-hoc* weighting provided by the pseudolikelihood approach, is to use Markov Chain Monte Carlo (MCMC) maximum likelihood (Geyer and Thompson, 1992; Snijders, 2002; Handcock, 2003). As noted earlier, the key difficulty with direct maximum likelihood estimation of Equation 2.31 is the presence of the partition function $\kappa(\boldsymbol{\theta}) = \sum_{\boldsymbol{G}\in\mathcal{G}} \exp\left\{\sum_A \theta_A S_A(\boldsymbol{G})\right\}$. This normalising constant is an intractable function of the parameter vector $\boldsymbol{\theta}$. In this estimation approach, MCMC techniques can be used to create an estimate of $\kappa(\boldsymbol{\theta})$ based on a sample of graphs drawn from $\mathcal{G}_N$.

The original log likelihood can be written as $L(\boldsymbol{\theta}) = \sum_A \theta_A S_A(\boldsymbol{G}) - \kappa(\boldsymbol{\theta})$. Maximising this is equivalent to maximising the likelihood ratio $LR = L(\boldsymbol{\theta}) - L(\boldsymbol{\theta}^{(0)})$ since the latter is just a constant for some arbitrary initial $\boldsymbol{\theta}^{(0)}$. Writing this out in full we get $LR = \sum_A \left[\theta_A - \theta_A^{(0)}\right] S_A(\boldsymbol{G}) - \left[\kappa(\boldsymbol{\theta}) - \kappa(\boldsymbol{\theta}^{(0)})\right]$. The second component can be approximated by drawing a sequence of W graphs, $(\boldsymbol{G}_1, ..., \boldsymbol{G}_W)$, from the ERGM under $\boldsymbol{\theta}^{(0)}$, and computing $\log \sum_{w \in W} \exp \left\{ \sum_A (\theta_A - \theta_A^{(0)}) S_A(\boldsymbol{G}^{(w)}) \right\}$ (see Kolaczyk (2009) pp185-187 for details). Under this procedure the maximiser of the approximated log likelihood will converge to its true value $\boldsymbol{\theta}$ as the number of sampled graphs W goes to infinity.

This approach has two major disadvantages. The first is that implementation of this method is very computationally intensive. Second, although this approach avoids the approximation of the likelihood by directly evaluating the normalising constant, its effectiveness depends significantly on the quality of the estimate of $\left[\kappa(\boldsymbol{\theta}) - \kappa(\boldsymbol{\theta}^{(0)})\right]$. If this cannot be approximated well then it is not clear that this approach, although more principled, should be preferred in practical applications.

Recent work by Bhamidi et al. (2008) and Chatterjee et al. (2010) suggests that in practice the mixing time – time taken for the Markov chain to reach its steady state distribution – of such MCMC processes is very slow (exponential time). This means that as the space of possible networks grows, the number of replications in the MCMC process that must be performed in order to achieve a reasonable approximation to $\left[\kappa(\boldsymbol{\theta}) - \kappa(\boldsymbol{\theta}^{(0)})\right]$ rises rapidly, making this approach difficult to justify in practice.

Statistical ERGMs Chandrasekhar and Jackson (2014) also note that practitioners often report obtaining wildly different estimates from repeated uses of ERGM techniques on the same set of data with the same model, with variation far exceeding that expected given the claimed standard errors. They propose a technique which they call *Statistical ERGM* (SERGM), which is easier to estimate, as an alternative to the usual ERGM. With this they are not able to recover the probability that we observe a particular network, but instead focus on the probability of observing a given realisation, $\boldsymbol{s}$, of the network statistics, $\boldsymbol{S}$.[71]

In an ERGM the sample space consists of the set of possible distinct networks on the N nodes. This set has $2^{N(N-1)}$ elements (in the case of a directed network), and we treat each isomorphic element as being equally likely. Our *reference distribution* is a uniform distribution

[71] $\boldsymbol{S}$ is a $|\mathcal{A}| \times 1$ dimensional vector stacking the network statistics S_A, and $\boldsymbol{\theta}$ a $1 \times |\mathcal{A}|$ dimensional vector of parameters.

across these $2^{N(N-1)}$ elements *i.e.* this is the null distribution against which we are comparing the observed network.

If our interest is only in the realisations of the network statistics, we can reduce the size of the sample space we are working with. Chandrasekhar and Jackson (2014) define SERGMs as ERGMs on the space of possible network statistics, $\mathcal{S}$. This sample space will typically contain vastly fewer elements than the space of possible networks.

We can then rewrite Equation 2.31 using the space of network statistics as sample space. In this case the probability of observing statistics $\boldsymbol{S}(\boldsymbol{G})$ taking value $\boldsymbol{s}$ is $\Pr_{\boldsymbol{\theta}}(\boldsymbol{S}(\boldsymbol{G}) = \boldsymbol{s}) = \frac{\#_{\boldsymbol{S}}(\boldsymbol{s})\exp(\boldsymbol{\theta s})}{\sum_{\boldsymbol{s}'} \#_{\boldsymbol{S}}(\boldsymbol{s}')\exp(\boldsymbol{\theta s}')}$, where $\#_{\boldsymbol{S}}(\boldsymbol{s}) = |\{\boldsymbol{G} \in \mathcal{G} : \boldsymbol{S}(\boldsymbol{G}) = \boldsymbol{s}\}|$ is the number of potential networks which have $\boldsymbol{S} = \boldsymbol{s}$.

So far we have only rewritten our originally ERGM by defining it over a new space. We defined our reference distribution in the ERGM to put equal weight on each possible *network*. To maintain this distribution when the sample space is the space of statistics, we must weight the usual (unnormalised) probability of observing network $\boldsymbol{G}$, $\exp(\boldsymbol{\theta s})$, by the number of networks which exhibit this configuration of statistics, $\#_{\boldsymbol{S}}(\boldsymbol{s}')$.

Much of the difficulty in estimating ERGM models comes from use of these weights, since we are required to know in how many networks a particular combination of statistics exists. Since this is typically not possible to calculate analytically, we discussed how MCMC approaches might be used to sample from the distribution of networks.

Chandrasekhar and Jackson (2014) complete their definition of SERGMs as a generalisation of ERGMs by allowing any reference distribution, $K_{\boldsymbol{S}}(\boldsymbol{s})$ to be used in the place of $\#_{\boldsymbol{S}}(\boldsymbol{s}')$. However, to ease estimation relative to ERGMs, they then define the 'count SERGM', which imposes $K_{\boldsymbol{S}}(\boldsymbol{s}) = \frac{1}{|\mathcal{S}|}$.[72] The key here is not that these weights are constant, but that they no longer depend on the space of networks. Since $K_{\boldsymbol{S}}(\boldsymbol{s})$ is now known, unlike $\#_{\boldsymbol{S}}(\boldsymbol{s}')$ which needed to be calculated, if $|\mathcal{S}|$ is sufficiently small, exact evaluation of the partition function $\tilde{\kappa}(\boldsymbol{\theta}) = \sum_{\boldsymbol{s}'} K_{\boldsymbol{S}}(\boldsymbol{s}') \exp\{\boldsymbol{\theta s}'\}$ is now possible.

Since count SERGMs – and any other SERGMs with known $K_{\boldsymbol{S}}(\boldsymbol{s}')$ – can be estimated directly and without approximation, they are easier to implement than standard ERGMs. Chandrasekhar and Jackson (2014) also provide assumptions under which the parameters of the SERGM, $\boldsymbol{\theta}_{SERGM}$, can be estimated consistently.

The key drawback to this method is in interpretation. The estimated parameters, $\boldsymbol{\theta}_{SERGM}$, are not the same as the parameters $\boldsymbol{\theta}$ in Equation 2.31, and the predicted probabilities are now the probability of a particular configuration of statistics, rather than of a particular network. Nevertheless, for a researcher interested in which network motifs are more likely to be observed

[72] Count SERGMs also restrict the set $\mathcal{A}$ to include only network motifs such as triangles and nodes of particular degree, which can be counted. This rules out, for example, statistics such as density.

than one would expect under independent edge formation, SERGMs offer an appropriate alternative.

2.4.2 Reduced form models of network formation

The methods discussed in the previous subsection focused on in-sample prediction of network edges. However, since they (mostly) predict these probabilities based on the structure of the networks, without use of other characteristics, they both fail to make use of all the information typically available to researchers, and also do not contain the necessary independent variation needed for use as the first stage of a social effects model with an endogenous network (of the sort discussed in Subsection 2.3.7). When our ultimate aim is to estimate a social effects model but we are concerned about the network being endogenous, one solution discussed in Subsection 2.3.7 is to estimate the edge probability using individual characteristics, including at least one covariate that is not included in the outcome equation (an exclusion restriction), as in a standard two-stage least squares setting. In this subsection we describe estimation of models that include individual (node) characteristics. As long as at least one of these is a valid instrument, then this approach to overcoming the endogeneity of network formation is possible.

A well-recognised feature of many kinds of interaction networks is the prevalence of homophily: a propensity to be linked to relatively similar individuals.[73] This observation may arise from a preference for interacting with agents who are similar to you (preference homophily), a lower cost of interacting with such agents (cost homophily), or a higher probability of meeting such agents (meeting homophily). However, they all have the reduced form implication that more similar agents are more likely to be linked.[74]

Fafchamps and Gubert (2007) provide a discussion of the conditions that must be fulfilled by a model used for *dyadic regression*, *i.e.* a regression model of edge formation when edges are being treated as observations and node characteristics are included in the regressors. They note the regressors must enter the model symmetrically, so that the effect of individual characteristics ($\boldsymbol{x}_i$, $\boldsymbol{x}_j$) on edge G_{ij} is the same as that of ($\boldsymbol{x}_j$, $\boldsymbol{x}_i$) on G_{ji}. Additionally the model may contain some edge-specific covariates, such as the distance between agents, which must by definition be symmetric $\boldsymbol{w}_{ij} = \boldsymbol{w}_{ji}$. If edges are modelled as directed, then the model takes the general form

$$G_{ij} = f\left(\lambda_0 + (\boldsymbol{x}_{1i} - \boldsymbol{x}_{1j})\boldsymbol{\lambda}_1 + \boldsymbol{x}_{2i}\boldsymbol{\lambda}_2 + \boldsymbol{x}_{3j}\boldsymbol{\lambda}_3 + \boldsymbol{w}_{ij}\boldsymbol{\lambda}_4 + u_{ij}\right) \tag{2.32}$$

[73] Homophily may be casually described as the tendency of ‘birds of a feather to flock together’.

[74] In Subsection 2.4.3 below, we consider homophily in more detail, and structural models that try to separate these causes of observed homophily.

This specification allows a term that varies with the difference between i and j in some characteristics, $(\boldsymbol{x}_{1i} - \boldsymbol{x}_{1j})$; terms varying in the characteristics of both the sender and the receiver of the edge, $\boldsymbol{x}_{2i}$ and $\boldsymbol{x}_{3j}$ respectively; some edge-specific characteristics, $\boldsymbol{w}_{ij}$; and an edge-specific unobservable, u_{ij}. There may be partial or even complete overlap between any of $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, and $\boldsymbol{x}_3$. Since G_{ij} is typically binary, the function $f(.)$ and the distribution of u are usually chosen to make the equation amenable to probit or logit estimation. However, in some cases other functional forms are chosen. For example, Marmaros and Sacerdote (2006) model $f(.)$ as exp(.) since they are working with email data, measuring edges by the number of emails between the individuals, which takes only non-negative values and varies (almost) continuously.

If edges are undirected, then $(\boldsymbol{x}_{1i} - \boldsymbol{x}_{1j})$ must be replaced with $|\boldsymbol{x}_{1i} - \boldsymbol{x}_{1j}|$;[75] $\boldsymbol{x}_2 = \boldsymbol{x}_3$ and $\boldsymbol{\lambda}_2 = \boldsymbol{\lambda}_3$; and $u_{ij} = u_{ji}$, so that G_{ij} necessarily equals G_{ji}. The identification of parameters $\boldsymbol{\lambda}_2$ and $\boldsymbol{\lambda}_3$ requires variation in degree. As Fafchamps and Gubert (2007) note, if all individuals in the data have the same number of edges, such as a dataset of only married couples, then it is possible to ask whether people are more likely to form edges with people of the same race, captured by $\boldsymbol{\lambda}_1$, but not possible to ask whether some races are more likely to have edges.

Careful attention needs to be paid to inference in this model, since there is dependence across multiple dyads for any individual, similar to the Markov random graph assumption discussed in the previous subsection. Fafchamps and Gubert (2007) show that standard errors can be constructed analytically using a 'four-way error components model'. This is a type of clustering, allowing for correlation between u_{ij} and u_{rs} if either of i or j is equal to either of r and s. The analytic correction they propose provides an alternative to using MRQAP, described in Subsection 2.4.1, which may also be used in this circumstance.

2.4.3 Structural models of network formation

Economic models of network formation consider nodes as motivated agents, endowed with preferences, constraints, and beliefs, choosing which edges to form. The focus for applied researchers is to estimate parameters of the agents' objective functions. For example, to understand what factors are important for students in deciding which other students to form friendships with.

These models allow us to think about counterfactual policy scenarios. For example, if friendships affect academic outcomes, then there might be a role for policy in considering how best to organise students into classrooms, given knowledge of their endogenous friendship formation response. If students tend to form homophilous friendships *i.e.* with others who have similar predetermined characteristics, but not to form friendships across classrooms, there may be a case for not streaming students into classes of similar academic abilities. This would create

[75] Or $(\boldsymbol{x}_{1i} - \boldsymbol{x}_{1j})^2$ may also be used.

more heterogeneity in the characteristics of friends than if streaming were used, which might improve the amount of peer learning that takes place.[76] We begin by discussing non-strategic models, in which these decisions depend only on the characteristics of the agents involved in the edge. We then discuss strategic network formation, which occurs when network features directly enter into the costs or benefits of forming particular edges.[77]

Structural Homophily

As noted above, a key empirical regularity which holds across a range of network types is the presence of *homophily.* This is related to the more familiar (in economics) concept of positive assortative matching, *i.e.* that people with similar characteristics form edges with one another. As we have already seen, many reduced form models include homophilic terms – captured by $\boldsymbol{\lambda}_1$ in Equation 2.32 – to allow the probability a tie exists to vary with similarity on various node characteristics.[78] In this subsection, we consider the economic models of network formation that are based on homophily.

We define homophily formally as follows. Let the individuals in a particular environment be members of one of H groups, with typical group h. Groups might be defined according to sex, race, height, or any other characteristics. Continuous characteristics will typically need to be discretised. We denote individual i's membership of group h as $i \in h$. Relationships for individuals in group h exhibit homophily if $\Pr(G_{ij} = 1|i \in h, j \in h) > \Pr(G_{ij} = 1|i \in h, j \notin h)$. In words, a group h exhibits homophily if its members are more likely to form edges with other members of the same group than one would expect if edges were formed uniformly at random among the population of nodes. In general there will be multiple characteristics $\{H^1, ..., H^K\}$ according to which individuals can be classified, and relationships may exhibit homophily on any number of these characteristics.

As noted earlier there are (at least) three possible sources of homophily: *preference homophily*, *cost homophily*, and *meeting homophily.*

Preference homophily implies that, conditional on meeting, people in a group are more likely to form edges with other members of the same group as they value these edges more. For example, within a classroom boys and girls might have equal opportunities to interact, but boys may choose to form more friendships with other boys (and *mutatis mutandis* for girls) if they have more similar interests.

[76] Clearly this is just an example, and there are many other factors to consider, such as the effectiveness of teachers when faced with more heterogeneous classrooms, the ability to tailor lessons to challenge high ability students, and other outcomes that might be influenced by changing friendships.

[77] See also a recent survey by Graham (2015), which became available after work on this manuscript.

[78] In principle this probability could be falling in similarity, known as *heterophily.* This may be relevant, for example in models of risk sharing with heterogeneous risk preferences and complete commitment.

Cost homophily occurs when the cost of maintaining an edge to a dissimilar agent is greater than the cost of maintaining an edge to a more similar agent. For example, one might have an equal preference for all potential friends, but find it 'cheaper' to maintain a friendship with individuals who live relatively nearer. Unlike preferences, which are in some sense fundamental to the individual, costs might be manipulable by policy. To the extent that they are environmental these can also change the value of an edge over time, *e.g.* a friend moving further away may lead to the friendship being broken.

Meeting homophily occurs when people of a particular group are more likely to meet other members of the same group. For example, if we thought of all students in a school year as being part of a single network, then there is likely to be meeting homophily within class groups, since students in the same class have more opportunities to interact. Again this is amenable to manipulation by policy, for example changing seating arrangements across desks in a classroom. However, unlike cost homophily, once individuals have met, changes in the environment should not change the value of a friendship.

These three sources of homophily all have the reduced form implication that the coefficient on the *absolute* difference in characteristics, $\boldsymbol{\lambda}_1$ in Equation 2.32, should be negative for any characteristics on which individuals exhibit homophily. However, since they may have different policy implications, there is a case for trying to distinguish which of these channels are operating to cause the observed homophily.

Currarini et al. (2009) suggest how one can distinguish between preference and meeting homophily under the assumption that cost homophily does not exist. They note that if group size varies across groups, then preference homophily should lead to more friendships among the larger group, whereas meeting homophily should not. Intuitively this is because under preference homophily, a larger own-group means there are more people with whom one might potentially form a profitable friendship. One could then use regression analysis to test for the presence of preference homophily by interacting group size with absolute difference in characteristics, and testing whether the estimated parameter is significantly different from zero.

Alternatively one might want to estimate the magnitude of the effect of changing particular features of the environment, such as the classrooms to which individuals are assigned. In this case one could parameterise an economic model of behaviour, and then directly estimate the parameters of the model. Currarini et al. (2009) do this using a model of network formation that incorporates a biased meeting process, so individuals can meet their own-type more frequently than other types, and differences in the value of a friendship depending on whether agents are the same type.[79] They simulate the model with a number of different parameters for meeting

[79] Again they do not allow for cost homophily.

probabilities and relative values of friendships, and use a minimum distance procedure to choose the parameters that best explain the data.

As ever with structural models, whilst this approach allows one to perform counterfactual policy experiments, the main cost is that the reasonableness and interpretation of results depend on the accuracy with which the imposed model fits reality. Also, without time series variation in friendships, one cannot also allow for cost heterogeneity, which might show up either in preferences by changing the value of forming an edge, or in meeting probabilities since those with lower meeting probabilities will typically have a greater cost to maintaining a friendship. Finally, it is important to note that estimation of such models requires the unobserved component of preferences to be independent of the factors influencing meeting. If the unobserved preference for partying is correlated with choosing to live in a particular dormitory, and hence meeting other people living here, then this will bias the parameter estimate of the probability of meeting in this environment.

Mayer and Puller (2008) develop an enriched version of this model which allows again for meeting and preference homophily, but they allow the bias in the meeting process to depend not only on exogenous characteristics, but also on sharing a mutual friend. Formally, $\Pr(meet_{ij} = 1|G_{ir} = G_{jr} = 1) > \Pr(meet_{ij} = 1)$, where $\Pr(meet_{ij})$ denotes the probability that nodes i and j meet (and hence have the opportunity to form an edge). This allows for the stylised fact that individuals who are friends often also share mutual friends, which helps the model match the observed clustering in the data.

However, although the model fit is improved, their model cannot distinguish whether this clustering is in fact generated by a greater probability of meeting such individuals, a greater benefit to being friends with someone you share a friend with already, or a lower cost of maintaining that friendship. They show how one can estimate their model using a simulated method of moments procedure. However, this method suffers from the same constraints as those in the model suggested by Currarini et al. (2009): the utility of the model for counterfactuals depends on how closely it matches reality; cost homophily is neglected; and it is important the unobserved component of preferences is independent of the meeting process.

In the next subsection we consider extensions to these models that allow network statistics, such as sharing a common friend, to enter into individuals' utility functions. These create strategic interactions which can complicate estimation.

Strategic network formation

Much of the theoretical literature on networks has emphasised the strategic nature of interactions, setting up games of network formation as well as games to be played on existing networks (as seen in Section 2.3 above). The empirical literature has recently begun to take a similar approach, trying to estimate games of network formation. The key extension of such models, beyond those already considered, is to include network covariates into the objective function of agents. This creates two complications: first such models may have zero, one, or many equilibria, and this must be accounted for in estimation; and second, as with ERGM models, the presence of network covariates necessitates the calculation of intractable functions of the unknown parameters.

Before considering estimation in more detail, we discuss the modelling choices that one needs to make. First, as with all structural modelling one must explicitly determine the nature of the objective function that agents are trying to maximise. For example one might have individuals with utility functions that depend on some feature of the network, who are trying to maximise this utility.[80] Second, the 'rules of the game': are decisions made simultaneously or sequentially? Unilaterally or bilaterally? What do agents know, and how do they form beliefs? Given that we typically only observe a single cross-section of data, additional assumptions about the nature of any meeting process are necessary. Similarly, data may be reported as directed or undirected, but whether we treat unreciprocated directed edges as measurement error or evidence of unilateral linking is an important consideration, particularly given the consequences of such measurement error (see Subsection 2.5.3). Finally, one needs to take a stand on the appropriate concept of equilibrium and the strategies being played. At the weakest, one could impose only that strategies must be rationalisable, and hence many strategy profiles are likely to be equilibria. On the other hand, depending on the information available to agents one could impose Nash equilibrium, or Bayes-Nash equilibrium where individuals have incomplete information and need to form beliefs. Alternatively one could use a partly cooperative notion of equilibrium such as pairwise stability (Jackson and Wolinsky, 1996), which models link formation as requiring agreement from both parties involved, although dissolution remains one-sided.[81]

Since these models are at the frontier of research on network formation, few general results are currently available. We therefore instead briefly discuss the approaches that have been taken so far to write estimable models, and estimate the parameters of these models. Our aim is to highlight some of the choices that need to be made, and their relative advantages and costs.

Christakis et al. (2010) and Mele (2013) both model network formation as a sequential game: there is some initial network, and then a sequential process by which edge statuses may be adjusted. Crucial, also, to their models, is that at each meeting agents only weigh the static benefits of updating the edge status (*i.e.* play a *myopic best response*), rather than taking

[80] For example their utility may depend on their centrality, or the number of edges they have subject to some cost of forming edges. It is important to note that although it is the *realised* network feature that typically enters an agent's objective function, their strategy will depend on their *beliefs* about how others will act.

[81] As in the literature on coalition formation, the issue of whether utility is transferable or not is also critical. Typically this issue is not discussed in networks papers (Sheng (2012) is an exception to this), and it is implicitly assumed that utility is not transferable.

into account the effect this decision will have on both their own and others' future decisions. Allowing for such forward-looking behaviour has so far proved insolvable from an economic theory perspective, and hence they rule this out.

Christakis et al. (2010) assume the initial network is empty, and allow each pair to meet precisely once, uniformly at random, in some unknown order. Mele (2013) also allows uniform at random meeting, but pairs may meet many times until no individual wants to change any edge. In both cases these assumptions about the meeting process – the number of meetings, order in which pairs meet, and probability with which each pair meets – will influence the set of possible networks that may result. However, in the latter case, the resulting network will be an equilibrium network, something which is not true in Christakis et al. (2010).

A different approach, taken by Sheng (2012), avoids making assumptions about the meeting order. Instead she uses only an assumption about the relevant equilibrium concept (pairwise stability). For the network to be pairwise stable, the utility an agent gets from each link that is present must be greater than the utility he would get if the link were not present, and conversely for a link which is not present at least one of the agents it would involve must not prefer it. Sheng uses the moment inequalities this implies for estimation, but is only able to find bounds on the probability of observing particular networks.[82] Hence assumptions about meeting order seem important for the point identification of the parameter of interest (we discuss this further below).

De Paula et al. (2014) also avoid assumptions on the meeting order. Rather than using individual-level data, they identify utility parameters by aggregating individuals into 'types', and looking at the share of each type that is observed in equilibrium. This can be seen as an extension of the work of Currarini et al. (2009). Individuals' characteristics are discretised, so that each individual can be defined as a single type. Agent characteristics might, for example, be sex and age. Typically age is measured to the nearest month or year, so is already discretised. However, if the number of elements in the support is large, broader discretisation might be desirable (*e.g.* in the age example, measure age in ten-year bands). Then we might define one type as (male, 25-35years) and another as (female, 15-25). De Paula et al. (2014) assume that agents have preferences only over the types they connect to both directly and indirectly, not who the individuals are, and that preference shocks are also defined in terms of type rather than individuals. They further assume that there is some maximum distance such that there is no value to a having connections beyond this distance, and there is a maximum number of direct connections that would be desired. Under these restrictions they can set identify the set of

[82]Sheng (2012) is actually only able to estimate an 'outer region' in which these probabilities lie, rather than a sharp set. More information is, in principle, available in the data, but making use of it would increase the computational burden.

parameters for which the observed outcome – distribution of network types – is an equilibrium, without making any assumptions on equilibrium selection. They are even able to allow for non-existence of equilibrium, in which case the identified set is empty. Estimation can be performed using a quadratic program.

Recent work by Leung (2015) takes a fourth approach, and is able to achieve point identification without assumptions on the meeting order. Instead the game is modelled as being simultaneous (so there is no meeting order to consider), but there is also incomplete information. Specifically, the unobserved (by the econometrician) link-specific component of utility is assumed to also be unobserved by other agents. Hence agents make their decisions with only partial knowledge about what network will form. Estimation proceeds using a so-called 'two-step' estimator, analogous to that used by Bisin et al. (2011a) in a different context. First agents' beliefs about the expected state of the network are estimated non-parametrically. The observed conditional probability of a link in the network is used as an estimate for agents' belief about the probability such a link should form. This estimated network is used to replace the endogenous observed network variables that enter the utility function. Then the parameters of the utility function can be estimated directly in a second step. One advantage of this approach is that only a single network is needed to be able to estimate the utility parameters, although the network must be large.

Whether edges should be modelled as directed has consequences for identification and estimation, as well as the interpretation of the results, and will depend on features of the data used. Both Christakis et al. (2010) and Mele (2013) use data on school students from the *National Longitudinal Study of Adolescent Health* (*Add Health*), but Christakis et al. (2010) assume friendship formation is a bilateral decision whilst Mele (2013) assumes it is unilateral. The data show some edges that are not reciprocated, and it is an issue for researchers how this should be interpreted.[83] Theoretically, networks based on unilateral linking are typically modelled as being Nash equilibria of the network formation game, whilst those based on bilateral edges use *pairwise stability* (Jackson and Wolinsky, 1996) as their equilibrium concept.[84]

Both Christakis et al. (2010) and Mele (2013) assume utility functions such that the marginal utility of an edge depends on characteristics of the individuals involved, the difference in their characteristics (homophily), and some network statistics. This has two crucial implications.

First, since they assume network formation occurs sequentially, they need to assume a meeting process to 'complete' their models. This process acts as an equilibrium selection mechanism.

[83] It is sometimes argued when data contain edges that are not reciprocated that the underlying relationships are reciprocal, but that some agents failed to state all their edges. The union of the edges is then used to form an undirected graph, so $g_{ij}^{undir} = max(g_{ij}, g_{ji})$.

[84] Loosely, an undirected network is pairwise stable if (i) $G_{ij} = 1$ implies that neither i nor j would prefer to break the edge, and (ii) $G_{ij} = 0$ implies that if i would like to edge with j then j must *strictly* not want to edge with i.

Although they do not discuss equilibrium, Christakis et al. (2010) use the meeting process to determine what network should be realised for a given set of covariates and parameters. Mele (2013) makes assumptions on the structure of the utility function to ensure that at least one Nash equilibrium exists, but potentially there are multiple equilibria. The meeting process is then used to provide an ergodic distribution over these equilibria. In both cases functional form assumptions and use of a meeting order are critical to identification.[85]

Second, both papers assume that the relevant network statistics are based on purely 'local' network features. By this we mean that the marginal utility to i of forming an edge with j depends only on edges that involve either i or j. This is equivalent to the *pairwise Markovian* assumption discussed in Subsection 2.4.1. Estimation of these models can therefore be performed using the MCMC techniques described there. It also suffers from the same difficulties, *viz.* that estimation is time-consuming, and often the parameter estimates are highly unstable between runs of the estimation procedure because of the difficulty in approximating the partition function.

Hence, although in principle, it has recently become possible to estimate economic models of strategic network formation, there is still significant scope for further work to generalise these results and relax some of the assumptions that are used.

2.5 Empirical Issues

The discussion thus far has taken as given some, possibly multiple, networks $g = \{1, .., M\}$ of nodes and edges. In this section we consider where this network comes from. We begin by outlining the issues involved in defining the network of interest. We then discuss the different methods that may be used to collect data on the network, focusing on practical considerations for direct data collection and sampling methods. Our discussion thereafter examines in detail the issue of measurement error in networks data. We divide issues into those where measurement error depends on the sampling procedure, and those from other sources. Since networks are composed of interrelated nodes and edges, random (*i.e.* i.i.d.) sampling of either nodes or edges imposes some (conditionally) non-random process on the other, which depends on the structure of the underlying network, thereby generating non-classical measurement error. We discuss the implications of measurement error arising from both these sources – sampling and other – on network statistics, and on parameter estimates of models that draw on these data. Researchers working in a number of disciplines including economics, statistics, sociology and statistical physics have suggested methods for dealing with measurement error in networks data, which are described in detail thereafter.

[85] Without a meeting order, both Sheng (2012) and de Paula et al. (2014) only achieve partial identification. Leung (2015) achieves point identification by assuming agents move simultaneously and have incomplete information.

2.5.1 Defining the network

A first step in network data collection is to define, based on the research question of interest, the interaction that one would like to measure. For example, suppose one were studying the role of social learning in the adoption of a new technology, such as a new variety of seeds. In this situation, information sharing with other farmers cultivating the new variety could be considered to be the most relevant interaction. The researcher would then aim to capture interactions of this type in a network of nodes and edges. It should be noted that different behaviours and choices will be influenced by different interactions. For example, amongst households in a village, fertiliser use might be affected by the actions of other farmers, whilst fertility decisions may be influenced by social norms of what the whole village chooses. Similarly, (extended) family members are more likely to lend one money, while friends and acquaintances are often better sources of information on new opportunities.[86]

Moreover, even when the interaction of interest is well-defined, *e.g.* risk-sharing between households, there is an additional question of whether *potential* network neighbours – that is households who are willing to make a transfer or lend to one's own household – or *realised* network neighbours – the households that one's household actually received transfers or loans from – are of interest. Hence the research question of interest and the context matter, and having detailed network data is not a panacea: one must still justify why the measured network is the most relevant one for the research question being considered.

In addition, researchers are typically also forced to define a *boundary* for the network, within which all interactions are assumed to take place. Geographic boundary conditions are very common in social networks – for instance, edges may only be considered if both nodes are in the same village, neighbourhood or town – supported by the implicit assumption that a majority of interactions takes place among geographically close individuals, households and firms. Such an assumption is questionable: for example, a household's risk sharing might depend more on its edges to other households outside the village, since the geographic separation is likely to reduce the correlation between the original household's shocks and the shocks of these out-of-village neighbours. However, if justifiable, it greatly eases the logistics and reduces the costs of collecting primary network data.

Network data collection involves collecting information on two interrelated objects – nodes and edges between nodes – within the pre-defined boundary. Data used in most economic applications are typically collected as a set of observations on nodes (individuals, households, or firms), with information on the network (or group(s)) they belong to, and perhaps with

[86] The classic example of this issue comes from Granovetter (1973), who shows the importance of 'weak ties' in providing job vacancy information.

information on other nodes within the network (or group) that they are linked to. As an example, in a development context, we may have a dataset with socio-economic information on households (nodes), the village or ethnic group they belong to (group), and potentially which other households within the village its members talk to about specific issues (edges). Our focus, as elsewhere in this paper, continues to be cases where detailed information on network neighbours (*i.e.* edges) is available, although where multiple group memberships are known these may also be used to implicitly define a set of neighbours, as in De Giorgi et al. (2010).

2.5.2 Methods for Data Collection

In practical terms, a range of methods can be and have been used to collect the information needed to construct network graphs. In order to construct undirected network graphs, researchers need information on the nodes in the network, and on the edges between nodes.[87] Depending on the interaction or relationship being studied, it may furthermore be possible to obtain information on the directionality of edges between nodes, and on the strength of edges, allowing for the construction of *directed* and *weighted* graphs. The methods include:

1. Direct Elicitation from nodes:

    (a) Asking nodes to report all the other nodes they interact with in a specific dimension within the specified network boundary, *e.g.* all individuals within the same village that one lends money to. In this case, nodes are free to list whomever they want. Information on the strength of edges can similarly be collected.[88]

    (b) Asking nodes to report for every other node in the network whether they interacted with that node (and potentially the strength of these interactions). In contrast to (a), nodes are provided with a list of all other nodes in the network. Though this method has the advantage of reducing recall errors, it may generate errors from respondent fatigue in networks with a large number of nodes.

    (c) Asking nodes to report their own network neighbours and their perception of edges between other nodes in the network. This method would presumably work reasonably well in settings where, and in interactions for which, private information issues are not very important (*e.g.* kinship relations in small villages in developing countries). Alatas et al. (2014) use this method to collect information on networks in Indonesian hamlets.

[87] Some features of network graphs can be obtained without detailed information on all nodes and the edges between nodes. Degree, for instance, can be captured by asking nodes directly about the number of edges they have, without enquiring further about who these neighbours are.

[88] In practice, edge strength is usually proxied by the frequency of interaction, or the amount of time spent together, or in the case of family relationships, by the amount of shared genetic material between individuals.

(d) Asking nodes to report their participation in various groups or activities, and then imposing assumptions on interactions within the groups and activities, *e.g.* two nodes are linked if they are members of the same group. The presence of multiple groups can generate a partially-overlapping peer group structure.

2. Collection from Existing Data Sources: Edges between nodes can be constructed from information in available databases *e.g.* citation databases (Ductor et al., 2014), corporate board memberships (Patnam, 2013), online social networks (*e.g.* LinkedIn, Twitter, Facebook).

   The resulting networks often have a partially-overlapping peer group structure, with agents that share a common environment (such as a university) belonging to multiple subgroups (*e.g.* classes within the university). Network structure is then imposed by assuming that an edge exists between nodes that share a subgroup. Examples include students in a school sharing different classes (*e.g.* De Giorgi et al., 2010) or company directors belonging to the same board of directors (*e.g.* Patnam, 2013) or households which, through marriage ties of members, belong to multiple families (*e.g.* Angelucci et al., 2010).

   Moreover, the directionality of the edge can sometimes, though not always, be inferred from available data, *e.g.* data from Twitter includes information on the direction of the edge, while the existence of an edge in LinkedIn requires both nodes to confirm the edge. However, it is not possible to infer directionality among, for instance, students in a school belonging to multiple classes, since we don't even know if they actually have any relationship.

In order to generate the full network graph, researchers would need to collect data on all nodes and edges, *i.e.* they need to collect a census. This is typically very expensive, particularly since a number of methods described above in Section 2.3 exploit cross-network variation to identify parameters, meaning that many networks would need to be fully sampled.

In general, it is very rare to have data available from a census of all nodes and edges. Even when a census of nodes is available, it is very common to observe only a subset of edges because of censoring in the number of edges that can be reported.[89] In practice, given the high costs of direct elicitation of networks, and the potentially large size of networks from existing data sources, researchers usually collect data on a sample of the network only, rather than on all nodes and edges.[90] Various sampling methods have been used, of which the most common are:

[89] This is a feature of some commonly used datasets, including the popular National Longitudinal Study of Adolescent Health (AddHealth) dataset.

[90] For instance, Facebook has over 1.8 billion monthly users, while Twitter reports having around 320 million regular users.

1. Random Sampling: Random samples can be drawn for either nodes or edges. This is a popular sampling strategy due to its low cost relative to censuses. Data collected from a random sample of nodes typically contain information on socio-economic variables of interest and some (or all) edges of the sampled nodes, although data on edges are usually censored.[91] At times, information may also be available on the identities, and in some rare cases, on some socio-economic variables of all nodes in the network. Data on outcomes and socio-economic characteristics of non-sampled nodes are crucial in order to be able to implement many of the identification strategies discussed in Section 2.3 above. Moreover, as we will see below, this information is also useful for correcting for measurement error in the network. Recent analyses with networks data in the economics literature have featured datasets with edges collected from random samples of nodes. Examples include data on social networks and the diffusion of microfinance used by both Banerjee et al. (2013) and Jackson et al. (2012); and data on voting and social networks used in Fafchamps and Vicente (2013).

   Datasets constructed through the random sampling of edges include a node only if any one of its edges is randomly selected. Examples of such datasets include those constructed from random samples of email communications, telephone calls or messages. In these cases researchers often have access to the full universe of all e-mail communication, but are obliged to work with a random sample due to computational constraints.

2. Snowball Sampling and Link Tracing: Snowball sampling is popularly used in collecting data on 'hard to reach' populations *i.e.* those for whom there is a relatively small proportion in the population, so that one would get an insufficiently large sample through random sampling from the population *e.g.* sex workers. Link tracing is usually used to collect data from vast online social networks. Under both these methods, a dataset is constructed through the following process. Starting with an initial, possibly non-random, sample of nodes from the population of interest, information is obtained on either all, or a random sample of their edges. Snowball sampling collects information on all edges of the initially sampled nodes, while link tracing collects information on a random sample of these edges. In the subsequent step, data on edges and outcomes are collected from any node that is reported to be linked to the initial sample of nodes. This process is then repeated for the new nodes, and in turn for nodes linked to these nodes (*i.e.* second-degree neighbours of the initially drawn nodes) and so on, until some specified node sample size is

[91] The network graph constructed from data where nodes are randomly sampled and where edges are included only if both nodes are randomly sampled is known as an induced subgraph. The network constructed from data where nodes are randomly sampled and all their edges are included, regardless of whether the incident nodes are sampled (*i.e.* if i is randomly sampled, the edge ij will be included regardless of whether or not j is sampled), is called a star subgraph.

reached or up to a certain social distance from the initial 'source' nodes. It is hoped that, after k steps of this process, the generated dataset is representative of the population *i.e.* the distribution of sampled nodes no longer depends on the initial 'convenience' sample. However, this typically happens only when k is large. Moreover, the rate at which the dependence on the original sample declines is closely related to the extent of homophily, both on observed and unobserved characteristics, in the network. In particular, stronger homophily is associated with lower rates of decline of this dependence. Nonetheless, this method can collect, at reasonable costs, complete information on local neighbourhoods, which is needed to apply the methods outlined in Section 2.3 above. Examples in economics of datasets collected by snowball sampling include that of student migrants used in Méango (2014).

The sampling method used has important implications for how accurately the network graph and its features are measured. In the next subsection we will discuss some of the common measurement errors arising from the above methods (as well as measurement error from non-sampling sources), their implications for model parameters, and methods for overcoming these often substantial biases.

2.5.3 Sources of Measurement Error

An important challenge that complicates identification of parameters using overlapping peer groups and detailed network data is the issue of measurement error. Measurement error can arise from a number of sources including: (1) missing data due to sampling method, (2) misspecification of the network boundary, (3) top-coding of the number of edges, (4) miscoding and misreporting errors, (5) spurious nodes and (6) non-response. We refer to the first three of these as sampling-induced error, and the latter three as non-sampling error. It is important to account for this, since as we will show in this subsection, measurement error can induce important biases in measures of network statistics and in parameter estimates.

Measurement error issues arising from sampling are very important in the context of networks data, since these data comprise information on interrelated objects: nodes and edges. All sampling methods – other than undertaking a full census – generate a (conditionally) non-random sample of at least one of these objects, since a particular sampling distribution over one will induce a particular (non-random) structure for sampling over the other.[92] This means that econometric and statistical methods for estimation and inference developed under classical sampling theory are often not applicable to networks data, since many of the underlying assumptions fail to hold. Consequently the use of standard techniques, without adjustments for the specific

[92] We consider a random sample to consist of units that are independent and identically distributed.

features of network data, leads to errors in measures of the network, and hence biases model parameters.

In practice, however, censuses of networks that economists wish to study are rare, and feasible to collect only in a minority of cases (*e.g.* small classrooms or villages). Frequently, it is too expensive and cumbersome to collect data on the whole network. Moreover, when data are collected from surveys, it is common to censor the number of edges that can be reported by nodes. Finally, to ease logistics of data collection exercises, one may erroneously limit the boundary of the network to a specified unit, *e.g.* village or classroom, thereby missing nodes and edges lying beyond this boundary. Subsection 2.5.3 outlines the consequences of missing data due to sampling on estimates of social effects arising from outcomes of network neighbours (such as those considered in Subsections 2.3.2, 2.3.3 and 2.3.4) and network statistics (as in Subsection 2.3.5). Until recently most research into these issues was done outside economics, so we draw on research from a range of fields, including sociology, statistical physics, and computer science.

Measurement error arising from the other three sources – misreporting or miscoding errors, spurious nodes, and non-response – which we label as non-sampling measurement error, can also generate large biases in network statistics and parameters in network models. Though there is a large literature on these types of measurement error in the econometrics and statistics (see, for example, Chen et al. (2011) for a summary of methods for dealing with misreporting errors in binary variables, also known as misclassification errors), these issues has been less studied in a networks context. Subsection 2.5.3 below summarises findings from this literature.

Finally, a number of methods have been suggested to help deal with the consequences of measurement error, whether due to sampling or otherwise. Subsection 2.5.4 outlines the various methods that have been developed for this purpose.

Measurement Error Due to Sampling

Node-Specific Neighbourhoods Collecting only a sample of data, rather than a complete census, can lead to biased and inconsistent parameter estimates in social effect models. This is because sampling of the network leads to misspecification of nodes' neighbours. In particular, a pair of nodes in the sampled network may appear to be further away than they actually are. Recall from Section 2.3 that with observational data, methods for identifying the social effects parameters in the local average, local aggregate and hybrid local model use the exogenous characteristics of direct, second- and, in some cases, third-degree neighbours as instrumental variables for the outcomes of a node's neighbours. Critically, these methods require us to know which edges are definitely *not* present to give us the desired exclusion restrictions. Misspecification of nodes' direct and indirect (*i.e.* second- and third-degree) neighbours may consequently result in mismeasured and invalid instruments.

Chandrasekhar and Lewis (2011) show that this is indeed the case for the local average model, where the instruments are the average characteristics of nodes' second- and third-degree neighbours. The measurement error in the instruments is correlated with the measurement error in the endogenous regressors, leading to bias in the social effect estimates. Simulations in their paper suggest that these biases can be very large, with the magnitude falling as the proportion of the network sampled increases, and as the number of networks in the sample increases.[93] Chandrasekhar and Lewis (2011) offer a simple solution to this problem when (i) network information is collected via a star subgraph – *i.e.* where a subset of nodes is randomly sampled ('sampled nodes') and all their edges are included in constructing the network graph; and (ii) data on the outcome and exogenous characteristics are available for all nodes in the network, or at least for the direct and second- and potentially third-degree neighbours of the 'sampled' nodes. In this case, all variables in the second stage regression (i.e. Equation 2.6) are correctly measured for the 'sampled' nodes, since for any node, the regressors, $\tilde{\boldsymbol{G}}_{i,g}\boldsymbol{Y}_g = \sum_{j \in nei_{i,g}} \tilde{G}_{ij,g} y_{j,g}$ and $\tilde{\boldsymbol{G}}_{i,g}\boldsymbol{X}_g = \sum_{j \in nei_{i,g}} \tilde{G}_{ij,g} \boldsymbol{x}_{j,g}$, are fully observed. Including only sampled nodes in the second stage thus avoids issues of erroneously assuming that nodes in the observed network are further away from one another than they actually are. The influence matrix constructed with the sampled network is, however still mismeasured, leading to measurement error in the instruments (which use powers of this matrix), and thus in the first stage. However, this measurement error is uncorrelated with the second stage residual, thus satisfying the IV exclusion restriction. Note though that the measurement error in the instruments reduces their informativeness (strength), particularly when the sampling rate is low. This is because this strategy requires the existence of nodes that have a (finite) geodesic of at least 2 or 3 between them. At low sampling rates there will be very few such pairs of nodes, since many sampled nodes will seem completely unconnected as the nodes that connect them will be missing from the data.

A similar issue applies to local aggregate and hybrid models. Simulations in Liu (2013) show that parameters of local aggregate models are severely biased and unstable when estimated with partial samples of the true network. In this model, however, as shown in Subsection 2.3.3, a node's degree can be used as an instrument for neighbours' outcomes. When the sampled data take the form of a star subgraph, the complications arising from random sampling of nodes can be circumvented by using the out-degree, which is not mismeasured, as an instrument for the total outcome of edges. This allows for the consistent estimation of model parameters. This is

[93] A limitation of these simulations is that the authors only considered simulations with either 1 or 20 networks. It is unclear how large such biases may be when a large number (*e.g.* 50) of networks is available.

supported by simulation evidence in Liu (2013), which shows that estimates of the local aggregate model computed using out-degrees as an additional instrument are very close to the parameters of a pre-specified data generating process. Other possible ways around this problem include the model-based and likelihood-based corrections outlined in Subsection 2.5.4.

Network Statistics Missing data arising from partial sampling generate non-classical measurement error in measured network statistics. This is an important issue in estimating the effects of network statistics on outcomes using regressions of the form seen in Subsection 2.3.5, because measurement error leads to substantial bias in model parameter estimates. A number of studies, primarily in fields outside economics, have investigated the consequences and implications of sampled network data on measures of network statistics and model parameters. The following broad facts emerge from this literature:

1. *Network statistics computed from samples containing moderate (30-50%) and even relatively high (∼70%) proportions of nodes in a network can be highly biased. Sampling a higher proportion of nodes in the network generates more accurate network statistics.* We illustrate the severity of this issue using a stylised example. Consider the network in panel (a) of Figure 2.4, which contains 15 nodes and has an average degree of 3.067. We sample 60%, 40% and 20% of nodes and elicit information on all their edges (*i.e.* we elicit a star subgraph). The resulting network graphs are plotted in panels (b), (c) and (d), with the unshaded nodes being those that were not sampled. Average degree is calculated based on all nodes and edges in the star subgraph, *i.e.* including all sampled nodes, the edges they report, and nodes they are linked with.[94] When only 20% of nodes are sampled, the average degree of the sampled graph is 2, which is around 35% lower than the true average degree.[95] However, when a higher proportion of nodes are sampled, average degree of the sampled graph becomes closer to that of the true graph. More generally, simulation evidence from studies including Galaskiewicz (1991), Costenbader and Valente (2003), Lee et al. (2006), Kim and Jeong (2007) and Chandrasekhar and Lewis (2011) have estimated the magnitude of sampling induced bias in statistics such as degree (in-degree and out-degree in the directed network case), degree centrality, betweenness centrality, eigenvector centrality, transitivity (also known as local clustering), and average path length.[96] They find biases that are very large in magnitude, and the direction of the bias varies depending

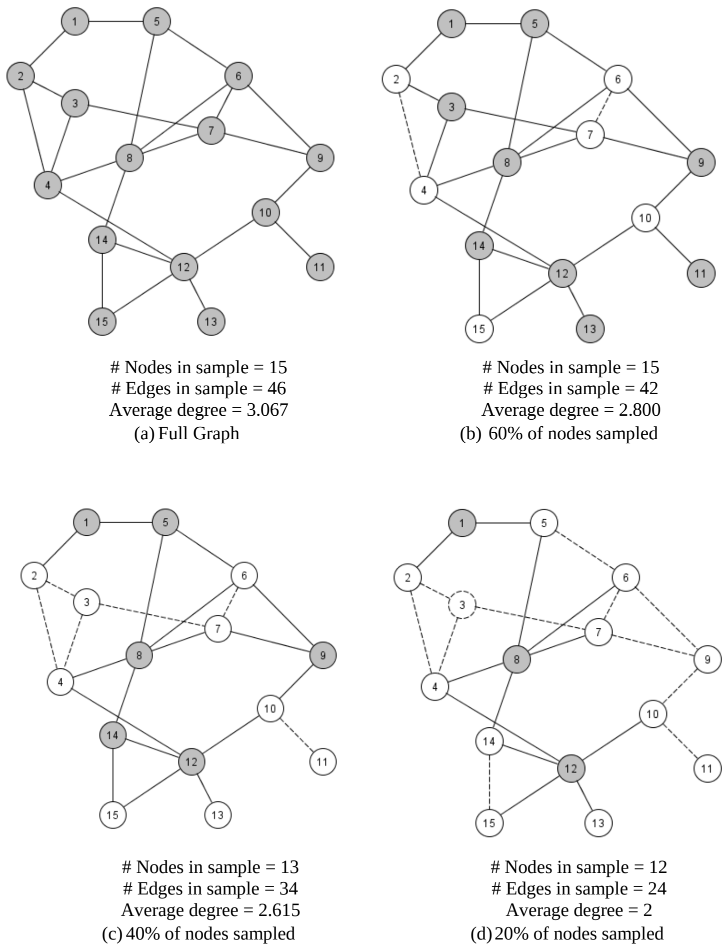
[94] This is equivalent to taking an average of the row-sums of the (undirected) adjacency matrix constructed from the sampled data, in which two nodes are considered to be connected if one reports an edge. This is a common way of constructing the adjacency matrix in empirical applications. However, for data collected through star subgraph sampling, an accurate estimate of average degree can be obtained by including only the sampled nodes in the calculation.

[95] We will discuss methods that allow one to correct for this bias in Subsection 2.5.4.

[96] Simulations are typically conducted by taking the observed network to be the true network, and constructing 'sampled' networks by drawing samples of different sizes using various sampling methods.

on the statistic. For example, the average path length may be over-estimated by 100% when constructed from an induced subgraph with 20% of nodes in the true network. This concern is particularly relevant for work in the economics literature: a literature review of studies in economics by Chandrasekhar and Lewis (2011) reports a median sampling rate of 25% of nodes in a network. Table 2.1 below summarises findings from these papers for various commonly used network statistics.

Figure 2.4: *Sampled networks with different sampling rates*

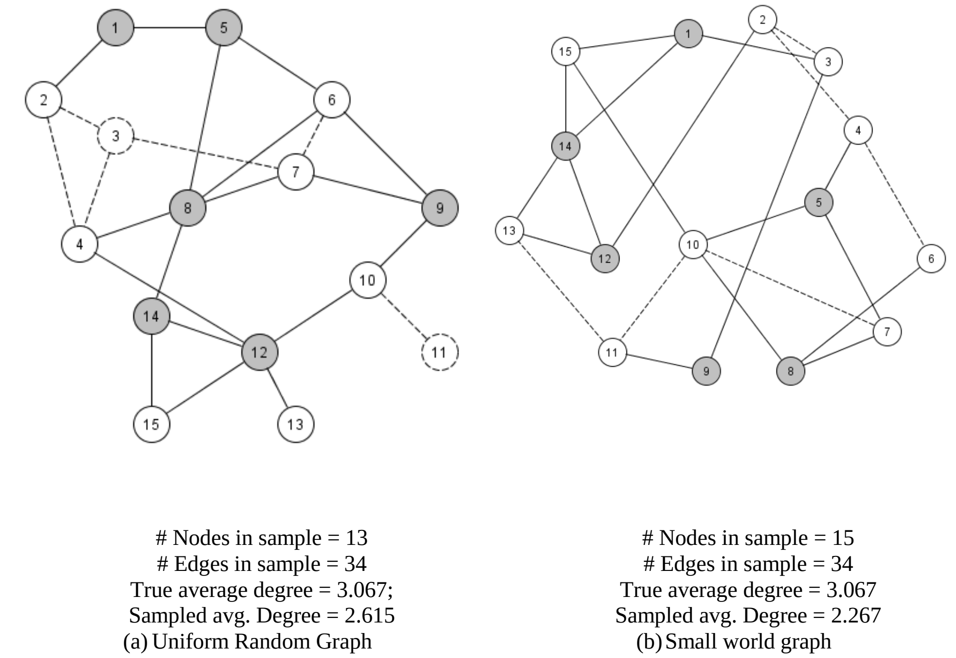


Notes: *This figure displays the full graph (panel (a)), and the star subgraphs obtained from sampling 60% (panel (b)), 40% (panel (c)) and 20% (panel (d)) of nodes. The unshaded nodes in panels (b), (c) and (d) represent nodes that were not sampled, and the dotted lines represent nodes and edges on which no data were collected. Though the average degree in the original graph is 3.067, that in the sampled graphs ranges from 2.8 to 2. The # Nodes, and # Edges indicated in the figure refer to the numbers included in the calculation of the displayed average degree.*

2. *Measurement error due to sampling varies with the underlying network topology (i.e. structure).* This is apparent from work by Frantz et al. (2009), who investigate the robustness of a variety of centrality measures to missing data when data are drawn from a range of underlying network topologies: uniform random, small world, scale-free, core-periphery and cellular networks (see Subsection 2.7.1 for definitions). They find that the accuracy of

centrality measures varies with the topology: small world networks, which have relatively high clustering and 'bridging' edges that reduce path lengths between nodes that would otherwise be far away from one another, are especially vulnerable to missing data. This is not surprising since key nodes that are part of a bridge could be missed in the sample and hence give a picture of a less connected network. By contrast, scale-free networks are less vulnerable to missing data. Such effects are evident even in the simple stylised example in Figure 2.5 below, where we sample the same nodes from networks with different topologies – uniform random, and small world. Though each network has the same average degree, and the same number of nodes is sampled in both cases, the average degree in the graph sampled from the uniform random network is closer to the true value than that sampled from the small world network.[97]

Figure 2.5: *Sampling from uniform random and small world networks*



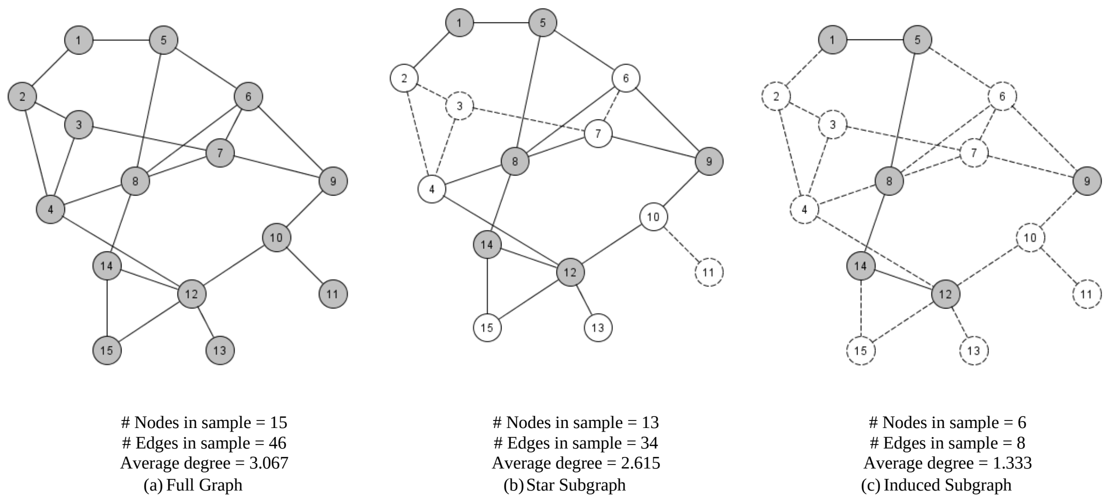
Notes: *This figure displays the star subgraphs obtained from sampling 40% of nodes in a network with a uniform random topology (panel (a)) and a small world topology (panel(b)). The unshaded nodes represent nodes that were not sampled, and the dotted lines represent nodes and edges on which no data were collected.*

3. *The magnitude of error in network statistics due to sampling varies with the sampling method.* Different sampling methods result in varying magnitudes of errors in network statistics. Lee et al. (2006) compare data sampled via induced subgraph sampling, random sampling of nodes, random sampling of edges, and snowball sampling, from networks with

[97]As in (1) above, average degree is calculated from the adjacency matrix with all nodes and edges in the sample (*i.e.* all the nodes and edges with firm lines).

a power-law degree distribution.[98] They show that the sampling method impacts the magnitude and direction of bias in network statistics. For instance, random sampling of nodes and edges leads to an over-estimation of the size of the exponent of the power-law degree distribution.[99] Conversely, snowball sampling, which is less likely to find nodes with low degrees, underestimates this exponent. We illustrate this fact further using a simple example that compares two node sampling methods common in data used by economists – *induced subgraph*, where only edges between sampled nodes are retained; and *star subgraph*, in which all edges of sampled nodes are retained regardless of whether or not the nodes involved in the edges were sampled. Consider again the network graph considered in panel (a) of Figure 2.4 above, and displayed again in panel (a) of Figure 2.6 below. We sample the same set of nodes – 1, 5, 8, 9, 12, and 14 – from the full network graph. Panels (b) and (c) of Figure 2.6 display the resulting network graphs under star and induced subgraph sampling respectively. Though the proportion of the network sampled is the same under both types of sampling, the resulting network structure is very different. This is reflected in the estimated network statistics as well: the average degree for the induced subgraph is just over a half of that for the star subgraph, which is not too different from the average degree of the full graph.[100]

Figure 2.6: *Sampling with star and induced subgraphs*



Notes: *Panel (a) of the figure displays the true network graph and panels (b) and (c) display the star and induced subgraph obtained when the darker-shaded nodes are sampled. The unshaded nodes in panels (b) and (c) represent nodes that were not sampled, and the dotted lines represent nodes and edges on which no data were collected. In the star subgraph, an edge is present as long as one of the two nodes involved in the edge is sampled. This is not the case in the induced subgraph, where an edge is present only if both nodes involved in the edge are sampled.*

[98] Power law degree distributions are those where the fraction of nodes having k edges, $P(k)$ is asymptotically proportional to $k^{-\gamma}$, where usually $2 < \gamma < 3$. Such a distribution allows for fat tails, *i.e.* the proportion of nodes with very high degrees constitutes a non-negligible proportion of all nodes.

[99] A larger exponent on the power law degree distribution indicates a greater number of nodes with large degrees.

[100] Average degree is calculated as above, including all nodes and edges in the sample, *i.e.* those with firm lines in Figure 2.6.

4. *Parameters in economic models using mismeasured network statistics are subject to substantial bias.* Sampling induces non-classical measurement error in the measured statistic; *i.e.*, the measurement error is not independent of the true network statistic. Chandrasekhar and Lewis (2011) suggest that sampling-induced measurement error can generate upward bias, downward bias or even sign switching in parameter estimates. The bias is large in magnitude: for statistics such as degree, clustering, and centrality measures, they find that the mean bias in parameters in network level regressions ranges from over-estimation bias of 300% for some statistics to attenuation bias of 100% for others when a quarter of network nodes are sampled.[101] As with network statistics, the bias becomes smaller in magnitude as the proportion of the network sampled increases. The magnitude of bias is somewhat smaller, but nonetheless substantial, for node-level regressions. Table 2.2 summarises the findings from the literature on the effects of random sampling of nodes on parameter estimates.

5. *Top-coding of edges or incorrectly specifying the boundary of the network biases network statistics.* Network data collected through surveys often place an upper limit on the number of edges that can be reported. Moreover, limiting the network boundary to an observed unit, *e.g.*, a village or classroom, will miss nodes and edges beyond the boundary. Kossinets (2006) investigates, via simulations, the implications of top-coding in reported edges and boundary specification on network statistics such as average degree, clustering and average path length. Both types of error cause average degree to be under-estimated, while average path length is over-estimated. No bias arises in the estimated clustering parameter if the consequence of the error is to simply limit the number of edges of each node.

Tables 2.1 and 2.2 below summarises findings on the consequences of missing data for both estimates of network statistics and parameter estimates when using data on networks collected through random sampling of nodes. We consider two types of graph induced by data collected via random node sampling: induced subgraph, and star subgraph, which are as shown in Figure 2.6 above.

[101] Simulations typically report bias in parameters from models where the outcome variable is a linear function of the network statistic.

Other Types of Measurement Error

Beyond sampling-induced measurement error, networks could be mismeasured for a variety of other reasons including:

1. Miscoding and misreporting errors: Edges could be miscoded, either because of respondent or interviewer error: respondents may forget nodes or interview fatigue may lead them to misreport edges. In some cases, there may be strategic reporting of edges, *e.g.*, respondents may report desired rather than actual edges, as in Comola and Fafchamps (2014).

2. Spurious Nodes: Spelling mistakes in node names or multiple names for the same nodes can lead to the presence of spurious nodes. This is a concern when edges are inferred from existing data.

3. Non-response: Edges are missing as a result of non-response from nodes.

Wang et al. (2012) consider, in a simulation study, the consequences of these types of measurement error on network statistics including degree centrality, the clustering coefficient and eigenvector centrality. They find that degree centrality and eigenvector centrality are relatively robust to measurement error arising from spurious nodes and miscoded edges, while clustering coefficient is biased by mismeasured data. Though there is a large literature on these types of measurement error in the econometrics and statistics (see, for example, Chen et al. (2011) for a summary of methods for dealing with misreporting errors in binary variables, also known as misclassification errors), these issues has been less studied in a networks context. An exception is Comola and Fafchamps (2014), who propose a method for identifying and correcting misreported edges.

2.5.4 Correcting for Measurement Error

Ex-post (*i.e.* once data have been collected) methods of dealing with measurement error can be divided into three broad classes: (1) design-based corrections, (2) model-based corrections, and (3) likelihood-based corrections. Design-based corrections apply primarily to correcting sampling-induced measurement error, while model-based and likelihood-based corrections can apply to both sampling-induced and non-sampling-induced measurement error. We briefly summarise the underlying ideas behind each of these, discussing some advantages and drawbacks of each.

Table 2.1: *Findings from literature on sampling-induced bias in estimated network statistics*

Statistic	Measurement error in statistic	
Network-Level Statistics	Star Subgraph	Induced Subgraph
Average Degree	Underestimated (–) if non-sampled nodes are included in the calculation. Otherwise sampled data provide an accurate measure.[a]	Underestimated (–).[a]
Average Path length	Not known.	Over-estimated (+); network appears less connected; magnitude of bias very large at low sampling rates, and falls with sampling rate.[b]
Spectral gap	Direction of bias ambiguous (±); depends on the relative magnitudes of bias in the first and second eigenvalues, both of which are attenuated.[a]	Direction of bias ambiguous (±): depends on the relative magnitudes of bias in the first and second eigenvalues, both of which are attenuated.[a]
Clustering Coefficient	Attenuation (–) since triangle edges appear to be missing.[a]	Little/no bias. Random sampling yields same share of connected edges between possible triangles.[a,b]
Average Graph Span	Overestimation (+) of the graph span: sampled network is less connected than the true network. At low sampling rates, graph span may appear to be small, depending on how nodes not in the giant component are treated.[a]	Overestimation (+) of the graph span: sampled network is less connected than the true network. At low sampling rates, graph span may appear to be small, depending on how nodes not in the giant component are treated.[a]

Notes. *Non-negligible, or little bias refers to | bias | of 0-20%, large bias to | bias | of 20%-50% and very large bias to | bias | > 50%.*
[a] *Source: Chandrasekhar and Lewis (2011);* [b] *Source: Lee et al. (2006).*

Table 2.1 cont.

Statistic	Measurement error in statistic	
Node - Level Statistics	Star Subgraph	Induced Subgraph
Degree (In and Out in directed graphs)	In-degree and out-degree both underestimated (–) if all nodes in sample included in calculation. If only sampled nodes included, out-degree is accurately estimated. In undirected graphs, underestimation (–) of degree for non-sampled nodes.[a]	Degree (in undirected graphs) of highly connected nodes is underestimated (–).[b]
Degree Centrality (Degree Distribution)	Not known.	Overestimation (+) of exponent in scale-free networks $\Rightarrow$ degree of highly connected nodes is underestimated. Rank order of nodes across distribution considerably mismatched as sampling rate decreases.[b]
Betweenness Centrality	Distance between true betweenness centrality distribution and that from sampled graph decreases with the sampling rate. At low sampling rates (*e.g.* 20%), correlations can be as low as 20%.[a]	Shape of the distribution relatively well estimated. Ranking in distribution much worse, *i.e.* nodes with high betweenness centrality appear to have low centrality.[d]
Eigenvector Centrality	Very low correlation between vector of true node eigenvector centralities and that from sampled graph.[a]	Not known.

Notes. [a]*Source: Costenbader and Valente (2003);* [b]*Source: Lee et al. (2006);* [c]*Source: Kim and Jeong (2007).*

Table 2.2: *Findings from literature on sampling-induced bias in parameter estimates*

Statistic	Bias in Parameter Estimates	
Network Level Statistics	Star Subgraph	Induced Subgraph
Average Degree	Scaling (+) and attenuation (–), both of which fall with sampling rate when all nodes in sample included in calculation; $\lvert\text{scaling}\rvert > \lvert\text{attenuation}\rvert$. No bias if only sampled nodes included.	Scaling (+) and attenuation (–), both of which fall with sampling rate; $\lvert\text{scaling}\rvert > \lvert\text{attenuation}\rvert$. Magnitude of bias higher than for star subgraphs.
Average Path length	Attenuated (–). Magnitude of bias large and falls with sampling rate.	Attenuated (–) (more than star subgraphs). Magnitude of bias is very large at low sampling rates, and falls with sampling rate.
Spectral gap	Attenuated (–), with bias falling with sampling rate. Bias magnitude large even when 50% nodes sampled.	Attenuated (–) (more than star subgraphs). Bias magnitude very large and falls with sampling rate.
Clustering Coefficient	Scaling (+) and attenuation (–); $\lvert\text{scaling}\rvert > \lvert\text{attenuation}\rvert$. Very large biases, which fall with sampling rate.	Attenuation (–), falls with sampling rate. Magnitude of bias non-negligible at node sampling rates of $<40\%$.
Average Graph Span	Estimates have same sign as true parameter if node sampling rate is sufficiently large; Can have wrong sign if sampling rate is too low, depending on how nodes not connected to the giant component are treated in the calculation.	Estimates have same sign as true parameter if node sampling rate is sufficiently large; Can have wrong sign if sampling rate is too low, depending on how nodes not connected to the giant component are treated in the calculation.

Notes. *Non-negligible bias refers to $\lvert bias\rvert$ of 0-20%, large bias to $\lvert bias\rvert$ of 20%-50% and very large bias to $\lvert bias\rvert > 50\%$. Source: Chandrasekhar and Lewis (2011).*

Table 2.2 cont.

Statistic	Bias in Parameter Estimates	
Node - Level Statistics	Star Subgraph	Induced Subgraph
Degree (In and Out in directed graphs)	Attenuation (–), with the magnitude of bias falling with the sampling rate. The magnitude of bias is large even when 50% of nodes are sampled.	Scaling (+), with the bias falling with the node sampling rate. Bias is very large in magnitude.
Degree Centrality (Degree Distribution)	Not known.	Not known.
Betweenness Centrality	Not known.	Not known.
Eigenvector Centrality	Attenuation (–), with magnitude of bias falling with the sampling rate. Magnitude of bias large even when 50% of nodes are sampled.	Attenuation (–), with magnitude of bias falling with the sampling rate. Magnitude of bias very large.

Notes. *Large bias refers to —bias— of 20%-50% and very large bias to —bias— ¿ 50%. Source: Chandrasekhar and Lewis (2011).*

Design-Based Corrections

Design-based corrections rely on features of the sampling design to correct for sampling-induced measurement error (Frank 1978, 1980a, 1980b, 1981; Thompson, 2006).[102] They are based on *Horvitz-Thompson* estimators, which use inverse probability-weighting to compute unbiased estimates of population totals and means from sampled data. This method can be applied to correct mismeasured network statistics that can be expressed as totals, such as average degree and clustering. We illustrate how Horvitz-Thompson estimators work using a simple example.

A researcher has data on an outcome y for a sample of n units drawn from the population. Under the particular sampling scheme used to draw this sample, each unit i in the population $U = \{1, ..., N\}$ has a probability p_i of being in the sample. The researcher wants to use the sample to compute an estimate of the sum of y in the population, $\tau = \sum_{i \in U} y_i$. The Horvitz-Thompson estimator for this total can be computed by summing the y's for the sampled units, weighted by their probability of being in the sample. That is, $\hat{\tau}_p = \sum\limits_{i \in U} \frac{y_i}{p_i}$. Essentially, the estimator computes an inverse probability-weighted estimate to correct for bias arising from unequal probability sampling. In the case of network statistics, this thus corrects for the non-random sampling of either nodes or edges induced by the particular sampling scheme. The key to this approach is the construction of the sample inclusion weights, p_i.

Formulae for node- and edge-inclusion probabilities are available for the random node and edge sampling schemes (see Kolaczyk (2009) for more details). Recovering sample inclusion probabilities when using snowball sampling is typically not straightforward after the first step of sampling. This is because every possible sample path that can be taken in subsequent sampling steps must be considered when calculating the sample-inclusion probability, making this exercise very computationally intensive. Estimators based on Markov chain resampling methods, however, make it feasible to estimate the sample inclusion probabilities. See Thompson (2006) for more details.

Frank (1978, 1980a, 1980b, 1981) derives unbiased estimators for graph parameters such as dyad and triad counts, degree distribution, average degree, and clustering under random sampling of nodes. Chandrasekhar and Lewis (2011) show that parameter estimates in network regressions using design-based corrected network statistics as regressors are consistent for three statistics: average degree, clustering coefficient, and average graph span. Their results show that the Horvitz-Thompson estimators can correct for sampling-induced measurement error. Numerical simulations suggest that this method reduces greatly, and indeed eliminates at sufficiently high sampling rates, the sampling induced bias in parameter estimates.

[102] Chapter 5 of Kolaczyk (2009) provides useful background on these methods.

There are two drawbacks of this procedure. First, it is not possible to compute Horvitz-Thompson estimators for network statistics that cannot be expressed as totals or averages. This includes node level statistics, such as eigenvector centrality, many of which are statistics of interest for economists. Second, they can't be used to correct for measurement error arising from reasons other than sampling (unless the probability of correct reporting is known). Model-based and likelihood-based corrections can, by placing more structure on the measurement error problem, offer alternative ways of dealing with measurement error in these cases.

Model-Based Corrections

Model-based corrections provide an alternative approach to correcting for measurement error. Such corrections involve specifying a model that maps the mismeasured network to the true network and have primarily been used to correct for measurement error arising from sampling related reasons. Thus the model is typically a network formation model of the type seen in Subsection 2.4.1 above. Parameters of the network formation model are estimated from the partially observed network, and available data on the identities and characteristics of nodes and edges; with the estimated parameters subsequently used to predict missing edges (in-sample edge prediction). Note that it is crucial to have information on the identities and, if possible, the characteristics (*e.g.* gender, ethnicity, *etc.*) of all nodes in the network. This is important from a data requirements perspective. Without this information, it is not possible to use this method to correct for measurement error.

In most economics applications, researchers would typically want to use the predicted networks to subsequently identify social effect parameters using models similar to those in Section 2.3 above. Chandrasekhar and Lewis (2011) show that the network formation model must satisfy certain conditions in order to allow for consistent estimation of the parameters of social effects models such as those discussed in Section 2.3.

They study a setting where data on the network is assumed to be missing at random, and where the identities and some characteristics of all nodes are observed. Data are assumed to be available for multiple, possibly large networks. This is necessary since in their results the rate of convergence of the estimated parameter to the true parameter depends on both the number of nodes within a network, and the number of networks in the data. Their analysis shows that consistent estimation of social effect parameters is possible with network formation models similar to those outlined in Subsection 2.4.1 above, as long as the interdependence between the covariates of pairs of nodes decays sufficiently fast with network distance between the nodes. This may not be satisfied for instance, in a model where a network statistic (such as degree distribution) is a sufficient statistic for the network formation process. In this case, Chandrasekhar and Lewis

(2011) show that parameters of the network formation process do not converge sufficiently fast to allow for consistent estimation of the social effect parameters in models at the node-level (*e.g.* Equation 2.1), though parameters of network-level models, such as Equation 2.5 can be consistently estimated. Their analysis also shows that network formation processes that allow for specific network effects in edge formation (*i.e.* some strategic models of network formation such as the model of Christakis et al., 2010) also satisfy conditions under which the social effect parameter can be consistently estimated.

Likelihood-Based Corrections

Likelihood-based corrections can be applied to correct for measurement error when only a subsample of nodes in a network are observed. Such methods have, however, been used to correct specific network-based statistics such as out-degree and in-degree, but may not apply to other statistics. Here, we discuss two likelihood-based methods to correct for measurement error: the first method from Conti et al. (2013), corrects for sampling related measurement error when data is available only for sampled nodes; while the second has been proposed and applied by Comola and Fafchamps (2014) to correct for misreporting.

Conti et al. (2013) correct for non-classical measurement error in in-degree arising from random sampling of nodes by adjusting the likelihood function to account for the measurement error. The method involves first, specifying the process for outgoing and incoming edge nominations, and as a result obtaining the outgoing and incoming edge probabilities. Specifically, Conti et al. (2013) assume that outgoing (incoming) edge nominations from i to j are a function of i's (j's) observable preferences, the similarity between i and j's observable characteristics (to capture homophily) and a scalar unobservable for i and j. Moreover, the process allows for correlations between i's observable and j's unobservable characteristics (and vice versa). When edges are binary, the out-degree and in-degree have binomial distributions with the success probability given by the calculated outgoing and incoming edge probabilities. Random sampling of nodes to obtain a star subgraph generates measurement error in the in-degree, but not in the out-degree. However, since the true in-degree is binomially distributed, and nodes are randomly sampled, the observed in-degree has a hypergeometric distribution conditional on the true in-degree. Knowledge of these distributions allows for the specification of the joint distribution of the true in-degree, the true out-degree and the mismeasured in-degree. Pseudolikelihood functions can therefore be specified allowing for parameters to be consistently estimated via maximum likelihood methods.[103]

[103] Conti et al. (2013) also account for censoring by using a truncated distribution in the likelihood function.

Comola and Fafchamps (2014) propose a maximum likelihood based framework to correct for measurement error arising from misreporting by nodes of their neighbours and/or flows across the edges. To illustrate this method, we take the case of binary edges. In survey data, where nodes are asked to declare the presence or not of an edge with other nodes, misreporting could mean that one of two nodes in any edge omits to report the edge; or both forget to report the edge even if it exists, or both report an edge when it doesn't exist or, one of the two nodes erroneously reports an edge when it doesn't exist. Misreporting in this case is a form of misclassification error. Assuming that the misreporting process is such that either nodes forget to declare neighbours, or they spuriously report neighbours, it is possible to use a maximum likelihood framework to correct for this misreporting bias. By assuming a statistical process for edges (*e.g.* Comola and Fafchamps (2014) assume that edges follow a logistic process, and are a function of observed characteristics), and given that the mismeasured variable is binary, it is possible to write down a likelihood function that incorporates the measurement error. Maximising this function provides the correct parameter estimates for the edge formation process, which can then be used to correct for misreporting.

2.6 Conclusion

Networks can play an important role both as a substitute for incomplete or missing markets and a complement to markets, for example, by transmitting information, or even preferences. Whether such effects exist in practice is an important empirical question, and recent work across a range of fields in economics has tried to provide some evidence about this. However, working with networks data creates important challenges that are not present in other contexts.

In this paper we outline econometric methods for working with network data that take account of the peculiarities of the dependence structures present in this context. It divides the issues into three parts: (i) estimating social effects given a conditionally exogenous observed network; (ii) estimating the underlying network formation process, given only a single cross-section of data; and (iii) accounting for measurement error, which in a network context can have particularly serious consequences.

When data are available on only agents and the reference groups to which they belong, researchers have for some time worried about how social effects might be identified. However, when detailed data on nodes and their individual links are present, identification of social effects (taking the network as conditionally exogenous) is generic, and estimation is relatively straightforward. Two broader conceptual issues exist in this case: First, theory is often silent on the precise form that peer effects should take when they exist. Since Manski (1993), many people have focused on the 'local average' framework, often without discussion of the implications for economic behaviour, but social effects might instead take a local aggregate, or indeed local maximum/minimum form where the best child in a classroom provides a good example to all others, or the worst disrupts the lesson. Until a non-parametric way of allowing for social effects is developed, researchers need to use theory to guide the empirical specification they use. Second, researchers typically treat the observed network as the network which mediates the social effect, and where many networks are observed the union of these is taken. Given what we know about measurement error in networks, this behaviour will generally create important biases in results, if the relevant network is a network defined by a different kind of relationship, or is actually some subset of the union taken. Here again it is important that some justification is given for why the network used should be the appropriate one.

In addition to these conceptual issue, the key econometric challenge in identifying social effects is allowing for network endogeneity. In recent years there have been attempts to account directly for network endogeneity. A natural first direction for this work has been to use exclusion restrictions to provide an instrument for the network structure. As ever, this requires us to be able to credibly argue that there is some variable that indirectly affects the outcome of interest, through its effect on the network structure, but has no direct effect. Whether this seems reasonable will depend on the circumstance, but an important issue here is that the network formation process must have a unique equilibrium for these methods to be valid.

This leads naturally to a discussion of network formation models that can allow for dependence between links. Drawing from work in a number of fields, this paper brings together the main estimation methods and assumptions, describing them in a common language. Although other fields have modelled network formation for some time, and developed methods to estimate parameters, they are often unsuitable when we treat the data as observations of decisions made by optimising agents. There is still much scope in this area to develop more general methods and results which do not rely on strong assumptions about the structure of utility functions or meeting processes in order to achieve identification.

Finally, the paper discussed data collection and measurement error. Since networks comprise of interrelated nodes and edges, a particular sampling scheme over one of these objects will imply a structure for sampling over the other. Hence one must think carefully in this context about how data are collected, and not simply rely on the usual intuitions that random sampling (which is not even well-defined until we specify whether it is nodes or edges over which we define the sampling) will allow us to treat the sample as the population. When collecting census data is not feasible, it will in general be necessary to make corrections for the induced measurement error, in order to get unbiased parameter estimates. Whilst there are methods for correcting some network statistics for some forms of sampling, again there are few general results, and consequently much scope for research.

Much work has been done to develop methods for working with networks data, both in economics and in other fields. Applied researchers can therefore take some comfort in knowing that many of the challenges they face using these data are ones that have been considered before, and for which there are typically at least partial solutions already available. Whilst the limitations of currently available techniques mean that empirical results should be interpreted with some caution, attempting to account for social effects is likely to be less restrictive than simply imposing that they cannot exist.

2.7 Appendix

2.7.1 Definitions

Here we provide an index of definitions for the different network representations and summary statistics used.

- **Adjacency Matrix:** This is an $N \times N$ matrix, $\boldsymbol{G}$, whose ij^{th} element, G_{ij}, represents the relationship between node i and node j in the network. In the case of a binary network, the elements G_{ij} take the value 1 if i and j are linked, and 0 if they are not linked; while in a weighted network, $G_{ij} = w(i, j)$, where $w(i, j)$ is some measure of the strength of the relationship between i and j. Typically, the leading diagonal of $\boldsymbol{G}$ is normalised to 0.

- **Influence Matrix:** This is a row-stochastic (or 'right stochastic') adjacency matrix, $\tilde{\boldsymbol{G}}$ whose elements are generally defined as $\tilde{G}_{ij} = {}^{G_{ij}}/_{\sum_j G_{ij}}$ if two agents are linked and 0 otherwise.

- **Degree:** A node's degree, d_i, is the number of edges of the node in an undirected graph. The degree of node i in the network with a binary adjacency matrix, $\boldsymbol{G}$, can be calculated by summing the elements of the i^{th} row of this matrix.[104] In a directed graph, a node's **in-degree** is the number of edges from other nodes to that node, and it's **out-degree** is the number of edges from that node to other nodes in the network. For node i, the former can be calculated by summing the elements of the i^{th} column of the binary adjacency matrix for the network, while the latter is obtained by summing the i^{th} row of this matrix.

- **Average degree:** The average degree for a network graph is the average number of edges that nodes in the network have.

- **Density:** The relative fraction of edges that are present in a network. It is calculated as the average degree divided by $N - 1$, where N is the number of nodes in the network.

- **Shortest path length (geodesic):** A path in a network g between nodes i and j is a sequence of edges, $i_1i_2, i_2i_3, ..., i_{R-1}i_R$, such that $i_ri_{r+1} \in g$, for each $r \in \{1, ..., R\}$ with $i_1 = i$ and $i_R = j$ and such that each node in the sequence $i_1, ..., i_R$ is distinct. The shortest path length or geodesic between i and j is the path between i and j that contains the fewest edges. The average geodesic of a network is the average geodesic for every pair of nodes in the network. For nodes for whom no path exists, it is common to either exclude them from the calculation of the average geodesic (i.e. to calculate the average geodesic

[104] Similarly, for a weighted graph, summing the elements for row i in the adjacency matrix yields the weighted degree.

from the connected part of the network) or to define the geodesic for these nodes to be some large number (usually greater than the largest geodesic in the network).

- **Diameter:** The diameter of a graph is the largest geodesic in the connected part of the network, where by connected, we refer to nodes for whom a path exists to get from one node to the other.
- **Component:** A connected component, or component, in an undirected network is a subgraph of a network such that every pair of nodes in the subgraph is connected via some path, and there exists no edge from the subgraph to the rest of the network.
- **Bridge:** The edge ij is considered to be a bridge in the network g if removing the edge ij results in an increase in the number of components in g.
- **Complete Network:** A network in which all possible edges are present.
- **Degree Centrality:** This is the node's degree divided by $N-1$, where N is total number of nodes in the network. It measures how well a node is connected in terms of direct neighbours. Nodes with a large degree have a high degree centrality.
- **Betweenness centrality**: This is a measure of centrality based on how well situated a node is in terms of the paths it lies on. The importance of node i in connecting nodes j and k can be calculated as the ratio of the number of geodesics between j and k that i lies on to the total number of geodesics between j and k. Averaging this ratio across all pairs of nodes yields the betweenness centrality of node i.
- **Eigenvector centrality**: A relative measure of centrality, the centrality of node i is the sum of the centrality of its neighbours. It can be calculated by solving the following equation in matrix terms, $\lambda C^e(\boldsymbol{G}) = \boldsymbol{G}C^e(\boldsymbol{G})$, where $C^e(\boldsymbol{G})$ is an eigenvector of $\boldsymbol{G}$, and λ is the corresponding eigenvalue.
- **Bonacich Centrality:** Another measure of centrality that defines a node's centrality as a function of their neighbours' centrality. It is defined as $\boldsymbol{b}(\boldsymbol{G}_g, \beta) = (\boldsymbol{I}_g - \beta\boldsymbol{G}_g)^{-1}.(\alpha\boldsymbol{G}_g\boldsymbol{\iota})$.
- **Dyad count:** A dyad is a pair of nodes. In an undirected network, the dyad count is the number of edges in the network.
- **Triad count:** A triad is a triple of nodes such that a path connecting all 3 nodes exists. The triad count of an undirected network is the number of such triples in the network.
- **Clustering coefficient:** For an undirected network, this measures the proportion of fully connected triples of nodes out of all potential triples in which at least two edges are present.

- **Support:** An edge $ij \in \mathscr{E}_g$ is supported if there exists an agent $k \neq i, j$ such that $ik \in \mathscr{E}_g$ and $jk \in \mathscr{E}_g$.

- **Expansiveness:** For subsets of connected nodes in the network, the ratio of the number of edges connecting the subset to the rest of the network to the number of nodes in the subset.

- **Sparseness:** A property of the network related with the length of all minimal cycles connecting triples of nodes in the network. For any integer, $q \geq 0$, a network is q-sparse if all minimal cycles connecting any triples of nodes (i, j, k) such that $ij \in \mathscr{E}_g$ and $jk \in \mathscr{E}_g$ have length $\leq q + 2$. See Bloch et al. (2008) for more details.

- **Graph span:** The graph span is a measure that mimics the average path length. It is defined as

$$span_g = \frac{log(N_g) - log(d_g)}{log(\tilde{d}_g) - log(d_g)} + 1$$

where N_g is the number of nodes in network g, d_g is the average degree of network g and $\tilde{d}_g$ is the average number of second-degree neighbours in the network.

Network Topologies

- **Bipartite network:** A network whose set of nodes can be divided into two sets, U and V, such that every edge connects a node in U to one in V.

- **Uniform random network:** A graph where edges between nodes form randomly.

- **Scale-free network:** A network whose degree distribution follows a power law, i.e. where the fraction of nodes having k edges, $P(k)$ is asymptotically proportional to $k^{-\gamma}$. Such a distribution allows for fat tails, *i.e.* the proportion of nodes with very high degrees constitutes a non-negligible proportion of all nodes.

- **Core-periphery network:** A network that can be partitioned into a set of nodes that is completely connected ('core'), and another set of agents ('periphery') who are linked primarily with nodes in the 'core'.

- **Cellular network:** Networks containing many sets of completely connected nodes (or 'cliques'), with few edges connecting the different cliques.

- **Small world network:** A network where most nodes are not directly linked to one another, but where geodesics between nodes are small, i.e. a node can reach every other node in the network by passing through a small number of nodes.

- **k-star:** A component with k nodes and $k-1$ links such that there is one 'hub' node who has a direct link to each of the $(k-1)$ other ('periphery') nodes.

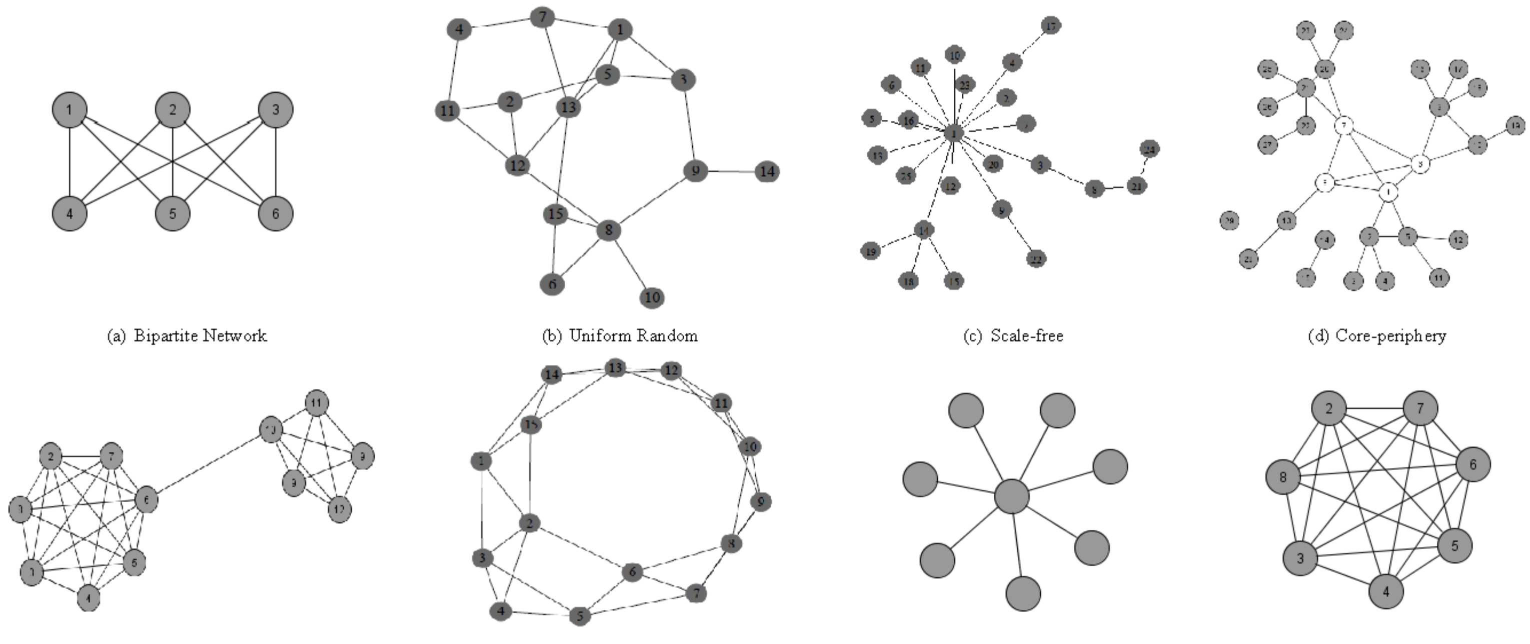
- **Cliques:** A clique is any induced subgraph of a network (*i.e.* subset of nodes and all edges between them) such that every node in the subgraph is directly connected to every other node in the subgraph.

- **Induced Subgraph:** The network graph constructed from data where nodes are randomly sampled and where edges are included only if both nodes are randomly sampled are known as induced subgraph.

- **Star Subgraph:** The network constructed from data where nodes are randomly sampled and all their edges are included, regardless of whether the incident nodes are sampled (*i.e.* if i is randomly sampled, the edge ij will be included regardless of whether or not j is sampled), is called a star subgraph.

- **Network Motif:** Any subgraph of the network which has a particular structure. For example, the reciprocated link motif is defined as any pair of nodes, $\{i, j\}$, such that both of the possible directed links between them, $\{ij, ji\}$, are present in the subgraph. Another example is the k-star motif, which is defined as any k nodes such that one of the nodes is linked to all (k-1) other nodes, and the other nodes are not linked to each other.

- **Isomorphic Networks:** Two networks are isomorphic iff we can move from one to the other only by permuting the node labels. For example, all six directed networks composed of three nodes and one edge are isomorphic. Isomorphism implies that all network statistics are also identical, since these statistics are measured at a network level so are not affected by node labels.



(a) Bipartite Network (b) Uniform Random (c) Scale-free (d) Core-periphery

(e) Cellular (f) Small world (g) K-Star (h) Clique

Figure 2.7: *Network Topologies*

2.7.2 Quadratic Assignment Procedure

The Quadratic Assignment Procedure (QAP) was developed originally by Mantel (1967) and Hubert and Schultz (1976).[105] It tests for correlation between a pair of network variables by calculating the correlation in the data, and comparing this to the range of estimates computed from the same calculation after permutation of the rows and columns of the adjacency matrix $\boldsymbol{G}$. For example, suppose we have two vectors $\mathbf{y}(\boldsymbol{G}) = \{y_i(\boldsymbol{G}_g)\}_{i \in \mathcal{N}_g}$ and $\mathbf{x}(\boldsymbol{G}) = \{x_i(\boldsymbol{G}_g)\}_{i \in \mathcal{N}_g}$ which are functions of the network. We first calculate $\hat{\rho}_{0,YX}$, the correlation between $\mathbf{y}$ and $\mathbf{x}$ observed in the data. In order to respect the dependencies between edges that involve the same node, we then jointly permute the rows and columns of the argument of $\mathbf{y}$. This amounts to effectively relabelling the nodes, so that we calculate a new estimate $\hat{\rho}_{w,YX}$: the correlation between $\mathbf{y}(\boldsymbol{G}_w)$ and $\mathbf{x}(\boldsymbol{G})$, where $\boldsymbol{G}_w$ is the permuted adjacency matrix. It is generally *not* the same as permuting the elements of the vectors $\mathbf{y}$. This is repeated W times, to give a range of estimates $\{\hat{\rho}_{w,YX}\}_{w=1,...,W}$. Under the null hypothesis of no correlation, we can perform, for example, a two-sided test at the 10% level, by considering whether $\hat{\rho}_{0,YX}$ lies between the 5th and 95th percentiles of $\{\hat{\rho}_{w,YX}\}_{w=1,...,W}$. If it does not, we can reject the null at the 10% level.

Ideally one would like to use all the possible permutations available, but typically this number is too large. Hence a random sample of permutations is typically used. This is done by drawing the from the set of nodes of the network, $\{1, ..., N\}$, without replacement. The order in which the indices are drawn is defined as the new, permuted ordering, for calculating $\mathbf{y}(\boldsymbol{G}_w)$.

Krackhardt (1988) extended QAP to a multivariate setting. Now we have variables $\{\mathbf{y}(\boldsymbol{G}), \mathbf{x}_1(\boldsymbol{G}), ..., \mathbf{x}_K(\boldsymbol{G})\}$ and are interested in testing whether there is a statistically significant correlation between $\mathbf{y}$ and the K other variables. To test for a relationship between $\mathbf{y}$ and $\mathbf{x}_1$, Krackhardt suggests we first regress $\mathbf{y}$ and $\mathbf{x}_1$, separately, on $(\mathbf{x}_2 ... \mathbf{x}_K)$ to give residuals $\mathbf{y}_1^*$ and $\mathbf{x}_1^*$. Then one can perform QAP on $\mathbf{y}_1^*$ and $\mathbf{x}_1^*$, as in the bivariate setting, where $\hat{\rho}_{0,Y^*X_1^*}$ is an estimate of the partial correlation between $\mathbf{y}$ and $\mathbf{x}_1$ conditioning on the other $(\mathbf{x}_2 ... \mathbf{x}_K)$. This process can be repeated for all K covariates.

[105] See Hubert (1987) for a review of developments of this method.

Chapter 3

Informal Insurance and Endogenous Poverty Traps[1]

3.1 Introduction

An old literature going back to Rosenstein-Rodan (1943) suggests that the failure of poor economies to develop comes from an inability to coordinate, where multiple simultaneous investments could be profitable, but alone none of these investments will be. However, there are many investments which are profitable even without others' investment, and yet do not take place. For example, in rural villages the purchase of small capital goods such as livestock is typically highly profitable, and yet we see little investment by most households (Bandiera et al., forthcoming; Banerjee et al., 2015b; and also De Mel et al., 2008 in the context of small businesses).

One obvious explanation for the lack of investment is that households are poor, and so neither have the resources to invest nor access to formal credit. However, households regularly borrow from (and lend to) friends and neighbours, using this to smooth consumption (Townsend, 1994; Udry, 1994). The puzzle, then, is why households are able to borrow informally for consumption but not for high return investments? It cannot be explained by a lack of resources (incomes and assets): whilst individual households in these 'risk-sharing networks' have few resources, collectively they have the resources needed for investment. So why don't households pool resources to allow some households to engage in investment?

[1]With thanks to Orazio Attanasio, Richard Blundell, Áureo de Paula, and Imran Rasul for their invaluable support and advice. Thanks also to Marco Bassetto, Martin Cripps, Mariacristina De Nardi, Costas Meghir, Nicola Pavoni, Bryony Reich, seminar participants at Bocconi, Bristol, Geneva, Gothenburg, Manchester, Stockholm School of Economics, Toulouse School of Economics, Universitat Autònoma de Barcelona, UCL, University of Southern California, and Warwick, and conference participants at European Econometric Society Winter Meeting.

This paper offers a new explanation, and empirical evidence, for this puzzle. The key idea is a that investment *reduces* the capacity of investing households to provide informal consumption smoothing. To see this, note that borrowing and lending for consumption smoothing – 'informal insurance' – is sustained by reciprocity: a household lends today because it wants the possibility of borrowing in the future, when it has a low income. Rather than writing formal contracts, borrowing occurs informally, with lenders motivated by loss of future access to borrowing if they do not lend when their incomes are relatively high. What makes borrowing for investment different is that an investing household will on average be better off in the future. Having investment income as well as labour income will reduce its need to borrow for consumption smoothing in future periods.[2] This reduced need for borrowing limits the amount it can be asked to lend — ask for too much and the household would rather just lose access to future insurance. The reduced capacity to provide other households with consumption smoothing prevents the other households from lending for investment.

In this paper I first develop a formal theoretical model that captures this mechanism, and then provide empirical evidence from a large scale randomised controlled trial (RCT). The model combines the key elements discussed above: informal insurance with limited commitment and lumpy (indivisible) investment. I show that the mechanism described, trading off insurance and investment, can lead to a *network-level* poverty trap i.e. the long run equilibrium level of income in the network will depend on the initial conditions. I also develop additional comparative static predictions specific to the frictions – limited commitment and lumpy investment – in my model. I then provide empirical evidence of the relevance of this mechanism. Using data from a large scale, long term randomised control trial in Bangladesh, I find evidence that networks in Bangladesh are indeed in a network-level poverty trap. I verify comparative static predictions of the model, in terms of both income inequality and network size. This provides additional support for my proposed mechanism, and allows me to rule out alternative competing hypothesis as the source of the network-level poverty trap.

More precisely, I develop a model which captures four important characteristics of households in village economies: (i) households are risk-averse and have volatile incomes; (ii) they are able to engage in consumption smoothing by making inter-household transfers; (iii) households have *limited commitment* in their risk-sharing arrangement i.e. at any point in time, the expected value of continuing any risk sharing must be at least as good as the value of walking away forever ('autarky'); (iv) households have the opportunity each period to invest in a 'lumpy' (indivisible) asset. The first three characteristics lead to models of risk sharing with (dynamic) limited commitment, as studied by Kocherlakota (1996) and Ligon et al. (2002). The fourth character-

[2] I discuss later conditions on the variance of investment returns. In my empirical context this will be lower than the variance of labour income, and the correlation between income from these sources is low.

istic has also been studied in a number of development contexts (Rosenzweig and Wolpin, 1993; Fafchamps and Pender, 1997; Munshi and Rosenzweig, 2016). My innovation is to combine these standard features and show that there are important interactions betweenthem, which can provide an explanation for the long-standing puzzle of underinvestment.

Analysis of the model provides three main findings. First, a poverty trap naturally arises in this model: the long run equilibrium income distribution depends on the initial level of capital invested. The 'depth' of this trap – the amount of income the network needs to escape the trap – is greater if commitment is limited, when no household can afford to invest in autarky. Second, investment has an inverted-U shape in income inequality. Third, investment becomes easier as network sizes increase. The latter two comparative static predictions are specific to the mechanism of my model, and are testable, so can be used to distinguish my explanation of a network-based poverty trap from alternative hypotheses such as coordination failure.

Formally, a network poverty trap exists if there is some level of aggregate income, such that equilibrium investment is different for networks whose maximum possible income is above or below this 'threshold'. Networks below the threshold will never have enough income to make even the first investment, and so remain persistently poor. Above the threshold, it will be possible for networks to initiate some investment. This raises future income, ensuring further investments are possible, and allowing all households to eventually invest. However, with only a 'small push' that provides some initial capital, the economy can be set on a path of further investment and income growth. Unlike so-called 'big push' models (Rosenstein-Rodan, 1943, Murphy et al., 1989), here it is not necessary for all households to be simultaneously coordinated in investment, nor is coordination alone – without the provision of assets – sufficient to generate further investment.

When no household has enough income to want to invest if in autarky, limited commitment reduces investment. Limited commitment makes resource pooling more difficult, hence investing households will only be able to credibly promise smaller transfers. A larger share of investment must therefore come out of their own pocket. This effect is important in explaining the puzzle with which I began. With full commitment, only networks with too few resources for investment would be in the poverty trap. Empirically many networks have resources that they do not use for investment, despite the high returns. Limited commitment risk sharing provides the necessary friction to explain why investment may not occur in these cases.

Under full commitment, the distribution of income would have no bearing on investment: all that matters is aggregate income. With limited commitment, households who receive temporarily high incomes might be better off leaving the initial insurance arrangement. This leads to a renegotiation of the arrangement, which might involve allowing them to invest. As inequality rises, it changes which households are better off from leaving the arrangement, unless it is renegotiated. I show that the changes in who is better off from leaving lead to rising inequality having an inverted-U shaped effect on investment.

Increasing network size will increase the amount of investment. This occurs for two reasons. First, aggregate income rises, providing more resources potentially available for investment. Second, as the risk-sharing pool grows, the quality of insurance is improved, increasing the opportunity cost of autarky. This is in contrast to models with coordination failure, where investment becomes more difficult as network size rises.

I verify these three findings using data from a large scale, long term randomised control trial (RCT) in Bangladesh. These data cover 27,000 households across 1,400 villages in the poorest districts of rural Bangladesh. They were collected as part of an asset transfer program by microfinance organisation BRAC. The intervention randomised villages into either treated or control status, and then provided assets (typically cows) to the poorest households in treated villages. These transfers were worth more than 50% of median income for the households that received them. Asset transfers took place after data collection in 2007, and follow-up surveys were carried out in 2009 and 2011. The program was evaluated by Bandiera et al. (forthcoming), who show that the program has large and sustained effects on both earnings and asset accumulation.

Four features of the data make them suitable for my context. First, the data are from one of the largest scale RCTs in a developing country, encompassing a large cross-section of networks, from more than 1,400 villages. This is important since my model predictions are at the network level. Second, in a subsample of my data, the data record the exact links used for risk-sharing transfers. I exploit this to construct a good proxy for the appropriate risk-sharing network in the full dataset, which recent work suggests it is important to measure well when studying risk sharing (Mazzocco and Saini, 2012). Third, the program provided large injections of lumpy capital (cows), with significant variation in the number of transfers across villages. This provides the exogenous variation in aggregate income necessary for my test of a network-level poverty trap. Finally, the data cover a long time scale, with a follow-up survey four years after the initial capital injection. This provides a large enough window to study how the initial injection affects *additional* investment, which is key to understanding whether a network has left the poverty trap.

The main empirical finding are as follows. First, aggregate investment in cows by risk-sharing networks between 2009 and 2011 is zero on average if the network received less than $\$3,500$ (PPP 2007) of capital from the program, 7% of median network income. This threshold is determined using a formal statistical test for a structural break with unknown break point i.e. a test for a change in the slope of additional investment with respect to the capital provided when the location of the slope change being unknown. Above this level, aggregate investment is increasing (and linear) in the aggregate amount of capital provided by the program. Second, I show that investment has an inverted-U shape in income inequality, with a third of networks having 'too much' inequality in terms of the effect on investment. Third, I show that investment is increasing in network size, and provide tentative evidence that this is caused by a shift in the location of the threshold. On average five additional households, a 10% increase in network size, are needed for one additional investment to be possible. These qualitative patterns are precisely as predicted by the model, and together they cannot be rationalised by any existing alternative mechanism.

This paper contributes to three distinct strands of literature. First, it contributes to the large literature on poverty traps. Whilst poverty traps are an old idea, empirical work has failed to find convincing evidence for any of the specific mechanisms that have been proposed.[3] The novel aspect of my model is that by introducing risk sharing, the poverty trap occurs at the network level, and by introducing limited commitment, risk-sharing networks with enough resources to invest might still choose not to. I use standard tools – non-convexity in production, of which lumpiness is a particular example, and a financial friction (limited commitment) – to generate the poverty trap (see for example Banerjee and Newman, 1993; Aghion and Bolton, 1997; and Ghatak, 2015). However, by embedding these in a risk-sharing framework, the poverty trap in my model occurs at the network level. I provide empirical evidence that we do indeed see a trap at this level, and my results are not consistent with a story of individual level traps. My mechanism is distinct from the group level poverty traps of Rosenstein-Rodan (1943) and Murphy et al. (1989), which are purely due to coordination failure. I provide evidence that allows me to rule out poverty trap models that rely on increasing returns to coordinated investment, including due to externalities, fixed costs, or learning.

Second, I contribute to the literature on risk sharing with frictions (Kocherlakota, 1996; Ligon et al., 2002; among others). In particular, there is a growing literature examining how endogenously incomplete insurance affects and is affected by opportunities in other markets (Attanasio and Rios-Rull, 2000; Attanasio and Pavoni, 2011; Ábrahám and Cárceles-Poveda, 2009; Ábrahám and Laczó, 2014; Morten, 2015). Attanasio and Pavoni (2011) highlight an important trade-off between using insurance and using (continuous) investment to provide consumption smoothing. A similar trade-off is present in my model, but the 'lumpiness' of investment in

[3]The main approach to testing for a poverty trap is to measure whether the elasticity of tomorrow's income with respect to today's income, via some channel, is greater than one. For example, the 'nutrition' poverty trap suggests that increased income would improve individual's nutrition, which increases their capacity to work and allows them to earn more. The test is then whether the product of the elasticity of nutrition with respect to income and elasticity of income with respect to nutrition is greater than one. Subramanian and Deaton (1996) estimate an elasticity of nutrition with repect to income of no more than .5, while Strauss (1986) estimates an elasiticity of income with respect to nutrition of .33: the product of these is far less than one. Estimated elasticities for other channels are also low. From Cohen, Dehejia and Romanov (2013) and Rosenzweig and Zhang (2009), the elasticity of child's income with respect to schooling would need to be greater than 33 to generate a demographic/education poverty trap.

my context (mirrored by many development applications) changes the nature of the decision-making, and creates the possibility of a poverty trap. Morten (2015) also considers a model with risk sharing and a binary decision, but where the decision only directly affects payoffs today. By contrast, in this paper investment has permanent effects on the distribution of income, allowing me to study questions of longer term development and growth. It also opens the door for the study of other long term discrete investment decisions, such as irrigation (Rosenzweig and Wolpin, 1993; Fafchamps and Pender, 1997), education (Angelucci et al., 2015), and permanent migration (Munshi and Rosenzweig, 2016), in the context of risk sharing with limited commitment.

Third, I contribute directly to the recent and growing work on asset transfer programs (Bandiera et al., forthcoming; Banerjee et al., 2015b; De Mel et al., 2008; Morduch et al., 2015). These studies find that in many cases, across a range of countries and contexts, asset transfer programs are very successful in increasing incomes. My paper provides a possible explanation for why such one-off transfers of assets appear to have larger effects on income growth than smaller, but longer term, cash transfer programs such as Progresa (Ikegami et al., 2016). Small increases in income will still be partly smoothed away, rather than providing the basis needed for additional investment. It also suggests a route for increasing the impact of these interventions: targeting at a network rather than a household level. By providing enough resources at an aggregate level, these programs can provide the 'small push' that networks need to get out of the poverty trap. My results highlight how a budget neutral redistribution of asset transfers across networks can increase additional investment. Restructuring the existing policy in this way could have increased additional investment four years after the program by 44%, relative to using household-level targeting.

The next section develops the model formally, and provides the theoretical results. Section 3.3 describes the data and context for my empirical work. Section 3.4 tests the key predictions of the model, and provides additional supportive evidence for the mechanism proposed. The final section concludes.

3.2 A Model of Insurance, Investment, and a Poverty Trap

Consider an infinite-horizon economy composed of N households. Households have increasing concave utility functions defined on consumption that satisfy the Inada conditions. They also have a common geometric discount rate, β.[4] Each period t, households receive endowment income $\mathbf{y}_t = \{y_t^1, \dots, y_t^N\}$ drawn from some (continuous) joint distribution $\mathcal{Y}$. Individual incomes are bounded away from 0, and aggregate income $Y_t := \sum_{i=1}^{N} y_t^i$ is bounded above by $Y^{\max}$. Income draws are assumed to be iid over time, but may be correlated across households within a period. I define $s_t^i := y_t^i / Y_t$ as household i's share of aggregate endowment income in period t. To ease notation, hereafter I suppress the dependence of variables on t.

The households belong to a single network, and they may choose to engage in risk sharing. Since households are risk-averse, and endowment incomes are risky, there is scope for mutually beneficial risk sharing. In particular, an informal agreement in which households with good income shocks in any period make transfers to those with bad income shocks will improve the expected discounted utility for all households. I model this risk sharing as net transfers, τ^i, made by households $i = 2, \dots, N$ to household 1.[5] Household consumption will then be $c^i = y^i - \tau^i$, and $\tau^1 \equiv \sum_{i=2}^{N} \tau^i$, where I suppress the dependence of all these objects on the shock $\mathbf{y}$ to ease notation.

An impediment to risk sharing is the presence of dynamic limited commitment (Kocherlakota, 1996; Ligon et al., 2002). Households may, in any period, choose to walk away from the arrangement, keeping all of their income that period and then being excluded from the arrangement thereafter. This will limit the amount of risk sharing that can take place.

Thus far, the model is an N household, continuous shocks version of the standard model of risk sharing with dynamic limited commitment. To this problem I introduce the possibility that households may engage in lumpy investment. Precisely, each period a household may choose whether or not to invest in a binary investment, κ. This has a one-off cost d, and pays a guaranteed return of R in all future periods.[6] Investment is an absorbing state, and households may hold at most one investment.[7] Additionally, investments must be held by the household that does the investment, although transfers may be made out of investment income. This rules

[4] For work considering risk sharing with heterogeneous preferences, see for example Mazzocco and Saini (2012). For work considering poverty traps with non-geometric discounting, see Banerjee and Mullainathan (2010) and Bernheim et al. (2015).

[5] In principle, each household could choose how much income to transfer to each other household. Since my interest is only on the total risk sharing that takes place, and not on the precise structure of transfers that are used, I model all transfers as going to or from household 1. For each household there is then a single decision about the net transfers to make (or receive). For work studying how network *structure* and risk sharing interact, see Ambrus et al. (2014) and Ambrus et al. (2015).

[6] There are two implicit assumptions here. First, the return does not vary with the number of investments that occur. This rules out both general equilibrium effects, where we might expect to see the return decline as the number of investments increases, and fixed costs, where we might expect to see the return increase. I will show later that in my empirical setting, these are both reasonable. Second, there is assumed to be no risk in the return on investment. This is done to distinguish my mechanism from an alternative mechanism, where a high return activity is also higher risk, so underinvestment occurs because of a *lack* of insurance (see for example Karlan et al., 2014). It is also appropriate to my context: as I document below, in my empirical setting, investment income will be less risky than non-investment income.

[7] The former is a simplifying assumption, which could be relaxed at the cost of adding more moving parts to the model. The latter could also be relaxed: all that is needed is some upper bound on the total number of investments a household can hold. This is reasonable in my context, where investments are in livestock: Shaban (1987) and Foster and Rosenzweig (1994) describe how moral hazard issues can limit the ability to hire labour from outside the family to manage livestock.

out cooperatives and other joint investment structures.[8] Now an uninvested household must choose each period what net transfers to make, τ^i, and whether to invest, $\Delta\kappa^i$.

Barring risk sharing and investment, no alternative forms of smoothing are permitted. This rules out external borrowing: whilst a household may engage in implicit borrowing from other households in the risk-sharing network, the network as a whole cannot borrow from the wider world. I will show that limited commitment problems make borrowing *within* the network difficult, even amongst households that interact regularly, so one would expect this problem to be even more severe for lenders from outside the community. I also rule out saving, so that investment is the only vehicle for transfering resources over time.[9] If private savings were introduced, they would provide an alternative means of transferring resources over time. In the model, a household would give or receive transfers, and then decide what share of resources (if any) to save. However, in the next period, the planner wants again to smooth consumption, and will look at the total cash-on-hand (income plus savings) that a household has in determining transfers. Hence a household that saves would effectively be 'taxed' on this saving, as it would be considered in the same way as any other income when the next period begins. This idea of a 'network tax' discouraging saving has been documented by Dupas and Robinson (2013), who show that poor households appear to have negative *nominal* returns to saving, and Jakiela and Ozier (2016), who show that households are willing to pay to prevent information about good income shocks being revealed.[10]

3.2.1 Risk Sharing under Limited Commitment without Investment

I first consider the limited commitment problem when all households have already invested. In this case there is no investment decision to make, and the problem has the same form as the many household version of Ligon et al. (2002), but with continuous income shocks. A solution to the model will provide a mapping from the complete history of income shocks, to the transfers that a household makes or receives today.

To find this solution, I first use the standard technique of writing the sequential problem i.e the choice of transfers in a given period conditional on the complete history of shocks, in a recursive formulation. Following Spear and Srivastava (1987) and Abreu et al. (1990), this simplifies the problem by encoding the dependence on the entire history into a single state variable, 'promised utility', ω, which summarises the relevant information.

[8] For a model with joint ownership of investment, see Thomas and Worrall (2016).

[9] For work studying limited commitment risk sharing with divisible saving, see for example Ligon et al. (2000) and Ábrahám and Laczó (2014).

[10] Allowing for hidden savings would complicate this argument slightly, but as long as investment cannot be hidden – which is likely in many contexts, such as when the investments are livestock – any systematic hiding of savings for investment purposes would be detectable and punishable once investment takes place.

I then take the usual approach (as in Ligon et al., 2002) of formulating the problem as a *planner's problem.* Without loss of generality, I assume household 1 is the hypothetical planner. Its role will be to choose the transfers that each household should make at each possible history, and provide promises of utility, in a way that meets certain constraints (described below).[11]

At any point in time, the planner's problem will then be to maximise its own utility, denoted by the value function $V(\mathbf{y}; \boldsymbol{\omega}(\mathbf{y}), \mathbf{1})$. This value function depends on the realised incomes, $\mathbf{y}$; the utility levels the planner promised to provide given the incomes, $\boldsymbol{\omega}(\mathbf{y}) = \{\omega^2(\mathbf{y}), \ldots, \omega^N(\mathbf{y})\}$; and the stock of investment, $\boldsymbol{\kappa}$, which here is equal to $\mathbf{1}$. The choices the planner makes are what transfers to ask each household to make today, $\boldsymbol{\tau}(\mathbf{y}) = \{\tau^2(\mathbf{y}), \ldots, \tau^N(\mathbf{y})\}$; what promises of expected utility to make for tomorrow, $\bar{\boldsymbol{\omega}}'(\mathbf{y}) = \{\bar{\omega}'^2(\mathbf{y}), \ldots, \bar{\omega}'^N(\mathbf{y})\}$; and how to deliver these promises, $\boldsymbol{\omega}'(\mathbf{y}, \mathbf{y}') = \{\omega'^2(\mathbf{y}, \mathbf{y}'), \ldots, \omega'^N(\mathbf{y}, \mathbf{y}')\}$ $\forall \mathbf{y}'$. The notation $'$ denotes that a variable relates to tomorrow.

So the planner's problem can be written as:

$$\max_{\left\{\tau^i(\mathbf{y}), \bar{\omega}'^i(\mathbf{y}), \{\omega'^i(\mathbf{y}, \mathbf{y}')\}_{\mathbf{y}'}\right\}_{i=2}^N} u\left(y^1 + R + \sum_{i=2}^N \tau^i(\mathbf{y})\right) + \beta \bar{V}(\bar{\boldsymbol{\omega}}'(\mathbf{y}), \mathbf{1}) \tag{3.1}$$

where

$$\bar{V}(\bar{\boldsymbol{\omega}}'(\mathbf{y}), \mathbf{1}) = \int V(t; \boldsymbol{\omega}'(\mathbf{y}, t), \mathbf{1}) \, \mathrm{d}F_{\mathbf{Y}'}(t) \tag{3.2}$$

denotes the expected *continuation value* for the planner when he has promised an expected utility of $\bar{\boldsymbol{\omega}}'(\mathbf{y})$ given current state $\mathbf{y}$, subject to three sets of constraints. Promised expected utility is defined as:

$$\bar{\omega}'^i(\mathbf{y}) = \int \omega'^i(\mathbf{y}, t) \, \mathrm{d}F_{\mathbf{Y}'}(t) \tag{3.3}$$

The first set of constraints, with multipliers $\lambda^i(\mathbf{y})$, are the *promise keeping constraints*:

$$[\lambda^i(\mathbf{y})] \qquad u(y^i + R - \tau^i(\mathbf{y})) + \beta \bar{\omega}'^i(\mathbf{y}) \geq \omega^i(\mathbf{y}) \qquad \forall i \in \{2, \ldots, N\} \tag{3.4}$$

These require that, at every possible realisation of income, $\mathbf{y}$, the planner actually provides (at least) the promised utility that he agreed to provide. The second set of constraints, with

[11] This will find an equilibrium set of contingent transfers (transfers that depend on the realised history) that is subgame perfect: no household would like to unilaterally deviate from the arrangement in any realised state of the world. However, a 'decentralised' approach, where one directly solved the repeated game representation, would generically have many possible equilibria, from which my approach will select a single one. For work studying the decentralisation problem, see Alvarez and Jermann (2000) and Ábrahám and Cárceles-Poveda (2009).

multipliers $\phi^i(\mathbf{y})$, are the limited commitment constraints:

$$[\phi^1(\mathbf{y})] \qquad u\big(y^1 + R + \sum_{i=2}^{N} \tau^i(\mathbf{y})\big) + \beta \bar{V}(\bar{\boldsymbol{\omega}}'(\mathbf{y}), \mathbf{1}) \geq \Omega(y^1, 1) \tag{3.5}$$

$$[\phi^i(\mathbf{y})] \qquad u(y^i + R - \tau^i(\mathbf{y})) + \beta \bar{\omega}'^i(\mathbf{y}) \geq \Omega(y^i, 1) \qquad \forall i \in \{2, \ldots, N\} \tag{3.6}$$

which require that each household (including the planner) gets at least as much expected discounted utility from the insurance arrangement as it would get if it walked away and took its outside option, $\Omega(\cdot)$. The outside option is a function of current income and current investment status, and for an invested household is calculated as the utility of consuming all its income today, and then the discounted expected utility given that it never again has insurance.[12] Formally:

$$\Omega(y^i, 1) := u(y^i + R) + \frac{\beta}{1-\beta} \int u(y' + R) \, \mathrm{d}F(y') \qquad \forall i \in \{1, \ldots, N\} \tag{3.7}$$

The third set of constraints, with multiplier $\beta\nu^i(\mathbf{y})$ is that for each household $i \in \{2, \ldots, N\}$ the planner must find some promise of utility for every possible income realisation, such that the average utility provided across all states is equal to the promised expected utility:

$$[\beta\nu^i(\mathbf{y})] \qquad \bar{\omega}'^i(\mathbf{y}) = \int \omega'^i(\mathbf{y}, t) \, \mathrm{d}F_{\mathbf{Y}'}(t) \qquad \forall i \in \{2, \ldots, N\} \tag{3.8}$$

The setup thus far is just the natural extension of Ligon et al. (2002) to the case with continuous shocks, plus the introduction of the "intermediate" variable, $\bar{\omega}^i$, which denotes the promised expected utility to household i. If I were to substitute the expression for $\bar{V}(\cdot)$ from Equation 3.2 in to the problem, and similarly for $\bar{\omega}'$ from Equation 3.8, the choice variables would be only transfers today and utility promises in each future state. This is as in Ligon et al. (2002), and the solution could be derived by using the first order conditions and by applying envelope theorem to this problem.

An alternative approach, which I pursue, is to note that the problem is separable: $\omega'^i(y')$ appears only in the definitions of $\bar{V}$ and $\bar{\omega}'^i$. Hence one can divide the problem into an "inner" part, which solves the allocation of utility across future states of the world given promised levels of expected utility, and an "outer" part which finds optimal transfers today and expected utility promises for tomorrow, given the shock today and that expected utility will be provided efficiently. This split is simply an application of Bellman's Principle of Optimality.

[12] This is the most extreme punishment that can be imposed on the household, without assuming there are also exogenous costs of relationship loss. It can therefore support the maximum amount of risk sharing. Weaker punishment strategies would provide additional, Pareto-dominated equilibria. I focus on a Pareto efficient insurance arrangement.

The inner problem studies how a given level of promised expected utility, $\bar{\boldsymbol{\omega}}'$, should be provided. Let $\mathcal{U}(\bar{\boldsymbol{\omega}}', \mathbf{1})$ denote the value function for a planner who has to provide promised expected utility $\bar{\boldsymbol{\omega}}'$, and can choose how this is delivered by selecting the utility to be delivered in each state of the world, $\boldsymbol{\omega}'(\mathbf{y}', \bar{\boldsymbol{\omega}}')$. $V(t; \tilde{\boldsymbol{\omega}}', \mathbf{1})$ denotes the continuation value of promising to deliver $\tilde{\boldsymbol{\omega}}'$ given the state is t.

Then the inner problem is:

$$\mathcal{U}(\bar{\boldsymbol{\omega}}', \mathbf{1}) = \max_{\{\omega'^i(\mathbf{y}'; \bar{\boldsymbol{\omega}}')\}_{i,\mathbf{y}'}} \int V(t; \boldsymbol{\omega}'(t; \bar{\boldsymbol{\omega}}'), \mathbf{1}) \, \mathrm{d}F_{\mathbf{Y}'}(t) \tag{3.9}$$

$$= \int \max_{\{\omega'^i(\mathbf{y}'; \bar{\boldsymbol{\omega}}')\}_{i,\mathbf{y}'}} V(t; \boldsymbol{\omega}'(t; \bar{\boldsymbol{\omega}}'), \mathbf{1}) \, \mathrm{d}F_{\mathbf{Y}'}(t) \tag{3.10}$$

s.t.

$$[\tilde{\nu}^i] \qquad \bar{\omega}'^i = \int \omega'^i(t; \bar{\boldsymbol{\omega}}') \, \mathrm{d}F_{\mathbf{Y}'}(t) \qquad \forall i \in \{2, \ldots, N\} \tag{3.11}$$

Appendix 3.7.1 provides a (heuristic) proof that one can move from 3.9 to 3.10.

Now the expected continuation value, $\mathcal{U}$, is defined as the integral over the continuation value in each possible realisation of the shock, $\mathbf{y}'$, where the planner can choose what utility to promise at each possible shock, subject only to these promised utilities integrating to the promised expected utility, $\bar{\boldsymbol{\omega}}'$.

The first order conditions and envelope condition for this problem are:

$$[\mathrm{FOC}(\omega'^i(\mathbf{y}'; \bar{\boldsymbol{\omega}}'))] \qquad \frac{\partial V(t; \boldsymbol{\omega}'(\mathbf{y}'; \bar{\boldsymbol{\omega}}'), \mathbf{1})}{\partial \omega'^i(\mathbf{y}'; \bar{\boldsymbol{\omega}}')} = \tilde{\nu}^i \tag{3.12}$$

$$[\mathrm{ET}(\bar{\omega}'^i)] \qquad \frac{\partial \mathcal{U}(\bar{\boldsymbol{\omega}}', \mathbf{1})}{\partial \bar{\omega}'^i} = \tilde{\nu}^i \tag{3.13}$$

Combining these, one gets that:

$$\frac{\partial \mathcal{U}(\bar{\boldsymbol{\omega}}', \mathbf{1})}{\partial \bar{\omega}'^i} = \frac{\partial V(\mathbf{y}'; \boldsymbol{\omega}'(\mathbf{y}'; \bar{\boldsymbol{\omega}}'), \mathbf{1})}{\partial \omega'^i(\mathbf{y}'; \bar{\boldsymbol{\omega}}')} \tag{3.14}$$

Then the "outer" problem is to choose transfers, $\boldsymbol{\tau}(\mathbf{y})$, and promised expected utilities, $\bar{\boldsymbol{\omega}}'(\mathbf{y})$, given that this promised expected utility will be delivered efficiently as in the inner problem. This alternative approach is just a rewriting of the original problem, and so gives an identical solution. However, as will be seen, this separation of the problem will allow the problem to be solved even when discrete choices and discrete state variables are introduced.[13]

[13] Although there is in principle already the discrete state variable of investment included here, with all households already invested it can never change.

For now, with this rewriting, the full problem is to maximise:

$$\max_{\{\tau^i(\mathbf{y}),\bar{\omega}'^i(\mathbf{y})\}_{i=2}^N} u\left(y^1 + R + \sum_{i=2}^N \tau^i(\mathbf{y})\right) + \beta\mathcal{U}(\bar{\boldsymbol{\omega}}'(\mathbf{y}), \mathbf{1}) \tag{3.15}$$

with respect to only transfers and promised expected utilities, subject to the constraints in Equations 3.4, 3.5, and 3.6, and with $\mathcal{U}(\cdot)$ defined as in Equation 3.10.

Taking first order conditions, using the envelope theorem for $\omega^i(\mathbf{y})$, this gives for $i \in \{2, \ldots, N\}$:

$$[\text{FOC}(\tau^i(\mathbf{y}))] \qquad \frac{\mathrm{d}u(c^1(\mathbf{y}))/\,\mathrm{d}\tau^i(\mathbf{y})}{\mathrm{d}u(c^i(\mathbf{y}))/\,\mathrm{d}\tau^i(\mathbf{y})} = \frac{\lambda^i(\mathbf{y}) + \phi^i(\mathbf{y})}{1 + \phi^1(\mathbf{y})} \tag{3.16}$$

$$[\text{FOC}(\bar{\omega}'^i(\mathbf{y}))] \qquad \frac{\partial\mathcal{U}(\bar{\boldsymbol{\omega}}'(\mathbf{y}), \mathbf{1})}{\partial\bar{\omega}'^i(\mathbf{y})} = -\frac{\lambda^i(\mathbf{y}) + \phi^i(\mathbf{y})}{1 + \phi^1(\mathbf{y})} \tag{3.17}$$

$$[\text{ET}(\omega^i(\mathbf{y}))] \qquad \frac{\partial V(\mathbf{y}; \boldsymbol{\omega}(\mathbf{y}), \mathbf{1})}{\partial\omega^i(\mathbf{y})} = -\lambda^i(\mathbf{y}) \tag{3.18}$$

Hence:

$$\frac{\partial\mathcal{U}(\bar{\boldsymbol{\omega}}'(\mathbf{y}), \mathbf{1})}{\partial\bar{\omega}'^i(\mathbf{y})} = -\frac{-\frac{\partial V(\mathbf{y};\boldsymbol{\omega}(\mathbf{y}),\mathbf{1})}{\partial\omega^i(\mathbf{y})}\phi^i(\mathbf{y})}{1 + \phi^1(\mathbf{y})} = -\frac{\mathrm{d}u(c^1(\mathbf{y}))/\,\mathrm{d}\tau^i(\mathbf{y})}{\mathrm{d}u(c^i(\mathbf{y}))/\,\mathrm{d}\tau^i(\mathbf{y})} \tag{3.19}$$

From the envelope theorem (Equation 3.18) it can be seen that the value function is decreasing in the promised utility $\omega^i(\mathbf{y})$ to each household i. When none of the limited commitment constraints bind, $\phi^1(\mathbf{y}) = \phi^i(\mathbf{y}) = 0$, the slope of the value function (the ratio of marginal utilities for i and 1) remains constant, and so the ratio of marginal utilities remain unchanged from the previous period. When a household's limited commitment constraint binds, the ratio of marginal utilities in that period and all future periods (until another constraint binds), is increasesd so that it receives an increased share of consumption.

3.2.2 Risk Sharing under Limited Commitment with Investment

Next I consider the case where $k < N$ investments have already been made, $k = \sum_{i=1}^N \kappa^i$. Now there is a meaningful investment decision for the planner, which is the chief innovation of the model. Precisely, the planner now has to choose the optimal number (and allocation) of investments $\Delta k(\mathbf{y}) \in \{1, \ldots, N-k\}$, as well as transfers and utility promises.[14] I first note that there is a weakly dominant allocation rule for assigning investments.

[14] In a full commitment setting it would not matter which households 'held' the investments, since the planner could always require them to make arbitrary transfers. With limited commitment this is no longer the case: if the planner requires too high a transfer, the household may prefer autarky.

Lemma 3.1. *There exists a unique weakly dominant investment allocation rule. Let* $\tilde{\omega}^i(\mathbf{y}) := \max\{\omega^i(\mathbf{y}), \Omega(y^i, 0)\}$*. Then if* $\Delta k(\mathbf{y})$ *investments are to occur, assign the investments to the* $\Delta k(\mathbf{y})$ uninvested *households with the highest values of* $\tilde{\omega}^i(\mathbf{y})$*.*

Proof. See Subsection 3.7.2. □

The planner's problem can therefore be simplified to choose only what transfers to make and *how many* investments to do, taking as given which households will be asked to do the investments. This reduces significantly the dimensionality of the choice problem, from $(N-k)!$ possible values for the discrete choice, to only $N-k$ values.

I next simplify the problem further, making use of additional separability in the structure of the problem. The planner's decision can be separated into first choosing what transfers to make *given a decision on the number of investments*, and then choosing the optimal number of investments. This follows from an application of Bellman's Principle of Optimality. So the planner's value function, given the shock, $\mathbf{y}$, the promised utility at that shock, $\boldsymbol{\omega}(\mathbf{y})$, and the existing distribution of investments, $\boldsymbol{\kappa}$, is:

$$V(\mathbf{y}; \boldsymbol{\omega}(\mathbf{y}), \boldsymbol{\kappa}) = \max_{\Delta k}\{V_{\Delta k}(\mathbf{y}; \boldsymbol{\omega}(\mathbf{y}), \boldsymbol{\kappa})\} \tag{3.20}$$

where $V_{\Delta k}(\mathbf{y}; \boldsymbol{\omega}(\mathbf{y}), \boldsymbol{\kappa})$ is the conditional value function when the planner requires Δk investments to occur (and be assigned as above), and chooses transfers optimally.

Before defining the planner's problem for the conditional value function, I define the expected continuation value, $\mathcal{U}$, when investment is possible:

$$\begin{aligned}\mathcal{U}(\bar{\boldsymbol{\omega}}', \boldsymbol{\kappa}) &= \int \max_{\{\Delta k(\mathbf{y}'; \bar{\boldsymbol{\omega}}'), \{\omega'^i(\mathbf{y}'; \bar{\boldsymbol{\omega}}')\}_i\}_{\mathbf{y}'}} V_{\Delta k}\left(t; \boldsymbol{\omega}'(t; \bar{\boldsymbol{\omega}}'), \boldsymbol{\kappa}'(t; \bar{\boldsymbol{\omega}}')\right) \mathrm{d}F_{\mathbf{Y}'}(t) \qquad (3.21)\\ \text{s.t. } \bar{\boldsymbol{\omega}}' &= \int \boldsymbol{\omega}'(t, \bar{\boldsymbol{\omega}}')\, \mathrm{d}F_{\mathbf{Y}'}(t), \qquad \Delta k \in \{0, 1, \ldots, N-k\}\end{aligned}$$

Now the benefit of writing the problem in terms of promised expected utilities can be seen. The expected value function is clearly differentiable with respect to $\bar{\omega}'^i$ $\forall i$, with the derivative equal to the value of the multiplier on the integral constraint for promised utilities.

That the expected value function should remain differentiable is not obvious. The discrete choice, Δk, introduces kinks into the value function defined in Equation 3.20. At the point where two conditional value functions cross (in $\omega^i(\mathbf{y})$ space), their slopes will be different. As the upper envelope of these conditional value functions, the overall value function will not be differentiable at these crossing points. With the standard approach to writing the problem, in terms of the promised utility at every state, it is not clear that the expected value function will be differentiable with respect to these promised utilities. However, with this redefinition of the problem, it is immediate that the value function will be differentiable with respect to promised expected utility.

The intuition for why this redefinition can be used to solve the problem of kinks in the value function comes from Prescott and Townsend (1984b, 1984a).[15] They model the allocation of resources in settings with moral hazard. Moral hazard introduces non-convexity into the set of feasible allocations, similar to the problem caused by kinks in my model. They show that, with a continuum of agents, they can solve the problem by introducing 'extrinsic uncertainty': randomness which has no bearing on economic fundamentals, but is nevertheless used in the allocation of resources conditional on all observables. More simply, they introduce lotteries which mean that, in some states (realised incomes in my model), observationally equivalent agents might receive different levels of resources. This 'convexifies' the problem, smoothing out any kinks. It works because the share of agents receiving a particular bundle of resources can be adjusted continuously, even when the bundles differ discretely.

In my context such extrinsic uncertainty is not needed. Randomness in the distribution of income shocks can be used instead to 'smooth out' the kinks. This is what Equation 3.21 is doing: by first choosing $\Delta k(\mathbf{y}'; \bar{\boldsymbol{\omega}}')$ and $\omega'^i(\mathbf{y}'; \bar{\boldsymbol{\omega}}')$ $\forall i, \mathbf{y}'$, and then integrating over the continuum of income shocks, the upper envelope function $\mathcal{U}(\cdot)$ is made convex in promised expected utility.[16]

Having rewritten the expected continuation value in this way, I can now set up the planner's problem with investment. The planner's value function, $V(\mathbf{y}; \boldsymbol{\omega}(\mathbf{y}), \boldsymbol{\kappa})$, is defined as in Equation 3.20 as the maximum over a set of conditional value functions, each for a different fixed number of investments. Given also the definition of the expected continuation value from Equation 3.21, these conditional value functions, $V_{\Delta k}(\mathbf{y}; \boldsymbol{\omega}(\mathbf{y}), \boldsymbol{\kappa})$, are given by:

$$V_{\Delta k}(\mathbf{y}; \boldsymbol{\omega}(\mathbf{y}), \boldsymbol{\kappa}) = \max_{\{\tau^i(\mathbf{y}), \bar{\omega}'^i(\mathbf{y})\}_{i=2}^N} u\left(y^1 + \kappa^1 R - \Delta\kappa^1(\mathbf{y})d + \sum_{i=2}^N \tau^i(\mathbf{y})\right) + \beta\mathcal{U}(\bar{\boldsymbol{\omega}}'(\mathbf{y}), \boldsymbol{\kappa}'(\mathbf{y})) \tag{3.22}$$

s.t.

$$[\lambda^i(\mathbf{y})] \qquad u(y^i + \kappa^i R - \Delta\kappa^i(\mathbf{y})d - \tau^i(\mathbf{y})) + \beta\bar{\omega}'^i(\mathbf{y}) \geq \omega^i(\mathbf{y}) \tag{3.23}$$

$$[\phi^1(\mathbf{y})] \qquad u\big(y^1 + \kappa^1 R - \Delta\kappa^1(\mathbf{y})d + \sum_{i=2}^N \tau^i(\mathbf{y})\big) + \beta\mathcal{U}(\bar{\boldsymbol{\omega}}'(\mathbf{y}), \boldsymbol{\kappa}'(\mathbf{y})) \geq \Omega^1\big(y^1, \kappa^1\big) \tag{3.24}$$

$$[\phi^i(\mathbf{y})] \qquad u(s^i Y + \kappa^i R - \Delta\kappa^i(\mathbf{y})d - \tau^i(\mathbf{y})) + \beta\bar{\omega}'^i(\mathbf{y}) \geq \Omega^i(y^i, \kappa^i) \tag{3.25}$$

[15] See also Phelan and Townsend (1991).

[16] A formal justification of this approach is provided by Lemma A1 and Lemma A2 of Pavoni and Violante (2007).

where $i \in \{1, \ldots, N\}$,

$$\Omega(y^i, \kappa^i) := u(y^i + \kappa^i R - \Delta\kappa^i_{aut}(y^i)d) + \beta \int \Omega(y', \kappa^i + \Delta\kappa^i_{aut}(y^i))\, \mathrm{d}F(y') \tag{3.26}$$

is the best outside option for household $i \in \{1, \ldots, N\}$, and the investment state is updated as:

$$\kappa'^i = \kappa^i + \Delta\kappa^i \qquad \text{where} \qquad \kappa'^i, \Delta\kappa^i \in \{0, 1\} \tag{3.27}$$

The main differences between these conditional value functions and the case without investment are that (i) some households will potentially now invest, adding the $-\Delta\kappa^i(\mathbf{y})d$ terms to household utility; (ii) the investment state $\boldsymbol{\kappa}$ must be updated when investment occurs; and (iii) the outside option for household i now allows for the option of future investment, if the household has not already invested.

As before this gives first order conditions for $i \in \{2, \ldots, N\}$, now (implicitly) conditional on both the income shock (as before), and also the investment decision, Δk:

$$[\text{FOC}(\tau^i(\mathbf{y}))] \qquad \frac{\mathrm{d}u(c^1(\mathbf{y}))/\,\mathrm{d}\tau^i(\mathbf{y})}{\mathrm{d}u(c^i(\mathbf{y}))/\,\mathrm{d}\tau^i(\mathbf{y})} = \frac{\lambda^i(\mathbf{y}) + \phi^i(\mathbf{y})}{1 + \phi^1(\mathbf{y})} \tag{3.28}$$

$$[\text{FOC}(\bar{\omega}'^i(\mathbf{y}))] \qquad \frac{\partial \mathcal{U}(\bar{\boldsymbol{\omega}}'(\mathbf{y}), \boldsymbol{\kappa}')}{\partial \bar{\omega}'^i(\mathbf{y})} = -\frac{\lambda^i(\mathbf{y}) + \phi^i(\mathbf{y})}{1 + \phi^1(\mathbf{y})} \tag{3.29}$$

$$[\text{ET}(\omega^i(\mathbf{y}))] \qquad \frac{\partial V(\mathbf{y}; \boldsymbol{\omega}(\mathbf{y}), \boldsymbol{\kappa})}{\partial \omega^i(\mathbf{y})} = -\lambda^i(\mathbf{y}) \tag{3.30}$$

Hence:

$$\frac{\partial \mathcal{U}(\bar{\boldsymbol{\omega}}'(\mathbf{y}), \boldsymbol{\kappa}'(\mathbf{y}))}{\partial \bar{\omega}'^i(\mathbf{y})} = -\frac{-\frac{\partial V(\mathbf{y}; \boldsymbol{\omega}(\mathbf{y}), \boldsymbol{\kappa}'(\mathbf{y}))}{\partial \omega^i(\mathbf{y})} + \phi^i(\mathbf{y})}{1 + \phi^1(\mathbf{y})} = -\frac{\mathrm{d}u(c^1(\mathbf{y}))/\,\mathrm{d}\tau^i(\mathbf{y})}{\mathrm{d}u(c^i(\mathbf{y}))/\,\mathrm{d}\tau^i(\mathbf{y})} \tag{3.31}$$

The first order conditions and envelope theorem result take the same form as without the investment decision. Hence the conditional value function is decreasing in promised utility, and the ratio of marginal utilities updated when a limited commitment constraint binds. This fully characterises the insurace transfers, given some exogenous investment decision. I next study the investment decision, and how it is influenced by various features of the model.

3.2.3 Poverty Trap

The first result from the model is that it naturally gives rise to the possibility of a poverty trap: a situation in which the long run equilibrium of the economy depends on its initial state. I will build this result in two steps. First, I suppose that households *are* able to commit i.e. the limited commitment friction is removed. In this case there will be a network-level poverty trap where some communities will be too poor to be able to ever invest. The structure of this trap will be analagous to a household-level trap: the only thing preventing investment is a lack of resources. However, this is insufficient to explain the observation that networks which have the resources choose not to invest. I then reintroduce limited commitment, and show conditions under which this can 'deepen' the poverty trap. Now networks which have sufficient resources to invest under full commitment may not invest, because the lack of commitment prevents resource pooling. This is the key mechanism driving the model.

Full Commitment

Under full commitment there exists a sequence of aggregate income thresholds $\widehat{Y}^{\mathrm{FC}}_{\Delta k}$, one between each possible level of investment and the level above it, such that if $\widehat{Y}^{\mathrm{FC}}_{\Delta k} < Y < \widehat{Y}^{\mathrm{FC}}_{\Delta k+1}$ then it will be optimal to make Δk investments this period. This leads to the possibility of a poverty trap: if an economy never receives a large enough level of aggregate income to reach the lowest threshold, i.e. $Y^{\max} < \widehat{Y}^{\mathrm{FC}}_{1}$ then it will forever remain with the current income distribution (absent external shocks), whilst if an external shock is provided to produce a 'small push' then further investment will be able to occur over time.[17]

Proposition 3.1. *There exists a unique threshold $\widehat{Y}^{FC}_{\Delta k} = \widehat{Y}^{FC}_{\Delta k}(\boldsymbol{\kappa}, N)$ such that with full commitment:*

1. *$\forall Y < \widehat{Y}^{FC}_{\Delta k}$, the optimal number of investments is no greater than $\Delta k - 1$;*
2. *at $Y = \widehat{Y}^{FC}_{\Delta k}$, $V_{\Delta k-1}(\cdot) = V_{\Delta k}(\cdot) \geq V_{\Delta k'}(\cdot)\ \forall \Delta k'$ i.e. the planner is indifferent between making $\Delta k-1$ and Δk investments and does not strictly prefer any other level of investment to these; and*
3. *$\forall Y > \widehat{Y}^{FC}_{\Delta k}$, the optimal number of investments is no fewer than Δk.*

There are $N-k$ such thresholds, with $\widehat{Y}^{FC}_{\Delta k-1} \leq \widehat{Y}^{FC}_{\Delta k}$, each implicitly defined by $\Gamma_{\Delta k}(\widehat{Y}^{FC}_{\Delta k}; \boldsymbol{\kappa}, N) \equiv 0$ where $\Gamma_{\Delta k}(\cdot) := V_{\Delta k-1}(\cdot) - V_{\Delta k}(\cdot)$.

Proof. See Subsection 3.7.3. □

Proposition 3.1 states that for an N-household economy in which $k = \sum_i \kappa^i_k$ investments have already been made, there are $N - k$ income thresholds, whose level depends on the number of existing investments and the network size, under the assumption that households can commit fully. Intuitively, when aggregate income is very low, it will be optimal to consume it all today, potentially after some redistribution. At higher levels of aggregate income, the utility cost of

[17] In contrast with 'big push' models (Rosenstein-Rodan, 1943; Nurkse, 1953; Murphy et al., 1989), here no coordination is needed between agents: an initial push that is large enough to allow one additional investment to occur will then automatically spillover, allowing further investments.

reducing total consumption by d today (the cost of an investment) is sufficiently low compared with the expected improvement in future expected utility, so it will become optimal to make one investment. At yet higher levels of aggregate income, additional investments become worthwhile. Network size scales down the per household cost (and return) of each investment.

This threshold result leads naturally to the possibility of a poverty trap, where the long run distribution of income depends on its initial state. When an economy has only a small number of initial investments (low level of capital), the highest possible aggregate income may be lower than $\widehat{Y}_1^{\mathrm{FC}}(\cdot)$, the level needed to make the first additional investment worthwhile. However, at a higher level of initial capital stock the maximum level of aggregate income is higher, allowing further investments to take place in some states of the world.

Lemma 3.2. *The threshold level of income needed to make Δk additional investments, $\widehat{Y}_{\Delta k}^{FC}$, is decreasing in the existing level of capital k, i.e. $\mathbf{D}_k \widehat{Y}_{\Delta k}^{FC} < 0$, where $\mathbf{D}_k$ is the finite difference operator (the discrete analogue of the derivative) with respect to k.*

Proof. See Subsection 3.7.4. □

Under full commitment the poverty trap result has very stark predictions: there are only two possible long run equilibria, $\boldsymbol{\kappa} = \mathbf{0}$ or $\mathbf{1}$.[18] This is because under full commitment only the level of aggregate income matters for whether investment takes place. Suppose there exists a state of the world in which, from a base of zero capital, making at least one investment is optimal for the planner. Then making an investment in the same state of the world, when the same combinations of endowment incomes are realised but when some investments have already occurred, must also be optimal (by Lemma 3.2). Hence either the economy will remain with zero capital or will converge to a state in which all households are invested.

Limited Commitment

I next consider how the above results are changed by limited commitment. I first show that limited commitment can change the 'depth' of the poverty trap: the threshold level of income needed such that doing some investment becomes optimal in equilibrium. To do this I consider how the investment threshold under autarky compares to that with full commitment. The results under limited commitment will fall somewhere between these, depending on the extent of limited commitment. I then show that with limited commitment, a wider range of equilibrium levels of

[18] $\boldsymbol{\kappa} = \mathbf{0}$ is the poverty trap long run equilibrium, while $\boldsymbol{\kappa} = \mathbf{1}$ is the 'good' long run equilibrium. If there were decreasing returns to investment at the aggregate level, the good equilibrium might be less extreme, with only some households ever investing, but there would be the same initial threshold needed to break out of the poverty trap equilibrium.

investment are possible.[19] This is important since in practice one observes intermediate levels of investment, which would never be a long run equilibrium with full commitment.

Under autarky, there will be an income threshold $\hat{y}$ such that if an (uninvested) individual household's income exceeds $\hat{y}$ it will invest, else it will not, and all non-investment income will be consumed.[20]

The first result is that, if all households have incomes below the level needed to invest in autarky i.e. $y^i < y^{\text{aut}}$, then the number of investments will necessarily (weakly) fall. To see this, note that limited commitment reduces the ability to make transfers today in expectation of receiving transfers in the future. Hence equilibrium outcomes under limited commitment are always between the full commitment and autarky outcomes. By assumption no household wants to invest in autarky, so if any investment were taking place under full commitment, under limited commitment it can only be weakly lower.

An analagous result can be seen when in autarky some subset of households would have chosen to invest. If under full commitment they were required to instead make transfers and not invest, then limited commitment moves them back towards their autarky choice. They are willing to leave the insurance arrangement, and if the cost (in terms of promised future utility) of asking them to not invest and make transfers instead is high enough, then they will again invest.

To see how the individual and aggregate thresholds compare, the threshold for a single investment to occur under full commitment, note that a move from autarky to full commitment insurance has two effects. First, it effectively scales the cost and return of the investment, as these can now be shared across households. Under full commitment, household j pays only $\alpha^j d$ per investment, where $\alpha^j = \alpha(\lambda^j, \underset{\sim}{\lambda}^{-j})$ is household j's share of aggregate consumption. In the limit as $N \to \infty$, holding the distribution of individual income constant $\alpha^j \to 0$ so collectively investment becomes infinitely divisible, and the problems of 'lumpiness' go away. Doing at least one investment (collectively) therefore becomes increasingly attractive relative to zero investments, reducing the threshold level of income needed for a single investment to occur.

Second, it reduces the variance of future consumption. Part of the value of the investment is that it is not perfectly correlated with households' endowment income, so provides some partial insurance.[21] Consequently, insurance from other households will reduce the demand

[19] Adding heterogeneity in investment returns to the model would also allow intermediate equilibria, where only some households were invested. However, as I show in the next subsection, limited commitment has implications for how the distribution of income and risk-sharing network size should matter for investment, which would not hold in a full commitment model with heterogeneity. I will show that these implications are borne out in the data, so that limited commitment is an important reason why such intermediate cases may be observed, although heterogeneity is certainly also present.

[20] To see that such a threshold exists, the same lines of reasoning used in the full commitment case can be replicated. The threshold is implicitly defined as $u(\hat{y}) + \beta\mathbb{E}\left[\Omega(y', 0)\right] = u(\hat{y} - d) + \beta\mathbb{E}\left[\Omega(y', 1)\right]$.

[21] In fact I model the return from investment as non-stochastic and hence entirely independent of the endowment income process, but this is not necessary.

for investment relative to current consumption. This effect will increase the threshold level of aggregate income needed for an investment to occur. Hence the overall effect of limited commitment may be to increase or decrease investment relative to full commitment insurance: determining the effect in a particular context will be an empirical question. Some intuition can be gained by considering properies of the cross-sectional income distribution.

To see how the relative magnitudes of these effects depend on the distribution of income, consider two polar cases: the case where all risk is aggregate (household incomes are perfectly positively correlated) and the case where all risk is idiosyncratic (aggregate income is fixed). In the former case, 'full commitment insurance' actually provides no insurance at all. However, it does allow households to use transfers to share the costs and returns of investment, so only the first of the above effects exists here. If individual households' incomes are not already high enough to make investment worthwhile, then pooling income may allow households to invest. In this case it is clear that as soon as limited commitment is introduced, the households are effectively in autarky. Without any idiosyncratic variation in income, there is no value to an invested household in remaining in the arrangement. Hence there is an immediate unravelling, with no households being willing to make transfers that support another household's investment, since they know that repayment is not credible.

Conversely, when aggregate income is fixed, and all variation is idiosyncratic, and hence insurable, the insurance arrangement has its maximum value. However, it is now possible that the availability of insurance can 'crowd out' investment: households with relatively high income shocks would invest in autarky, but with insurance they are required to instead make transfers. If the fixed level of aggregate income is below $\widehat{Y}_1$, then with full commitment insurance no investments will ever take place. In this case limited commitment weakens the insurance arrangement, which might allow *more* investment to take place. Households who receive relatively good shocks might not be able to commit to providing full smoothing to those who were unlucky: instead they may also invest. Since insurance is here at its most valuable, this is the case where households are most willing to forgo investment to ensure continued access to the insurance arrangement. At intermediate levels of correlation, the effect of the LC constraint is in between these extreme cases.

As well as changing the level of the lowest threshold, $\widehat{Y}_1$, limited commitment can change the distance between the thresholds. This is important, as it can create long run equilibria where the economy has an intermediate level of capital, rather than the all or nothing result seen under full commitment.

To see this, consider the situation where one household has invested. Under full commitment, I showed that the threshold level of income needed to do one additional investment has now fallen.

Under limited commitment there is an additional effect: relative to the case where no-one has invested, the household with an investment has an improved outside option. This endogenously restricts the set of possible equilibrium transfers. Since household consumption will no longer be a constant share of aggregate consumption, and since owning an investment increases the consumption share for a household, there will no longer necessarily be increasing differences in the planner's utility when another household invests. The limited commitment analogue of Lemma 3.2 may therefore not hold: an increase in the level of capital will not necessarily reduce the income thresholds for investment. Instead there are now parameters which can support 'intermediate' equilibria, where the long run share of households who are invested is strictly between zero and one.[22]

3.2.4 Comparative Statics

I now consider two additional testable predictions of the model: the effect of changes in the distributions of income, and in the size of the risk-sharing network. These predictions arise specifically from the interaction of lumpy investment with limited commitment, and would not be present either with full commitment, or with a single alternative source of insurance market incompleteness (e.g. hidden effort, hidden income).

How Does Inequality in Initial Income Affect Investment?

In the presence of limited commitment, income inequality affects the decision to invest. The intuition of this result is straightforward: increased income inequality affects the set of limited commitment constraints which are binding, by changing the outside options for households. As shown above, limited commitment has a direct impact on investment decisions.

More concretely, consider a redistribution of endowment income from a (poorer) uninvested household, whose limited commitment constraint does not bind, towards a (richer) uninvested household whose constraint is binding. The increase in income for the richer household improves that household's outside option, making the limited commitment constraint for that household more binding. If the arrangement previously required the household to invest, this will remain unchanged, whilst if the household was not previously asked to invest, the planner may now find this an optimal way to provide utility to the household.

If the redistribution had been from a poorer to richer household where *both* households had binding limited commitment constraints, the same argument would hold for the richer household, but now the reverse may occur for the poorer household: since the planner need to

[22] Hence although all households are ex ante identical, there are long run equilibria where they necessarily have different levels of expected utility. Matsuyama (2002, 2004, 2011) provides other examples of models which have this 'symmetry-breaking' property.

transfer less utility to this household, it may no longer provide the household with investment as a way to transfer some of this utility. Depending on which household is at the margin of investment, an increase in inequality can therefore increase or decrease total investment. For a given level of aggregate income, a small increase in inequality (starting from a very equal initial distribution) will lead to some LC constraints starting to bind. As this inequality increases and these constraints become increasingly binding, this increases the number of investments that occur. Eventually, further increases in income inequality lead to a reduction in investments, as they are effectively redistributions from one constrained household to another, reducing the need to provide the poorer of these households with an investment. Hence there will be an 'inverted-U' shape effect of inequality, where initial increase in inequality will increase investment, but too much concentration in just a few hands will again reduce the level of investment.

Proposition 3.2. *Consider an initial distribution of income,* $\mathbf{s}$*. Let* $\mathbf{s}'$ *be an alternative, more unequal distribution, such* $\mathbf{s}'$ *is a mean preserving spread of* $\mathbf{s}$*. For relatively equal distribution,* $\mathbf{s}$*, investment will be weakly greater under* $\mathbf{s}'$*. For relatively unequal distribution,* $\mathbf{s}$*, investment will be weakly lower under* $\mathbf{s}'$*.*

Proof. See Subsection 3.7.5. □

How Does Network Size Affect Investment?

Given a fixed distribution for individual income, increasing network size has two complementary effects on investment. First, it raises expected aggregate income. Second, it reduces the variance of average income (assuming that incomes are not perfectly correlated) and increases the variance of aggregate income. Even if the mean of aggregate income were fixed, a mean preserving spread of aggregate income would increase investment, since there would be more extreme high income shocks. Under full commitment these periods provide a large incentive to invest to smooth income across time. Under limited commitment there is an additional effect that, with a lower variance for mean income, the value of the insurance arrangement is improved, so autarky is relatively less attractive. This makes it easier to sustain investment.

Proposition 3.3. *An increase in the number of households reduces the threshold level of aggregate income needed for investment by improving the value of the insurance arrangement. It also increases the likelihood of aggregate income exceeding even the initial threshold.*

Proof. See Subsection 3.7.6. □

This prediction would not be true if moral hazard or hidden information were the friction driving incomplete insurance. It is also the opposite of what one would see in a model where the network (or some share of it) needs to coordinate for investment to be profitable: then larger group sizes would make coordination more difficult.

3.3 Data from a Randomised Control Trial in Bangladesh

3.3.1 Data Source

I use data from a large scale, long term randomised control trial in rural Bangladesh, collected in partnership with microfinance organisation BRAC. The data cover 27,000 households across 1,409 villages, in the poorest 13 districts of rural Bangladesh.

The villages were selected as follows. From each district, one or two subdistricts (upazilas) were randomly sampled. From each of these, two BRAC branch offices were randomly selected for the program, one to be treated, the other as a control (for more details see Bandiera et al., forthcoming). All villages within 8km of a sampled branch office were then included in the final sample, giving the total of 1,409 villages, with a median of 86 households.

A census of households in each village took place in 2007. This asked questions on demographics of household members, and their education and employment statuses, as well as collecting detailed information on household assets. This was used both to construct a sampling frame for the further surveys, and for targeting the program.

A sample of households was then selected from each village. A participatory wealth ranking in the census divided households into one of four wealth categories. All households in the lowest wealth grouping – which includes all households eligible for the program – were sampled, along with a 10% random sample of all remaining households. This gives a sample of 7,111 eligible households, 13,704 'ineligible poor' households (in the bottom two wealth ranks), and 6,162 'non-poor' households. Sampled households were given a baseline survey in 2007, with follow up surveys in 2009 and 2011. In these surveys detailed data were collected on household income, investment, and risk sharing.

Table 3.1 provides some key descriptives about these households, grouping them into the above categories. Households comprise around four members, but poorer households are smaller as they are more likely to not have a working age man present. This is particularly true in eligible households, where it was used in program targeting (see below for details). Eligible households are very poor, with almost half below the poverty line, and hardly any already own cows. Ineligible poor households, and then non-poor households, do indeed have higher incomes, consumption, and assets (cows), providing evidence that the participatory wealth ranking provides a good measure of relative material standard of living.

Four features of the data make them suitable for my context. First, the data cover a large cross-section of networks, encompassing more than 1,400 villages. This is important since the model predictions are at the network level. Second, in a subsample of the data, exact links used for risk-sharing transfers are measured. This makes it possible to construct a good proxy for the appropriate risk-sharing network in the full dataset, addressing concerns that the whole village is not the level at which risk sharing takes place. Third, the program provided large injections of lumpy capital, with significant variation in the number of transfers across villages (see Figure 3.1). This provides the exogenous variation in aggregate income necessary for my test of a network-level poverty trap. Fourth, households were surveyed again two and four years after the transfers were made. This allows study of how the initial transfers affect later investment decisions, which are necessarily long term.

3.3.2 Program Structure

The intervention carried out by BRAC was an asset transfer program. Using information from the census survey, household eligibility for the program was determined. Eligibility depended on a number of demographic and financial criteria. A household was automatically ineligible for the program if any of the following were true: (i) it was already borrowing from an NGO providing microfinance; (ii) it was receiving assistance from a government antipoverty program; or (iii) it has no adult women present.[23] If none of these exclusion criteria were met, a household was deemed eligible if at least three of the following inclusion criteria were satisfied: (i) total household land was less than 10 decimals (400 square metres); (ii) there is no adult male income earner in the household; (iii) adult women in the household work outside the home; (iv) school-aged children have to work; (iv) the household has no productive assets.

After the baseline survey, eligible households in treated villages were given a choice of asset bundles.[24] All bundles were worth approximately the same amount, $515 in 2007 PPP. 91% of treated households choose a bundle with cows, 97% with cows or goats. In the following analysis I treat all treated households as though they actually received cows, but my results are robust to treating those who did not choose livestock as though they received no transfers. Along with assets, treated households also receive additional training from BRAC officers over the following two years. By the 2009 survey, all elements of the program had ceased, except that treated households now had the additional capital they had been provided with. After the 2011 survey, eligible households in control villages also received asset transfers.

[23] The last criterion exists because the asset transfers were targeted at women.

[24] At the time of the asset transfers, eligibility was reassessed and 14% of households that were deemed eligible at the census no longer met the eligibility criteria. However, there is significant variation in the share of households no longer deemed eligible across branches, suggesting that implementation of reassessment varied across branches. To avoid the concern that this introduces unwanted variation, in what follows I continue to use the initial eligibility status.

One limitation of the program structure, for the purpose of this study, is that while entire villages are either treated or control, variation in the *intensity* of treatment – the value of transfers to a risk-sharing network, which is proportional number of households in the network who receives transfers – is endogenous, since it depends on characteristics of the households. The ideal experiment for my context would have been to directly randomise villages into G groups, where group 1 has zero households receiving asset transfers, group 2 has 1 household receiving transfers, and so on. Then the marginal effect of having $g+1$ households treated rather than only g households could be estimated by comparing outcomes for households or networks in groups $g+1$ and g. In Subsection 3.4.1 I discuss two different approaches I take to handle this, one exploiting the available randomisation and the other using the non-linearity of the relationship being tested for.

3.3.3 Defining Risk-Sharing Networks

The predictions of the model concern behaviour at the risk-sharing network level. Early work on informal risk sharing assumed that the relevant group in which risk sharing takes place is the village (Townsend, 1994). Implicitly this assumes there are frictions preventing risk sharing with households outside the village, and that within the village all households belong to a common risk-sharing pool. Recent evidence suggests that in some context risk-sharing networks might be smaller than the village. Using data from Indian villages, Mazzocco and Saini (2012) and Munshi and Rosenzweig (2016) both find that caste groups within a village are the appropriate risk-sharing network i.e. there are important frictions preventing risk sharing across caste lines within a village.

To determine the appropriate group for risk sharing, I use a subsample of 35 villages in which, rather than the stratified random sampling scheme used elsewhere, a census of all households was taken at all waves. Households were asked whether they suffered a 'crisis' in the last year. If they did, they were asked how they coped with it, and where transfers or informal loans were used for coping, they were asked who the transfers or loans were from. Additionally all households were asked who (if anyone) they borrowed food from or lent food to. I combine these various dimensions of household links into a single dimension, which I term 'sharing risk'. I then study what grouping can be constructed in the full sample, that provides a good proxy for being a risk-sharing partner of an eligible household, since my interest is in constructing the risk-sharing network for these households.

Table 3.8 provides evidence on this question. The first point of note is that almost all of eligibles' risk sharing (94%) is done with other households in the same village. Second these links are highly concentrated among other households in the lowest two wealth classes. In particular, 70% of eligibles risk-sharing links are with other households from the bottom two wealth classes, compared with only 55% that would be expected under random linking. This motivates me to focus on the poorest two wealth classes as the relevant group for risk sharing.

To further test this definition of the risk-sharing network, I perform Townsend tests (Townsend, 1994) under the different groupings. These involve regressions of the following form:

$$
\begin{aligned}
\Delta \log c_{hgt} &= \beta_0 + \beta_1 \Delta \log y_{hgt} + \boldsymbol{\beta} \Delta \mathbf{z}_{hgt} \\
&\quad + \delta_0 D_{hg} + \delta_1 D_{hg} \Delta \log y_{hgt} + \boldsymbol{\delta} \Delta D_{hg} \mathbf{z}_{hgt} + \gamma_{gt} + \varepsilon_{hgt}
\end{aligned}
\tag{3.32}
$$

where $\Delta \log c_{hgt}$ is the change in log expenditure for household h in risk-sharing group g at time t; $\Delta \log y_{hgt}$ is the change in log income; $\Delta \mathbf{z}_{hgt}$ are changes in demographic charateristics; $D_{hg} = 1$ if household h is not an eligible household; and γ_{gt} are group dummies. The idea of the test is that, if eligible households in group g are able to fully smooth consumption, their expenditure should not respond to changes in their household income, i.e. $\beta_1 = 0$, once changes in demographics and group-level shocks, γ_{gt}, which cannot be smoothed, are accounted for. Including the interactions with D_{hg} allows ineligible households (poor and non-poor) to potentially be in the same risk-sharing group as the eligible households but without imposing that they respond to shocks in the same way. This ensures that the results of the test are not confounded by changes in sample composition: β_1 always measures the response of eligibles' expenditure to their income. The appropriate risk-sharing network for eligibles will then be the grouping such that, including fewer households gives a larger β_1, but including additional households does not further reduce β_1. If all of eligibles' risk sharing is with the bottom two wealth classes, then excluding ineligible poor households should make risk sharing appear worse, since part of the aggregate shock is being excluded. Conversely, including the whole village should not improve measured risk sharing, because the additional households are irrelevant.

To estimate this I use data on expenditure and income for all households in control villages in the main sample over the three waves of data collection. Both the observations and variables used in this test are separate from the previous approach, so this provides independent evidence about the appropriate group. Equation 3.32 is estimated for the different definitions of group previously considered. Table 3.9 shows the results of this test. Consistent with the earlier result, it can be seen that including ineligible poor households into the risk-sharing network for the eligibles improved measured consumption smoothing (p-value=.026). However, including the rest of the village does not further improve smoothing (p-value=.403), justifying their exclusion from the risk-sharing network.

To the extent that using wealth groupings is an imperfect proxy for the true risk-sharing network, it will introduce some noise into my later work. As a robustness check, in my empirical test for a poverty trap I will show that qualitatively similar results would be found if the entire village were used, or only eligible households are used.

3.3.4 Final Sample Descriptives

My final sample, focusing on the bottom two wealth classes, includes 20,815 households across 1,409 villages, although as a robustness check I show results including all households. Table 3.2 provides some key descriptives about the (poor) risk-sharing networks. Note that these means and aggregates (and all further ones) are constructed using sample weights to provide statistics representative of the underlying population.

Aggregate income is $\$53,600$ for the median risk-sharing network, which has 50 households, while the median asset transfer is worth 4% of this. There is also variation in income inequality and network size, allowing me to test the additional predictions of the model.

3.3.5 Verifying Model Assumptions

I first verify that the context matches the modelling framework in five dimensions: (1) households have variable incomes; (2) household savings are small relative to income; (3) households engage in risk sharing; (4) households have potentially productive lumpy investments available; (5) risk-sharing networks have the resources needed to be able to invest.

1. *Households have variable incomes.* Using only the time series variation for households in the poor risk-sharing networks in control villages, the median coefficient of variation is .35 (mean is .41).

2. *Household savings are small relative to income.* At baseline, the median household in the villages covered by my data has cash savings totalling .5% of their income. Including also jewellery and ceremonial clothing, this rises to 3.6%, and including other household assets (including consumer durables) it reaches 11.8% of income. Savings, even including jewellery, are therefore an order of magnitude smaller than income shocks, and so have limited scope for providing consumption smoothing.

3. *Households engage in risk sharing.* As described above, households were asked whether they suffered a crisis, and if so how they coped with it. They may report multiple methods. Potential crises include crop loss, serious illness or death of household member, and damage to house due to natural disaster. To avoid confounding with the asset transfer program, I consider only households in all control villages, and I pool their responses over the

three waves. In each wave, about half of all households report suffering some kind of crisis. Of those who report suffering a crisis, 38% receive loans or transfers from other households to provide smoothing. 50% of households also use their own savings to provide some smoothing, although as noted these savings are small relative to the size of shocks households face. 36% of households also report reducing consumption during a crisis. Taken together, these results indicate that households use risk-sharing transfers as an important channel of consumption smoothing, but consumption smoothing is incomplete.

4. *Households have potentially productive lumpy investments available.* Bandiera et al. (forthcoming) document that for these data that the mean internal rate of return on cows is 22%. In 2007 USD PPP terms, a cow costs around $257. This is 18% of median household income in a village, and 29% of median household income among the households eligible for the program.

5. *Risk-sharing networks have the resources needed to be able to invest.* Figure 3.2 shows the distribution of aggregate wealth holdings across risk-sharing networks, as defined in Subsection 3.3.3. Wealth is broken down into a number of categories, and the cost of a cow is marked on the figure. This gets to heart of the puzzle this paper seeks to explain: more than 75% of risk-sharing networks have available to them enough *cash*, let alone other assets, needed to be able to invest in cows. Yet despite this, and the high returns, these savings are not pooled across households to purchase cows.

3.4 Empirical Evidence

First I provide evidence of a network-level poverty trap. Second, I test the additional comparative static predictions of the model, to provide supportive evidence of limited commitment. Third, I consider three leading alternative explanations for a network-level poverty trap, and show that their predictions are not borne out in my empirical context.

3.4.1 Evidence for a Network-Level Poverty Trap

The prediction of the model is that there should exist some aggregate income threshold such that (i) below the threshold the network is in a poverty trap and we see no investment, (ii) above the threshold we see investment taking place, with investment increasing in the value of transfers. To test this, I use exogenous variation in the amount of capital (and hence, implicitly, income) provided *at the network level* by the asset transfer program. As described above, the program provided the same value of assets to all eligible households in treated villages. However, there is variation in the number of eligible households within a village. Hence the comparison I make is between risk-sharing networks with the same number of eligible households across treatment and control villages.

Non-parametric Relationship Between Investment and Capital Injection

To investigate the prediction of a network-level poverty trap, I begin by testing non-parametrically the reduced form effect. I study how investment by the network between 2009 and 2011, $\Delta k_{v,2011}$, varies with the value of the capital injection provided by the program, $\Delta k_{v,2009}$ (both measured in 2007 USD PPP).[25] Precisely I estimate the following local mean regression:

$$\Delta k_{v,2011} = m(\Delta k_{v,2009}) + \epsilon_{v,2011} \tag{3.33}$$

separately for treated and control networks, where $m(\cdot)$ is unknown and estimated using a Nadaraya-Watson kernel-weighted local mean estimator.

Figure 3.3 plots the conditional mean, and 95% confidence interval. It can be seen that investment in further cow ownership is close to zero and does not vary with the value of the capital injection up to a value of around \$4,000. When more than this level of capital was provided by the program, there appears to be an increasing (and approximately linear) relationship between the capital provided and the amount of additional investment takes place. This is precisely the relationship predicted by the model.

As discussed earlier, the ideal design for my context would be to have experimental variation in the intensity of treatment, as measured by the number of households that receive transfers. Since the number of transfers to a village is endogenous, conditional on being in a village that is treated, there are two approaches I take to provide support for this result, each dealing with a different potential worry.

To test whether the observed relationship is due to the program, or just due to underlying heterogeneity, I plot the same relationship for the control sample. Figure 3.7 plots additional investment between 2009 and 2011 against the value of the capital injection that *would have been* provided had the risk-sharing networks been in treated villages. It is clear from this that in the absence of actual asset transfers, investment is zero on average, and does not vary with the placebo value of capital injection.

To check for robustness of the relationship to definitions of the risk-sharing network, I re-estimate the relationship for the treated sample, using different levels of aggregation. Figure 3.8 estimates the relationship in Equation 3.33 where all households in the village are assumed to belong to the risk-sharing network, rather than only those in the lower two wealth classes.

[25]Since the program provided some consumption support and training between 2009, I do not try to disentangle what occurs between 2007 and 2009. Instead I study the additional investment that takes place after 2009, by when the program is no longer active and no additional support is being provided.

Three points are of note. First, the general shape of the relationship remains the same, and the apparent threshold is at the same location. Second, the entire graph has been translated upwards by $1,000. This implies the richer households in these villages were doing some investment, but this is not responsive to the amount of capital injected, consistent with them not being part of the risk-sharing network. Third, the confidence intervals are now wider. If the richer households are not part of the risk-sharing network of the eligible households, then including them should just add noise to the estimated effects, as can be seen.

At the other extreme, Figure 3.9 estimates the relationship supposing that eligible households are part of a common risk-sharing network that excludes all other households. Again the same shape of relationship is visible, with a similar apparent location for the threshold. However, the slope of the relationship above the threshold is much flatter, so that at $7,000 capital injection, the total additional investment is now $1,000. In Figure 3.3, investment at this level of capital injection was $4,000, indicating that *other households* were also investing at these higher levels of capital injection, but not investing at low levels. This apparent spillover – with additional investment by ineligible poor households depending on the number of eligible households – is direct evidence that consideration of the risk-sharing network is important when studying the impact of this type of program. It also helps rule out explanations based on household-level poverty traps: if these were the only explanation for the initial lack of investment, then households which don't benefit from the program should not be responding.[26]

Testing Formally for a Threshold Effect

The non-parametric results suggest that, among the treated networks, there exists a threshold value of aggregate capital injection needed to spur additional investments by the risk-sharing network. To test this relationship formally, I estimate for the treated sample a regression of the form:

$$\begin{aligned}\Delta k_{v,2011} = \alpha_1 &+ \delta_1 \Delta k_{v,2009}.\mathbf{1}\{\Delta k_{v,2009} < \Delta k^*\} \\ &+ \delta_2 \Delta k_{v,2009}.\mathbf{1}\{\Delta k_{v,2009} \geq \Delta k^*\} + \boldsymbol{\gamma}_1 \mathbf{X}_{v,2009} + \boldsymbol{\gamma}_1 \mathbf{X}_{v,2007} + \epsilon_{v,2011}\end{aligned} \tag{3.34}$$

where again $\Delta k_{v,2011}$ is the increase in cow investment by the network as a whole (v) between 2009 and 2011, $\Delta k_{v,2009}$ is the value of the asset transfers provided to the network by the program, Δk^* is a proposed threshold value of asset transfers, and $\mathbf{X}$ is a vector of controls. Note that since the asset transfer by the program takes place in 2007, $\Delta k_{v,2011}$ does not include

[26]In Subsection 3.4.3 I rule out other alternative explanations, including the possibility that other households' investment can be explained by general equilibrium effects.

the initial injection. All monetary values are in 2007 USD at purchasing power parity exchange rates.

This specification captures the idea that there is some threshold level of asset transfers, Δk^*, needed to push a network out of the poverty trap. Below this threshold there should be no additional investment, $\alpha_1 = 0$ and $\delta_1 = 0$, and above this threshold we should see additional investment increasing in the value of capital injection, $\delta_2 > 0$. Whilst the model does not predict the functional form for how additional investment varies with the capital injection, the estimated non-parametric relationship, Figure 3.3, suggests that at least over the support of my data, linearity does not seem unduly restrictive.

Since the threshold, Δk^*, is unknown, I use an iterative regression procedure designed to test for a structural break (a change in the slope of the relationship) with unknown break point. This involves running a sequence of such regressions over a prespecified range of possible values for Δk^*, and then testing for significance of the test statistic against an adjusted distribution, to account for the repeated testing.

I use two different statistics, both the Quandt Likelihood Ratio test (see Quandt, 1960; Andrews, 1993) and the Hansen test (Hansen, 1999). The former selects as the threshold location the point which maximises the absolute value of the t-statistic on δ_2. The latter uses a criterion based on the residual sum of squares, so accounts more directly for the relative explanatory power of the regression as a whole.

Precisely, for the Quandt Likelihood Ratio test I calculate for each possible threshold the F-statistic for the comparison between the model with and without the threshold. I then select from among these regressions, the one with the highest F-statistic. The corresponding threshold in that regression is then taken as the estimated location of the threshold. To test whether this threshold value is 'significant', I compare the F-statistic to the limiting distribution for this statistic under the null (Andrews, 1993), thus correcting for the multiple testing.

Table 3.3 shows the results of this test using different control variables, $\mathbf{X}$. In all cases the most likely location for a threshold is at \$3,500 of asset transfers, equal to 6.5% of income for the median network, and close to the level suggested by visual inspection of Figure 3.3.[27] This is equivalent to treating 14% of households in the median network. Figure 3.4 shows non-parametrically the relationship between the value of capital provided to the network by the the program (in 2007), and the additional investment by the network between 2009 and 2011, splitting the non-parametric plot at \$3,500. This makes the relationship clear to see.

[27]The discrepancy between the visual estimate and the formal method is caused simply by the non-parametric smoothing: by using observations below the threshold when estimating the local mean above the threshold, the figure makes the threshold look later and less sharp than it is.

Testing whether this potential threshold is itself statistically significant, I can reject at the 5% level the hypothesis that there is no threshold effect. This is true with additional controls, but when district fixed effects are included the qualitative patterns remain unchanged but the estimates become noisier. Studying the regression results, one can see that below the threshold the level of investment is close to zero, and above the threshold it is increasing, consistent with the model predictions.[28]

For the Hansen test, I estimate the same regression specifications as for Columns (1) and (3) in Table 3.3 above. For each possible threshold I calculate the residual sum of squares (RSS). I select among the regressions the one (or set) with the lowest RSS. The corresponding threshold in that regression is the estimated location for the threshold using this method. To test whether the threshold is significant, I test construct the Hansen statistic. This is, at any possible threshold, the difference between the RSS at that threshold and the minimum RSS from all thresholds considered, divided by the minimum RSS and multiplied by the sample size. This is necessarily equal to zero at the proposed threshold. If it is below .05 at any other tested threshold, then that threshold cannot be rejected as a possible location for the threshold.

Figure 3.10 shows the value of the likelihood ratio statistic from running the Hansen test for possible thresholds between \$2,000 and \$5,000, at intervals of \$100.[29] From this it is clear that the most likely location of the threshold is between \$3,700 and \$4,100, or 6.9-7.6% of income in the median network, close to the estimate of \$3,500 from using the Quandt Likelihood Ratio approach. Henceforth I use \$3,500 as the estimate of the threshold location, but my results are qualitatively robust to choosing instead a point in [\$3,700, \$4,100].

Impact of Capital Injection on Investment

Having identified the location of the threshold, I then estimate the following regression on the sample including both treated and control variables:

$$\begin{aligned}\Delta k_{v,2011} = \alpha_0 + \alpha_1 T_v + \beta_1 \Delta k_{v,2009} + \delta_1 T_v \Delta k_{v,2009}.\mathbf{1}\{\Delta k_{v,2009} < 3500\} \\ + \delta_2 \Delta T_v k_{v,2009}.\mathbf{1}\{\Delta k_{v,2009} \geq 3500\} + \boldsymbol{\gamma}_1 \mathbf{X}_{v,2009} + \boldsymbol{\gamma}_2 \mathbf{X}_{v,2007} + \epsilon_{v,2011}\end{aligned} \tag{3.35}$$

where T_v is an indicator for village treatment status, and all other variables are as before. Now the specification makes use of the exogeneity due to randomisation of villages: the coefficients α_1, δ_1, δ_2, are identified from the difference between the treated and control risk-sharing networks.

[28]Note however that since this regression is chosen using the iterative procedure described above, it would not be correct to use the standard errors provided directly for inference.

[29]Note that since the variation in the aggregate value of the capital injection comes only from variation in the number of treated households, there are only data points at intervals of \$515 (the value of the asset transfer to one household). I show the test using intervals of \$100 just to make clear the region of possible values that this test cannot reject.

I use again three possible aggregations of households: the risk-sharing networks I constructed, the entire village, and only the eligible households. It is important to note that the standard errors estimated here do not account for the prior estimation of the threshold location.

Table 3.4 shows the results of this estimation. The results, now using the randomisation for identification, are similar to what was seen non-parametrically: additional investment, $\Delta k_{v,2011}$, is flat with respect to the capital injection below the threshold \$3,500 ($\delta_1 = 0$), and increasing after the threshold ($\delta_2 > 0$). The threshold is robust to the different definitions of risk-sharing network.

The estimated coefficient δ_2 suggests that, after the first \$3,500 worth of asset transfers to a risk-sharing network, every additional \$500 generates a further \$750 in investment, although the confidence intervals are large.[30] Amongst the eligible households the effect of an additional \$500 after the threshold is between \$40 and \$280 of additional investment.

3.4.2 Evidence for Limited Commitment

Having provided evidence of a network-level poverty trap, I next show support for the limited commitment channel developed in the model. To do this I test the two comparative static predictions I developed from the model: (i) investment has an inverted-U shape in income inequality; and (ii) investment is increasing in network size.

Income Inequality

Although the program does affect income inequality, it does so in a way that simultaneously also changes the level of income and the value of risk sharing, since other households' income distributions have changed. Hence I will not be able to use the program to directly provide evidence for the inequality effect, because the program does not vary inequality independently of other relevant variables.

Instead I show how investment varies with income inequality in the control sample, using the variation available in the cross-section. An important limitation of this approach is the possibility that income inequality is endogenous. Two factors help mitigate this worry.

First, my prediction is on realised income inequality in a period, rather than underlying difference in expected incomes. All poor households are all engaged in very similar activities: 80% of hours worked by women are spent either in casual wage labour or rearing livestock, and 80% of hours worked by men are in casual wage labour, rearing livestock, or driving a rickshaw. Hence the income distributions from which households are drawing are likely to be

[30] Given how much more precise the estimates are among the eligible only group, this suggests that even the 'all poor' group I construct as a proxy for the risk-sharing network might be too large, containing some irrelevant households and making the estimates less precise.

similar, conditional on number of working age household members and whether they engage in livestock rearing or rickshaw driving. One approach then is to condition on these variables at the household level, and then construct inequality in residual income. At the network level, differences in (residual) income inequality should then be reflective only of the realisation of income shocks that period.

Second, to the extent that there are unobserved systematic factors which drive both realised income inequality and investment, to confound my test they would need to also act in a non-linear way. Since the relationship for which I am testing is an inverted-U shape, any unobserved variable which generates a monotone relationship between inequality and investment would not remove the relationship I am testing for. Whilst this would prevent me from treating the estimated parameters as causal, my intention is only to test the shape of the relationship, not to use the parameter estimates directly.

To investigate the prediction of an inverted-U relationship, I begin by semi-parametrically estimating the relationship between aggregate investment, $\Delta k_{v,2011}$, and inequality $I_{v,2009}$, among risk-sharing networks in control villages, controlling linearly for variables $\mathbf{X}$. I estimate the following specification:

$$\Delta k_{v,2011} = m(I_{v,2009}) + \boldsymbol{\gamma}_1 \mathbf{X}_{v,2009} + \boldsymbol{\gamma}_2 \mathbf{X}_{v,2007} + \epsilon_{v,2011} \tag{3.36}$$

using Robinson's (1988) partially linear estimator, where $\mathbf{X}$ contains the value of income, savings, and livestock, and network size (number of households – this is only included once since it does not vary over time).[31] All monetary values are in 2007 USD at purchasing power parity exchange rates.

Since the theory does not provide a precise measure of inequality for this test, I use two standard measures of inequality, and show that the inverted-U relationship is robust to either of these definitions. The measures I use are the interquartile range – the difference between the 75$^{\text{th}}$ and 25$^{\text{th}}$ percentiles – and standard deviation of the income distribution. These two measures have differing advantages: the interquartile range will be more robust to outliers, but the standard deviation will be better able to capture inequality at the top of the income distribution. To improve robustness to outliers, I first winsorise the income data, replacing any values below the first percentile (or above the 99$^{\text{th}}$ percentile) with the value at the first (99$^{\text{th}}$) percentile.

[31] Precisely, first $\Delta k_{v,2011}$, and each element of $\mathbf{X}_{v,2009}$ and $\mathbf{X}_{v,2007}$, are each non-parametrically regressed on $I_{v,2009}$. For eah variable $Z \in \{\Delta k_{v,2011}, \mathbf{X}_{v,2009}, \mathbf{X}_{v,2007}\}$ a residual $\eta_z = Z - \widehat{m_z(I_{v,2009})}$ is calculated. Then regression of $\eta_{\Delta k}$ on the η_X variables recovers estimates of $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$. Finally, non-parametric regression of $\Delta k_{v,2011} - \hat{\boldsymbol{\gamma}}_1 \mathbf{X}_{v,2009} - \hat{\boldsymbol{\gamma}}_2 \mathbf{X}_{v,2007}$ on $I_{v,2009}$ provides an estimate of the local mean.

Figure 3.5 shows this relationship graphically for the interquartile range, with the best fitting quadratic overlaid. An inverted-U shape is clearly visible. Figure 3.11 shows the relationship again, now using standard deviation as the measure. Again the inverted-U shape is clear, and the relationship looks close to quadratic. One-third of networks had realised income inequality that is past the peak level for investment.

Since the relationship looks well-approximated by a quadratic, I next estimate the following specification, still using only risk-sharing networks in control villages, and with variables as before:

$$\Delta k_{v,2011} = \alpha_0 + \beta_1 I_{v,2009} + \beta_2 I^2_{v,2009} + \boldsymbol{\gamma}_1 \mathbf{X}_{v,2009} + \boldsymbol{\gamma}_2 \mathbf{X}_{v,2007} + \varepsilon_{v,2011} \tag{3.37}$$

This estimates parametrically a quadratic relationship between inequality and investment.

Table 3.5, Columns (1)-(3) show that for both measures of inequality, and across all specifications, the coefficient on the linear term is positive and the coefficient on the quadratic term is negative, consistent with the predicted inverted-U shape. Column (4) replaces the inequality measure with inequality in *residualised* income. Household income is first regressed on household size and number of cows, goats, and chickens, to control for permanent differences in household incomes. Inequality is then calculated using the residuals from these regressions. Residualisation changes the magnitudes of the coefficients, but again the same shape emerges: a positive coefficient on the linear term, and a negative one on the quadratic term.

Finding robust evidence of this inverted-U shape relationship justifies the choice of limited commitment as the relevant friction in my model. Alternative frictions used in the literature on risk sharing, such as hidden action and hidden income, wouldn't give generate this prediction.

Network Size

I next test the prediction that investment is increasing in network size. The program does not provide variation in network size, so I cannot use it directly to provide exogenous variation here. Instead I perform two tests, one using the control villages and the other using the treated villages.

First, using only the control villages, I study the relationship between investment, $\Delta k_{v,2011}$, and network size, N_v. I begin by estimating the relationship between these non-parametrically, using kernel-weighted local mean regression:

$$\Delta k_{v,2011} = m(N_v) + \epsilon_{v,2011} \tag{3.38}$$

Figure 3.12 plots the relationship: it is increasing and approximately linear.

I therefore estimate linearly the relationship between between investment, $\Delta k_{v,2011}$, and network size, N_v, controlling for the value of income, savings, and livestock, and a quadratic in income inequality. All monetary values are in 2007 USD at purchasing power parity exchange rates. This estimation equation is given by:

$$\Delta k_{v,2011} = \alpha_0 + \beta_1 N_v + \boldsymbol{\gamma}_1 \mathbf{X}_{v,2009} + \boldsymbol{\gamma}_2 \mathbf{X}_{v,2007} + \varepsilon_{v,2011} \tag{3.39}$$

Table 3.6, Column (1) shows the unconditional relationship between network size and investment, providing a parametric estimate of the relationship seen in Figure 3.12. Columns (2) and (3) again add controls and then also district fixed effects. Throughout the relationship remains positive and weakly significant. For an additional five households in the risk-sharing network (a 10pp increase in network size), there is a \$250 increase in investment, which is the value of one cow.

This provides some evidence against a group-level poverty trap driven by coordination failure: larger group sizes might be expected to find cooperation more difficult, in which case the relationship should be negative (Murphy et al., 1989).

A second way to test this relationship, now using the treated sample, is to note that the prediction of the model is that the threshold level of income needed for investment to be possible should be declining in network size. To test this, I re-estimate the local mean regression from Equation 3.33, which showed how additional investment varies with the value of the capital injection provided, but splitting networks into above and below median.

Figure 3.6 shows the estimated local mean for each group. Here one can see visually that the point at which investment begins increasing is at a lower level of asset transfers for larger networks than it is for smaller ones, as suggested by the model. However, no formal way to test the difference in the thresholds has yet been developed, so this should be considered merely indicative.

3.4.3 Evidence for Alternative Explanations

I consider two alternative explanations. First, I investigate whether the network-level poverty trap could be generated by some form of increasing returns to cows at the network level. Second, I study whether asset transfers caused a non-linear effect via some other channel. Specifically I consider general equilibrium and aspirations.

Increasing Returns

The classic model of a group-level poverty trap is the 'big push model' of Rosenstein-Rodan (1943) (later formalised by Murphy et al., 1989). The key mechanism underlying it is that the return to investment is *increasing* in the number of other agents who engage in investment. Whilst the original model is motivated by concerns about industrial structure, and generates the poverty trap through demand, which are not relevant in a village economy, network-level increasing returns might still exist for a number of other reasons.[32] One possibility is group level fixed costs. For example, the price of milk may be higher (or even just more stable) in nearby markets than within the village, but there is a fixed cost of travel so that it is only productive if enough milk is being taken. Another explanation might be that there is learning across households: the more households that engage in livestock rearing, the more sources of information and advice there are, helping to better look after the cows. Such non-linear effects of social learning about investment are documented by Bandiera and Rasul (2006).

These mechanisms are distinct from my model. In my model households are unable to invest due to constraints, namely the inability to commit to future transfers, but the returns from investment are independent of the number of investments. A direct test of these increasing-returns based alternatives, is to see whether the return on cows is increasing in the number of households that received cows.

To test this, I estimate non-parametrically, again using a kernel weighted local mean smoother, the mean return on a cow against the number of eligible.[33] Since the value of capital provided to an eligible household is fixed, the aggregate capital injection maps linearly to the number households that receive cows.

From Figure 3.13 it can be seen that the mean return on cows appears to be declining in the number of cows transfered, at low number of transfers, and then to be flat and stable. These results are inconsistent with a story of increasing returns, ruling out the possibility that the observed poverty trap could be driven by network level increasing returns.

[32] The motivation for the original model is that production using 'modern' techniques involves fixed costs (e.g. administration of factories) but has a higher productivity than 'traditional' production. The existence of a fixed cost means that investment in modern production methods is only profitable if demand for output is high enough. Decisions about whether to invest are made sector-by-sector within the economy. If a sector invests, it becomes more productive and pays its workers more (the sector is made up of competitve firms, so that wage equals marginal return). However, the workers spend their income equally across all sectors. So a single sector investing may not generate enough additional demand for its own output to cover the fixed cost of investment. Only if a large enough share of sectors coordinate and invest simultaneously, will the increase in aggregate demand be enough to justify the investment. Hence a poverty trap may exist if sectors cannot coordinate on investment.

[33] Increasing returns mean that the return on the *marginal* cow is higher than on the previous cow, in which case the mean return should also be rising.

Prices and Aspirations

If real returns to cows are unchanged, an alternative explanation for the increase in investment might be that some other channel is activated once a sufficiently large number of households receive transfers. This could generate a threshold effect.

The first possibility is that general equilibrium effects might occur in some non-linear way. An immediate piece of evidence that suggests this is unlikely to be the case is that the threshold relationship documented is in terms of the *aggregate* number of households/value of capital provided, consistent with the model. General equilibrium effects, by contrast, should depend instead on the *share* of households treated. Figure 3.15 shows the non-parametric estimate of Equation 3.33 but where the independent variable of interest is the share of poor households who are treated. Plotted against the share treated, there does not appear to be any clear relationship. This provides additional evidence against a model of aggregate demand spillovers, as in Murphy et al. (1989).

An alternative way to test for general equilibrium effects is to estimate how prices vary with the value of the capital injection. Whilst this is more direct test, it requires us to know in what markets to look, and to have good measures of prices in those markets. Three possible prices of interest are the price of milk, which is the output price for cow owners; the price of cows, which is the cost of additional investment; and the wage, which is both the source of income for investment and the opportunity cost of time spent looking after cows.[34]

To test empirically whether either of these effects can explain the non-linearity in investment, I estimate the following specification using the full sample (treated and control villages):

$$\begin{aligned}\Delta\text{Channel}_{v,2009} = \alpha_0 + \alpha_1 T + \beta_1 \Delta k_{v,2009} + \delta_1 T_v \Delta k_{v,2009}.\mathbf{1}\{\Delta k_{v,2009} < 3500\} \\ + \delta_2 T_v \Delta k_{v,2009}.\mathbf{1}\{\Delta k_{v,2009} \geq 3500\} + \gamma_1 X_{v,2009} + \epsilon_{v,2009}\end{aligned} \tag{3.40}$$

where $\Delta\text{Channel}_{v,2009}$ measures the change, between 2007 and 2009, in the price being considered and other variables are as before.

The results are shown in Table 3.7 Columns (1)-(3). The results show no effect of the program on the price of milk in 2009; a reduction in the value of cows on average, but with no threshold effect; and an increase in the average wage, again with no threshold effect. Hence none of these markets appear to be the channel through which any general equilibrium effects could be driving the threshold in aggregate investment.

[34]Bandiera et al. (forthcoming) note that many individuals appear to be underemployed, having many days a year on which they cannot find work, due to the seasonality of labour demand. In this case, the wage may not be a good measure of the opportunity cost of time, as individuals are constrained in the amount of labour they can supply, due to insufficient demand.

A second possible source of non-linearity might be driven by changes in aspirations. The worry would be that there is a non-linear increase in the demand for cows as the number of neighbours owning cows rises. This could happen because households perceptions of the return on cows increases in the prevalence of ownership, or because households receive direct utility from cow ownership – beyond the financial returns – and this rises when ownership becomes more prevalent, as in a model of 'Keeping up with the Joneses'.

Table 3.7 Columns (4) shows the results of estimating Equation 3.40 with the change, between 2007 and 2009, in the share of ineligible poor households without livestock in 2009 who aspire to own livestock as the dependent variable.[35] Although the program raises aspirations on average, there is again no evidence of any threshold effect.

3.5 Conclusion

Poor households often do not undertake profitable investments, even when they belong to networks which could pool resources to invest. This paper provides a novel explanation for this puzzle: informal risk sharing can crowd out investment. To show this, I extend the classic model of risk sharing with limited commitment (Ligon et al., 2002) to also allow for lumpy investment. I show that with this addition, the model generates a poverty trap at the level of the risk-sharing network: unless aggregate income is above some threshold, the network will never be able to invest. The key insight is that once a household invests, it has less need for insurance and is more willing to walk away from the risk-sharing arrangement. This limits the investor's ability to credibly promise future transfers, so its risk-sharing partners demand transfers today, limiting investment. Hence, in the absence of institutions enforcing joint property rights, a network can be in a poverty trap despite having the resources to be able to collectively invest.

To provide evidence for this mechanism I used data from a long term, large scale randomised control trial in Bangladesh. The program randomised 1,400 villages into treatment or control status, and provided assets to the poorest households in half of these villages. I exploit variation in the aggregate level of transfers provided to risk-sharing networks to show evidence for a network-level poverty trap. Precisely, I showed empirically a threshold level of aggregate capital provision needed for the program to generate further investment: networks that received more than $3,500 were 'pushed' out of the trap. I also showed empirical evidence for additional predictions of the model, that are not implied by leading alternative models of poverty traps.

My findings have important implications for policy. The asset transfer program from which my data were drawn has now been expanded to more than half a million households in Bangladesh,

[35] Eligible households in treated villages are automatically excluded from the sample because they own cows in 2009. I exclude them from the sample in control villages to avoid composition bias.

and similar programs have begun in 33 countries worldwide. This expansion is motivated by the consistent and robust results that these programs create sustained income growth (Bandiera et al., forthcoming). My results explain *why* we see these large and long run effects, and crucially also how these programs can be further improved. If the program targeting took into account not only household characteristics, but also network characteristics and the size of the aggregate transfer being provided, more networks could be pushed out of the poverty trap, and set on a path of sustained growth.

An important direction for future research is to quantify the trade-off faced by designers of such programs between reducing poverty and growing incomes. Using the reduced form estimates of the effect of asset transfers, a budget-neutral redistribution of asset transfers in my data could generate additional investment of 44%. However, this would be achieved by reducing transfers to inframarginal networks, which are far from the poverty trap threshold, and providing them instead to marginal networks just below the threshold. Whilst this increases the number of networks pushed out of the trap, it also increases inequality *across* networks, reducing consumption in those which lose transfers. Directly estimating the parameters of the model would allow the study of the welfare gains from alternative targeting policies, taking account of this trade-off, and maximising the gains from these promising new interventions.

3.6 Tables and Figures

Table 3.1: *Household characteristics, by wealth grouping*

Sample: All villages, baseline (weighted)

Sample of households:	(A) Eligible		(B) Ineligible poor		(C) Non-poor	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
	(1)	(2)	(3)	(4)	(3)	(4)
Household size	3.29	1.70	4.09	1.61	4.74	1.76
Total income	960	690	1480	1040	2890	2440
Total consumption	1570	880	2200	1250	3740	2500
below poverty line ($1.25) [Yes=1]	.46	.50	.39	.49	.21	.41
Owns cows [Yes=1]	.06	.24	.28	.45	.63	.48
Population share	.26		.51		.23	
Observations	7,111		13,704		6,162	

Notes: *All statistics are constructed using baseline (pre-program) data for all households in both treated and control villages across the full sample. Observations are at the household level. Data were collected using a stratified random sampling scheme, so weights are used throughout to make the sample representative of the population. Wealth classes are determined using a Participatory Rural Assessment, which classifies households into four or five wealth classes. Additional criteria are used to determine which households in the poorest wealth class are 'Eligible' for treatment (Panel A). 'Ineligible Poor' (Panel B) includes all households in the bottom two wealth classes (bottom three when five wealth ranks were used) who were not eligible for the program, while 'Non-poor' (Panel C) includes all households in the top two wealth classes. All financial measures are in 2007 USD terms, converted to dollars using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. Column 1 shows the share of all links from eligible households to households in these other 'wealth class' categories.*

Table 3.2: *Risk-sharing network characteristics*

Sample: All villages, baseline

	Mean (1)	Std Dev (2)	Median (3)
Income distribution (USD):			
Aggregate	57,700	39,300	53,600
Standard deviation	721	371	643
Interquartile range	811	425	739
Value of capital injection	2,740	2200	2060
Number of households in:			
Village	87.8	16.5	86
Risk sharing network	51.7	20.1	50
Total observations	1,409	1,409	1,409

Notes: *All statistics are constructed using baseline (pre-program) data for all villages, both treated and control, across the full sample. Observations are at the risk sharing network level. Within a village, a risk sharing network is the set of low wealth households: those in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Data were collected using a stratified random sampling scheme, so weights are used throughout to aggregate household data to the risk sharing network level. All asset values are in 2007 USD terms, converted to dollars using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. The value of the capital injection is the value of the assets transfered to an eligible household (*515 *USD PPP) multiplied by the number of eligible households in the risk sharing network. The value of transfer to an eligible household is calculated by converting the value of the assets in Bangladeshi Taka (9500TK) to 2007 USD terms.*

Table 3.3: *Location of the poverty trap threshold*

Dependent Variable: Increase in Total Cow Assets (2007 USD PPP)
Standard errors (in parentheses) clustered at district level

	Unconditional (1)	With Controls (2)	With Controls & District F.E. (3)
Slope below threshold (δ_1)	-.057	-.087	-.217
	(.430)	(.454)	(.361)
Slope above threshold (δ_2)	1.57***	1.86***	2.68*
	(.55)	(.77)	(1.49)
Optimal Threshold Level	3500	3500	3500
F-statistic	5.45	5.89	3.24
5% critical value for F-statistic	5.24	5.24	5.24
Observations (Clusters)	689 (13)	689 (13)	689 (13)

Notes: **** denotes significance at 1%, ** at 5%, and * at 10% level, when treated as a standalone regression. The sample comprises low wealth households in treated villages across the full sample. Observations are at the risk-sharing network level. Within a village, a risk-sharing network is the set of low wealth households: those in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Data were collected using a stratified random sampling scheme, so weights are used to aggregate household data to the risk-sharing network level. All financial variables are in 2007 USD terms. Where necessary, they are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation. They are then converted to dollars using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. A thin tail of networks receive more than* $8000 *worth of assets. Since the density on this part of the support is low (fewer than five networks for any value of asset transfers), I trim these networks (2% of the sample). The outcome measures the increase in the value of cow assets between 2009 and 2011. Table shows regression of the outcome on the aggregate value of capital provided to the risk-sharing network in 2007. Column (1) is a regression of the outcome on a constant, and the value of aggregate asset transfers interacted with a threshold dummy (allowing the slope to vary at this point). Column (2) allows for additional controls (lagged income and asset variables, and network size). Column (3) additionally allows for district level fixed effects (so the constant is not reported). These specifications are run sequentially at different values of the threshold, varying the threshold between* $2000 *and* $5000, *at intervals of* $500. *I provide the results for the threshold among these which produced a regression with the highest F-statistic when tested against the null of no threshold. The value for this threshold, the estimated F-statistic, and the 5% critical value for this statistic (which corrects for the repeated testing, see Andrews, 1993) are provided at the bottom of the table.*

Table 3.4: *Using the program randomisation to test the poverty trap*

Dependent Variable: Increase in Total Cow Assets (2007 USD PPP)
Standard errors (in parentheses) clustered at district level

Risk-sharing Network:	(A) All Poor		(B) Whole Village	(C) Eligibles Only
	Unconditional	Controls & District F.E.	Controls & District F.E.	Controls & District F.E.
	(1)	(2)	(3)	(4)
Slope below threshold for Treated	-.242	-.110	-.785	-.076
	(.515)	(.516)	(1.112)	(.087)
Slope above threshold for Treated	1.801**	1.435*	1.517	0.316***
	(.816)	(.813)	(1.711)	(.121)
Treated	792	859	1914	-107
	(1058)	(1063)	(2299)	(187)
Observations (Clusters)	1360 (13)	1360 (13)	1360 (13)	1260 (13)

Notes: **** denotes significance at 1%, ** at 5%, and * at 10% level, when treated as a standalone regression. Observations are at the risk-sharing network level. The network in Panel A comprises low wealth households across the full sample. A household is low wealth if it is in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. The network in Panel B comprises all households across the full sample. The network in Panel C comprises eligible households across the full sample. The sample size here is smaller because 100 villages which have no eligibles are now automatically excluded. Data were collected using a stratified random sampling scheme, so weights are used to aggregate household data to the risk-sharing network level. All financial variables are in 2007 USD terms. Where necessary, they are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation. They are then converted to dollars using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. A thin tail of networks receive more than $8000 worth of assets. Since the density on this part of the support is low (fewer than five networks for any value of asset transfers), I trim these networks (2% of the sample). The outcome measures the increase in the value of cow assets between 2009 and 2011. Table shows regression of the outcome on the aggregate value of capital provided to the risk-sharing network in 2007. Column (1) is a regression of the outcome on a constant, a treatment dummy, the value of aggregate transfers that would be provided if a network is treated, and the value of aggregate transfers interacted with a treatment dummy and an indicator for whether aggregate transfers exceed $3,500. Columns (2)-(4) allow for additional controls (lagged income and asset variables, and network size), and district level fixed effects.*

Table 3.5: *How does investment vary with income inequality*

Dependent Variable: Increase in Total Cow Assets between 2009-11 (2007 USD PPP)
Sample: Control villages
Standard errors (in parentheses) clustered at district level

	Unconditional	With Controls	And District FE	Resid. Inc. Ineq.
	(1)	(2)	(3)	(4)
Panel A. Interquartile Range				
IQR of 2009 Income Distribution	8.32**	8.27**	6.48*	.270**
	(3.35)	(2.80)	(3.05)	(.123)
(IQR of 2009 Income Distribution)^2	-.003**	-.003***	-.002**	-.0001**
	(.001)	(.001)	(.001)	(.0000)
Observations (Clusters)	696 (13)	696 (13)	696 (13)	696 (13)
Panel B. Standard Deviation				
SD of 2009 Income Distribution	4.99	7.12***	6.09**	.382**
	(3.22)	(2.66)	(2.60)	(.136)
(SD of 2009 Income Distribution)^2	-.002*	-.002**	-.002**	-.0001**
	(.001)	(.001)	(.001)	(.0000)
Observations (Clusters)	696 (13)	696 (13)	696 (13)	696 (13)

Notes: **** denotes significance at 1%, ** at 5%, and * at 10% level. Standard errors clustered at the village level. Constructed using data on low wealth households in control villages across the full sample. Observations at the risk sharing network level. Within a village, a risk sharing network is the set of low wealth households: those in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Data were collected using a stratified random sampling scheme, so weights are used throughout to aggregate household data to the risk sharing network level. All financial variables are in 2007 USD terms. Where necessary, they are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation. They are then converted to dollars using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. The outcome measures the increase in the value of cow assets between 2009 and 2011. Panel A uses as the variable of interest the interquartile range of income: the difference between the* 75^{th} *and* 25^{th} *percentiles of the cross-sectional income distribution in the network in 2009. Panel B uses as the variable of interest the standard deviation of the cross-sectional income distribution in the network in 2009. In both Panels, Columns (1) shows the unconditional regression of additional investment on income inequality and income inequality squared. Column (2) includes as controls total income, total saving, and the value of cows, all in 2009 and 2007, and also network size. Column (3) includes additionally district fixed effects. Column (4) includes the same controls as Column (3), but replaces the inequality measure with inequality in* residualised *income, where household income is first regressed on household size and number of cows, goats, and chickens, to control for permanent differences in household incomes, and inequality is then calculated using the residuals from these regressions.*

Table 3.6: *How does investment vary with network size*

Dependent Variable: Increase in Total Cow Assets between 2009-11 (2007 USD PPP)
Sample: Control villages
Standard errors clustered at district level

	Unconditional (1)	With District FE (2)	With Controls (3)
Network size	52.9* (27.1)	55.4** (20.8)	44.8* (24.1)
Observations (Clusters)	696 (13)	696 (13)	696 (13)

Notes: **** denotes significance at 1%, ** at 5%, and * at 10% level. Standard errors clustered at the village level. Constructed using data on low wealth households in control villages across the full sample. Observations at the risk sharing network level. Within a village, a risk sharing network is the set of low wealth households: those in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Data were collected using a stratified random sampling scheme, so weights are used throughout to aggregate household data to the risk sharing network level. All financial variables are in 2007 USD terms. Where necessary, they are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation. They are then converted to dollars using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. The outcome measures the increase in the value of cow assets between 2009 and 2011. The interquartile range of income for a network is the difference between the* 75^{th} *and* 25^{th} *percentiles of the cross-sectional income distribution in 2009. Columns (1) shows the unconditional regression of additional investment on network size. Column (2) includes as controls total income, total saving, and the value of cows, all in 2009 and 2007, and the interquartile range (IQR) and IQR squared in 2009. Column (3) includes additionally district fixed effects.*

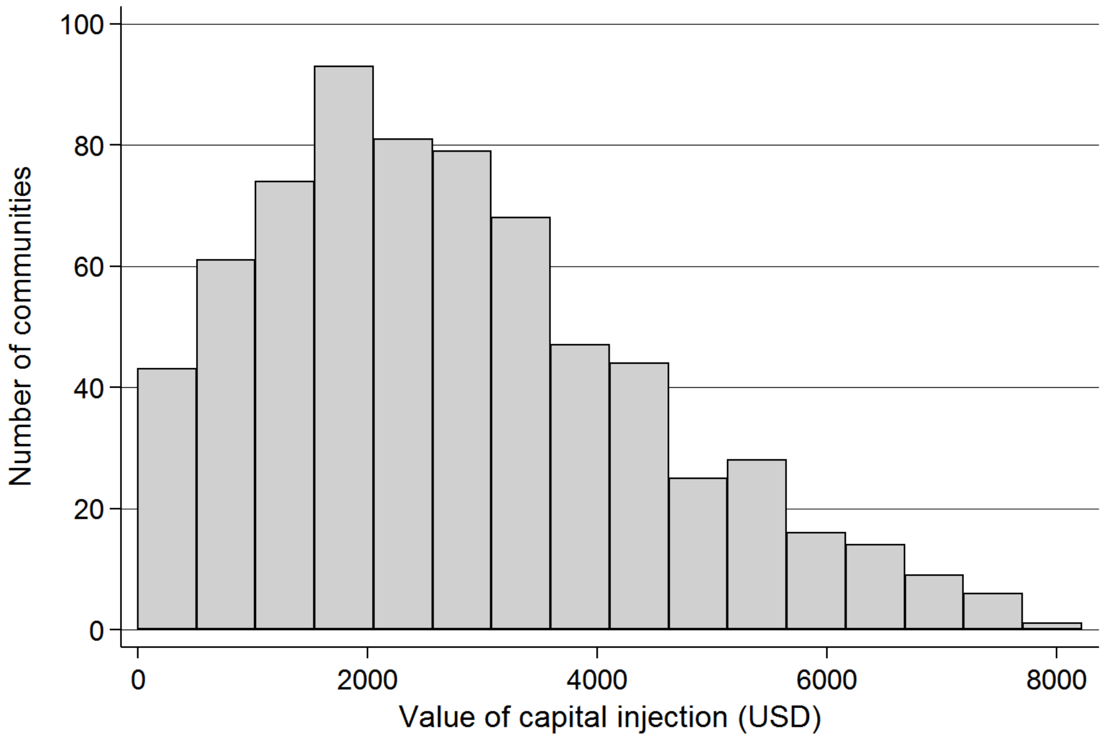
Table 3.7: *How does capital injection affect prices and aspirations*

Standard errors (in parentheses) clustered at district level

	Milk Price	Cow Price	Wage	Aspirations
	(1)	(2)	(3)	(4)
Slope below threshold for Treated	-.055	-2.17	.327	-.035
	(.034)	(12.84)	(.351)	(.010)
Slope above threshold for Treated	.005	6.58	.221	-.024
	(.031)	(9.32)	(.178)	(.015)
Treated	.110	-7.97**	.871**	.200**
	(.060)	(29.75)	(.365)	(.029)
Mean Baseline Level among Controls	1.28	475	6.32	.766
Financial Controls	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes
Observations (Clusters)	1190 (13)	1190 (13)	1190 (13)	938 (13)

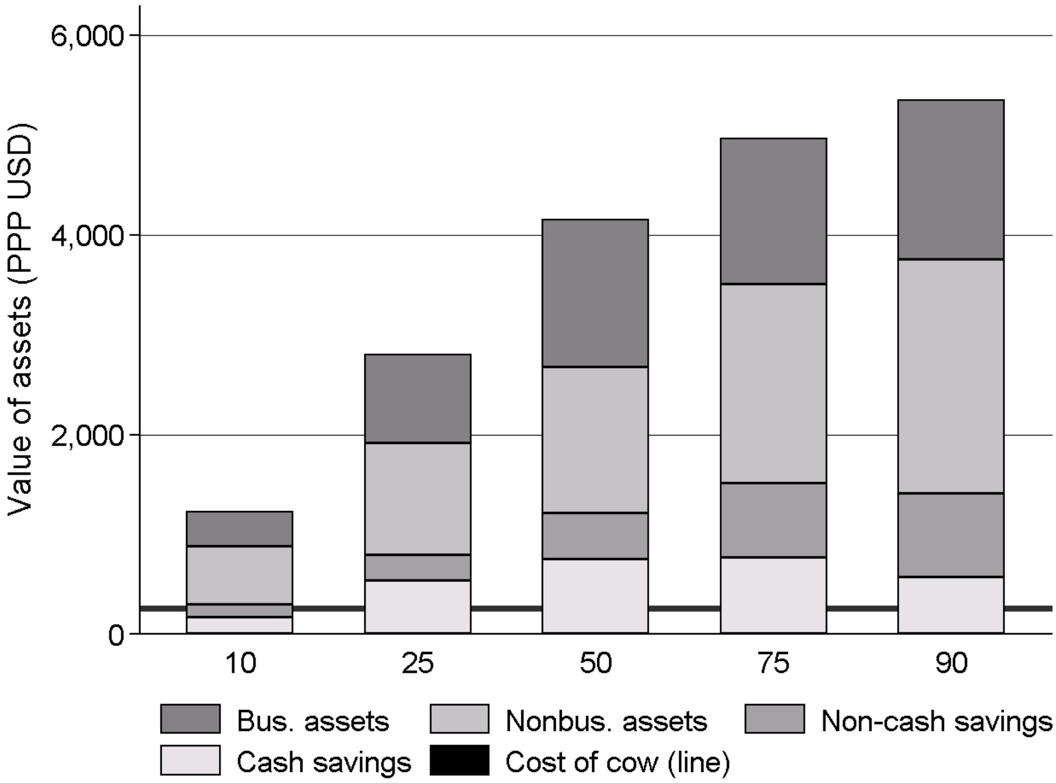
Notes: **** denotes significance at 1%, ** at 5%, and * at 10% level. The sample comprises all households in all villages across the full sample in 2009. Data were collected using a stratified random sampling scheme, so weights are used to aggregate household data to the risk-sharing network level. All financial variables are in 2007 USD terms. Where necessary, they are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation. They are then converted to dollars using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. In Column (1) the outcome measures the change in the price of milk. The price of milk is calculated by dividing for each household their total expenditure on milk by the total quantity they report purchasing, and then averaging this across households in the village. In Column (2) the outcome measures the change in the price of cows. The price of cows is calculated by dividing for each village the total reported value of cows by the total reported number of cows. In Column (3) the outcome measures the change in the wage. The wage is calculated by dividing for each household the total reported income by the total reported number of wage hours worked. In Column (4) outcome measures the change in the share of all poor households who are without cows and report aspiring to own cows. The table shows regression of the outcome in 2009 on the aggregate value of capital provided to the risk sharing network in 2007. Financial controls are the total income, total savings, and total value of cows in 2007.*

Figure 3.1: *Frequency distribution of aggregate value of capital injection provided by the program*



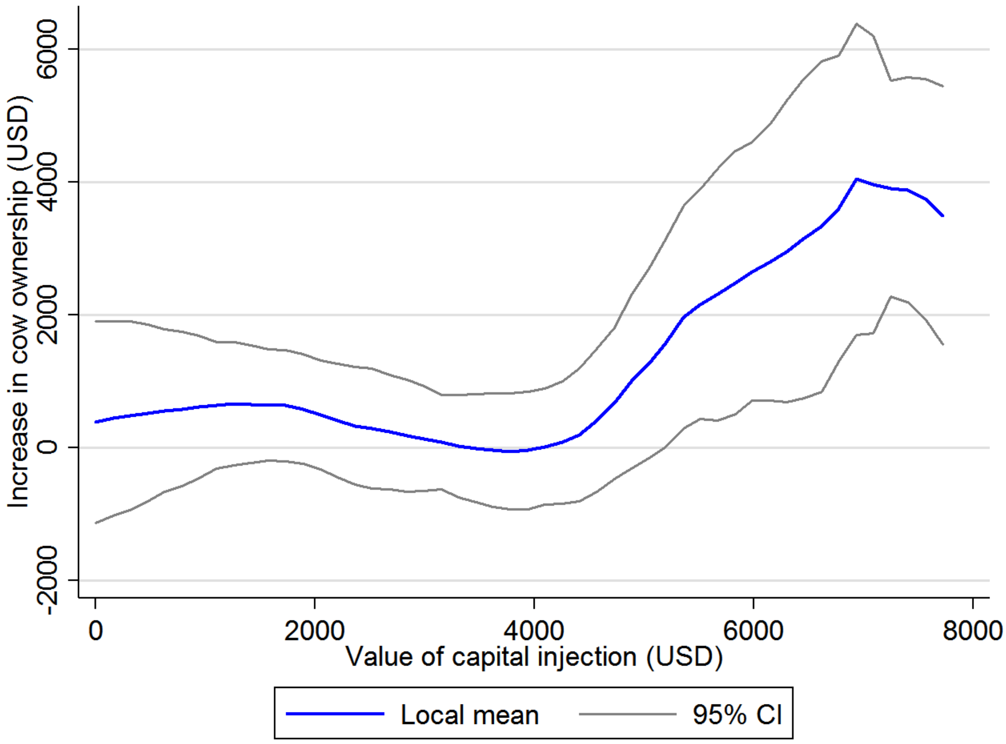
Notes: *Constructed using data on households in treated villages across the full sample. Observations at the risk-sharing network level. Within a village, a risk-sharing network is the set of households in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Data were collected using a stratified random sampling scheme, so weights are used throughout to aggregate household data to the risk-sharing network level. The value of the capital injection is the value of the assets transfered to an eligible household (515 USD PPP) multiplied by the number of eligible households in the risk-sharing network. The value of transfer to an eligible household is calculated by converting the value of the assets in Bangladeshi Taka ($9,500$ TK) to 2007 USD terms, using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. A thin tail of networks receive more than \$8,000 worth of assets. Since the density on this part of the support is low (fewer than five networks for any value of asset transfers), I trim these networks (2% of the sample).*

Figure 3.2: *Distribution of aggregate assets across risk-sharing networks, broken down by type*



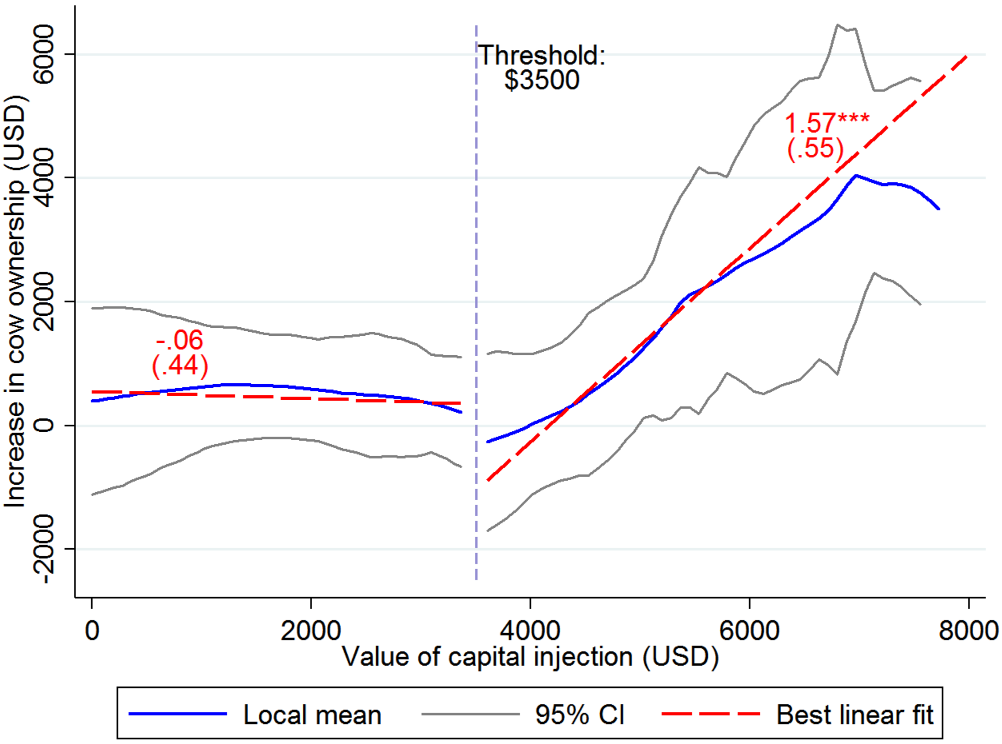
Notes. *All statistics are constructed using baseline (pre-program) data for all villages, both treated and control, across the full sample. Observations are at the risk-sharing network level. Within a village, a risk-sharing network is the set of households in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Data were collected using a stratified random sampling scheme, so weights are used throughout to aggregate household data to the risk-sharing network level. All asset values are in 2007 USD terms, converted to dollars using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. Cash savings include savings held at home, in any bank, with any NGO or microfinance institution, and with any savings guard. Non-cash savings include the value of jewellery and ceremonial sarees. Nonbusiness assets include electrical devices (radios, televisions, refrigerators), personal vehicles (bicycles, motorbikes), and furniture. Business assets include animals, farm infrastructure and machinery, and productive vehicles (rickshaw, van, cart).*

Figure 3.3: *Impact of capital injection on further investment*



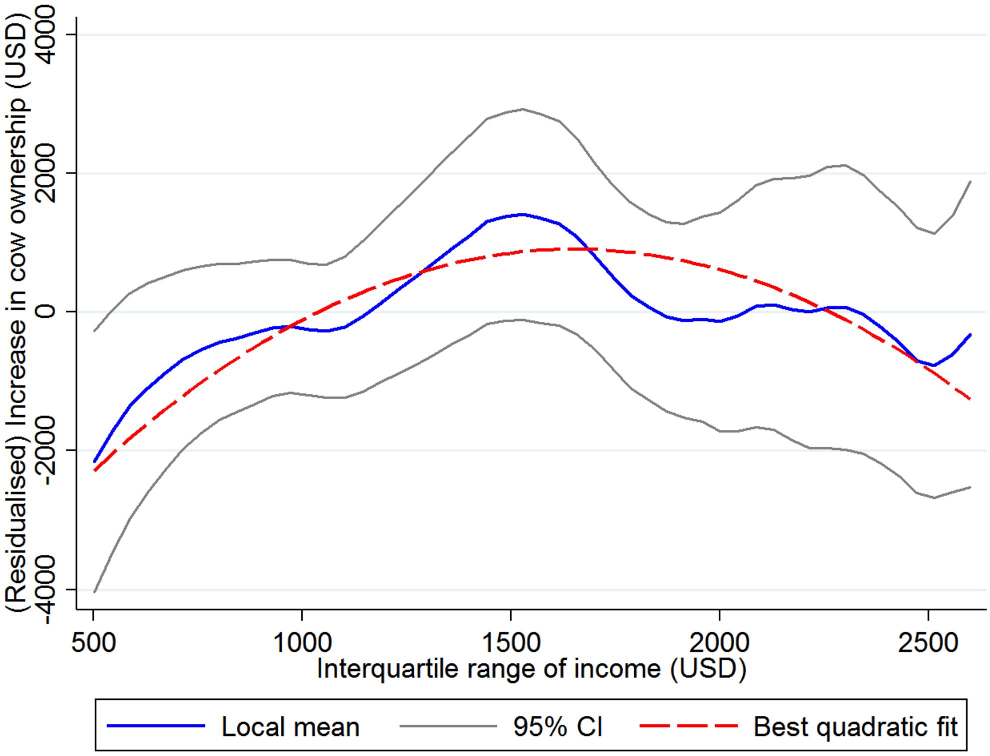
Notes: *Constructed using data on all poor households in treated villages across the full sample. Observations at the risk-sharing network level. Within a village, a risk-sharing network is the set of households in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Data were collected using a stratified random sampling scheme, so weights are used throughout to aggregate household data to the risk-sharing network level. The value of the capital injection is the value of the assets transfered to an eligible household (515 USD PPP) multiplied by the number of eligible households in the risk-sharing network. The value of transfer to an eligible household is calculated by converting the value of the assets in Bangladeshi Taka (9,500TK) to 2007 USD terms, using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. Increase in cow ownership is measured as the increase in the value of cows between 2009 and 2011 owned by all households in the risk-sharing network. The values in Bangladeshi Taka are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation, and then converted to USD PPP. A thin tail of networks receive more than $8,000 worth of assets. Since the density on this part of the support is low (fewer than five networks for any value of asset transfers), I trim these networks (2% of the sample). The graph shows the kernel-weighted local mean, estimated using an Epanechnikov kernel with bandwidth $800. The outer region provides the 95% confidence interval.*

Figure 3.4: *Impact of capital injection on further investment, either side of threshold*



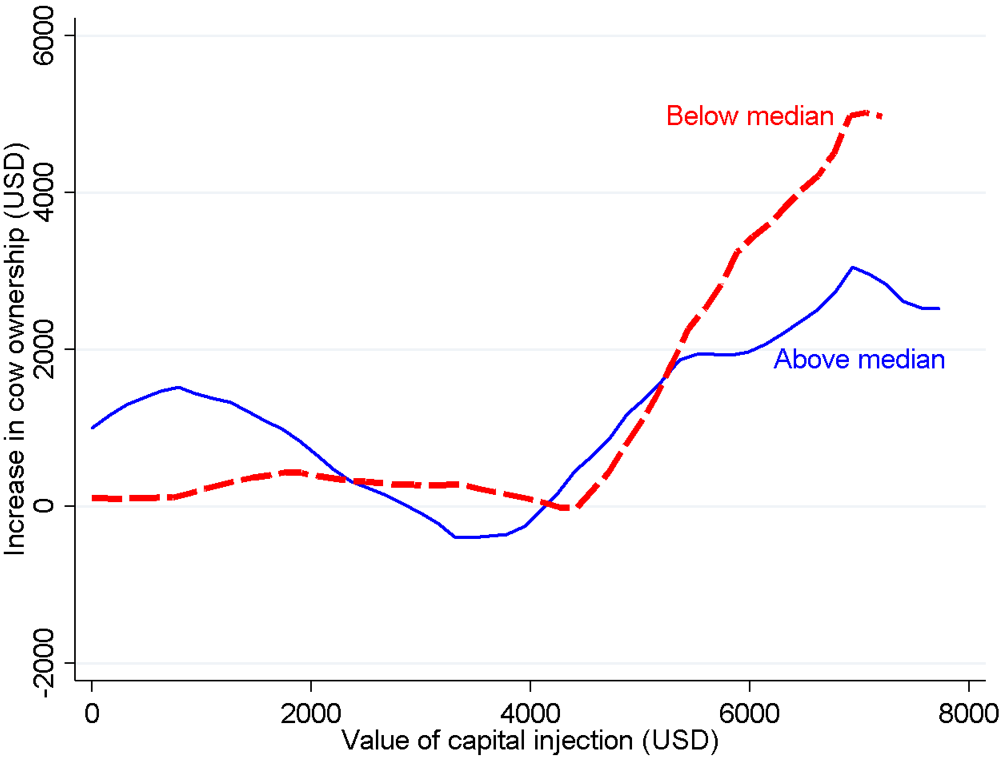
Notes: *Constructed using data on all poor households in treated villages across the full sample. Observations at the risk-sharing network level. Within a village, a risk-sharing network is the set of households in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Data were collected using a stratified random sampling scheme, so weights are used throughout to aggregate household data to the risk-sharing network level. The value of the capital injection is the value of the assets transfered to an eligible household (515 USD PPP) multiplied by the number of eligible households in the risk-sharing network. The value of transfer to an eligible household is calculated by converting the value of the assets in Bangladeshi Taka (9,500TK) to 2007 USD terms, using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. Increase in cow ownership is measured as the increase in the value of cows between 2009 and 2011 owned by all households in the risk-sharing network. The values in Bangladeshi Taka are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation, and then converted to USD PPP. A thin tail of networks receive more than $8,000 worth of assets. Since the density on this part of the support is low (fewer than five networks for any value of asset transfers), I trim these networks (2% of the sample). The figure shows non-parametrically the relationship between the increase in the aggregate value of cows in a risk-sharing network between two years and four years after transfers were made, and the value of capital provided to the network by the program. This is plotted either side of the estimated threshold of $3,500. This threshold was selected by linear regressions of investment on capital injection, at a sequence of possible values for the threshold. The most likely value for the threshold is then the proposed value in the regression which had the largest F-statistic for a change in the slope. I use the Quandt Likelihood Ratio test, as described in Section 3.4, to test for significance of the threshold. The non-parametric relationship shown is a kernel-weighted local mean, estimated using an Epanechnikov kernel with bandwidth $800. The outer region provides the 95% confidence interval. The best linear fit is plotted either side of the threshold, with slope coefficient noted and standard errors in parentheses. *** denotes significance at 1%, ** at 5%, and * at 10% level, when treated as a standalone regression.*

Figure 3.5: *Investment is an inverted-U in income inequality, as measured by interquartile range*



Notes: *Constructed using data on all poor households in control villages across the full sample. Observations at the risk-sharing network level. Within a village, a risk-sharing network is the set of households in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Data were collected using a stratified random sampling scheme, so weights are used throughout to aggregate household data to the risk-sharing network level. The interquartile range of income for a network is the difference between the* 75^{th} *and* 25^{th} *percentiles of the cross-sectional income distribution in 2009. It is converted to 2007 USD terms, using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. Increase in cow ownership is measured as the increase in the value of cows between 2009 and 2011 owned by all households in the risk-sharing network. The values in Bangladeshi Taka are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation, and then converted to USD PPP. Residualised increase in cow ownership is the residuals from first regressing increase in cow ownership on total income, total saving, and the value of cows in 2009 and 2007. The kernel-weighted local mean, estimated using an Epanechnikov kernel with bandwidth $140. The outer region provides the 95% confidence interval.*

Figure 3.6: *Higher investment threshold for smaller network size*



Notes: *Constructed using data on all poor households in treated villages across the full sample. Observations at the risk-sharing network level. Within a village, a risk-sharing network is the set of households in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Data were collected using a stratified random sampling scheme, so weights are used throughout to aggregate household data to the risk-sharing network level. The value of the capital injection is the value of the assets transfered to an eligible household (*515 *USD PPP) multiplied by the number of eligible households in the risk-sharing network. The value of transfer to an eligible household is calculated by converting the value of the assets in Bangladeshi Taka (*9,500*TK) to 2007 USD terms, using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. Increase in cow ownership is measured as the increase in the value of cows between 2009 and 2011 owned by all households in the risk-sharing network. The values in Bangladeshi Taka are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation, and then converted to USD PPP. A thin tail of networks receive more than* $8,000 *worth of assets. Since the density on this part of the support is low (fewer than five networks for any value of asset transfers), I trim these networks (2% of the sample). Network size is measured as the number of households in the network. The data are split into above and below median network size. The kernel-weighted local means are then plotted separately for each case, estimated using an Epanechnikov kernel with bandwidth $800.*

3.7 Theoretical Appendix

3.7.1 Proof that one can take the maximum inside the integral

To show that the definitions of $\mathcal{U}(\bar{\boldsymbol{\omega}}', \mathbf{1})$ in Equations 3.9 and 3.10 are equivalent, I first note that:

$$V(t; \boldsymbol{\omega}'(t; \bar{\boldsymbol{\omega}}'), \mathbf{1}) \leq \sup_{\mathbf{w}'(\cdot, \bar{\boldsymbol{\omega}})} V(t; \mathbf{w}'(t, \bar{\boldsymbol{\omega}}'), \mathbf{1}) \qquad \forall t, \bar{\boldsymbol{\omega}}', \boldsymbol{\omega}'(t; \bar{\boldsymbol{\omega}}') \tag{3.41}$$

Integrating both sides:

$$\int V(t; \boldsymbol{\omega}'(t; \bar{\boldsymbol{\omega}}'), \mathbf{1}) \, \mathrm{d}F_{\mathbf{Y}'}(t) \leq \int \sup_{\mathbf{w}'(\cdot, \bar{\boldsymbol{\omega}})} V(t; \mathbf{w}'(t, \bar{\boldsymbol{\omega}}'), \mathbf{1}) \, \mathrm{d}F_{\mathbf{Y}'}(t) \quad \forall \bar{\boldsymbol{\omega}}', \boldsymbol{\omega}'(t; \bar{\boldsymbol{\omega}}') \tag{3.42}$$

Then:

$$\sup_{\mathbf{w}'(\cdot, \bar{\boldsymbol{\omega}})} \int V(t; \mathbf{w}'(t; \bar{\boldsymbol{\omega}}'), \mathbf{1}) \, \mathrm{d}F_{\mathbf{Y}'}(t) \leq \int \sup_{\mathbf{w}'(\cdot, \bar{\boldsymbol{\omega}})} V(t; \mathbf{w}'(t, \bar{\boldsymbol{\omega}}'), \mathbf{1}) \, \mathrm{d}F_{\mathbf{Y}'}(t) \qquad \forall \bar{\boldsymbol{\omega}}' \tag{3.43}$$

Note also that on the LHS of Equation 3.43, choice of $\mathbf{w}'(\cdot, \bar{\boldsymbol{\omega}})$ is essentially choice of the integrand $V(\cdot; \mathbf{w}'(\cdot; \bar{\boldsymbol{\omega}}'), \mathbf{1})$ to maximise the value of the integral. The choice of V that was made on the RHS is still available, although in principle some other choice could be better. Hence:

$$\sup_{\mathbf{w}'(\cdot, \bar{\boldsymbol{\omega}})} \int V(t; \mathbf{w}'(t; \bar{\boldsymbol{\omega}}'), \mathbf{1}) \, \mathrm{d}F_{\mathbf{Y}'}(t) \geq \int \sup_{\mathbf{w}'(\cdot, \bar{\boldsymbol{\omega}})} V(t; \mathbf{w}'(t, \bar{\boldsymbol{\omega}}'), \mathbf{1}) \, \mathrm{d}F_{\mathbf{Y}'}(t) \qquad \forall \bar{\boldsymbol{\omega}}' \tag{3.44}$$

Combining Equations 3.43 and 3.44 it must be that the two sides are equal, so taking the integral of the maximum, as in Equation 3.10, gives the same result as taking the maximum of the integral, Equation 3.9.

3.7.2 Proof of Lemma 3.1: Weakly dominant investment allocation rule

Suppose this rule were not weakly dominant. Then for some income shock $\mathbf{y}$, and some desired number of investments, $\Delta k(\mathbf{y}, \boldsymbol{\kappa})$, there exists an alternative investment allocation strategy which is *strictly* better than the one proposed i.e. there exists a pair of households i, j such that $\tilde{\omega}^i(\mathbf{y}, \boldsymbol{\kappa}) > \tilde{\omega}^j(\mathbf{y}, \boldsymbol{\kappa})$, $\Delta\kappa^i = 0$, $\Delta\kappa^j = 1$.

To show this cannot be the case, first note that households draw from a common income distribution, so the probability of any income draw is as likely for i and j. I define $\tilde{y}^i(\boldsymbol{\kappa})$ implicitly as $\tilde{\omega}^i(\mathbf{y}, \boldsymbol{\kappa}) = \Omega^i(\tilde{y}^i, 1)$, the value of income such that the household's limited commitment constraint when invested just binds. In effect this is the maximum income draw the household could get, if assigned an investment, before its promised utility would need to be increased to keep it in the risk-sharing arrangement. Since $\Omega(y^i, 1)$ is increasing in y^i, $\tilde{y}^i$ is increasing in $\tilde{\omega}^i(\mathbf{y}, \boldsymbol{\kappa})$. Then $\tilde{\omega}^i(\mathbf{y}, \boldsymbol{\kappa}) > \tilde{\omega}^j(\mathbf{y}, \boldsymbol{\kappa})$ implies that $\tilde{y}^i > \tilde{y}^j$. Hence there exists a region of individual income shock, $[\tilde{y}^j, \tilde{y}^i]$ such that an income of this size to household i would not increase the utility it is promised, but a shock of this size to j would increase the utility it is promised. There is no income shock that could occur in the next period that would increase the promised utility to i, that wouldn't increase the promised utility to j at least as much if j received that income shock. Then, since providing utility is costly to the planner, this allocation rule is more costly than instead allocating the investment to i.

3.7.3 Proof of Proposition 3.1: Threshold income level for investment

The proof of Proposition 3.1 involves three steps. I first show that when the planner is choosing $\Delta k(\mathbf{y}, \boldsymbol{\kappa})$ optimally, $\frac{\partial V_{\Delta k}}{\partial Y} > 0$. Next I show that $\frac{\partial V_{\Delta k}}{\partial Y} > \frac{\partial V_{\Delta k-1}}{\partial Y} > 0$, so that $V_{\Delta k-1}$ and $V_{\Delta k}$ cross at most once. Finally I show that $V_{\Delta k-1}$ and $V_{\Delta k}$ do cross at least once, and hence there is a unique $\widehat{Y}_{\Delta k}$ s.t. $\Gamma_{\Delta k}(\widehat{Y}_{\Delta k}) \equiv V_{\Delta k-1}(\widehat{Y}_{\Delta k}) - V_{\Delta k}(\widehat{Y}_{\Delta k}) = 0$.

Conditional value functions are increasing in Y

I want to show that $\frac{\partial V_{\Delta k}}{\partial Y} > 0$. Taking the derivative of the conditional value function $V_{\Delta k}$ wrt Y, and applying the envelope theorem, I get:

$$\begin{aligned} \frac{\partial V_{\Delta k}}{\partial Y} &= t^1 \frac{\mathrm{d}u(c^1)}{\mathrm{d}c^1} + \sum_{i=2}^{N} t^i \lambda^i \frac{\mathrm{d}u(c^i)}{\mathrm{d}c^i} \\ &= \frac{\mathrm{d}u(c^1)}{\mathrm{d}c^1} \end{aligned}$$

where for notational convenience I define $t^1 = \left(1 - \sum_{i=2}^{N} t^i\right)$, and the second equality comes from use of the FOCs wrt τ^i. Hence from the properties of $u(\cdot)$, $\frac{\partial V_{\Delta k}}{\partial Y} > 0$.

Slopes of conditional value functions are increasing in Δk

By the budget constraint, aggregate consumption $C = \sum_{j=1}^{N} c^j$ is total income less spending on investment: $C = Y + kR - \Delta k d$. From the first order conditions wrt τ^i, c^i is strictly increasing in c^1, so all households' consumptions must be strictly increasing in aggregate consumption. Hence, since aggregate consumption is strictly decreasing in Δk, the number of investments, $c^1_{\Delta k} < c^1_{\Delta k-1}$. Then by concavity of $u(\cdot)$, $\frac{\partial V_{\Delta k-1}}{\partial Y} - \frac{\partial V_{\Delta k}}{\partial Y} = u'(c^1_{\Delta k-1}) - u'(c^1_{\Delta k}) < 0$. As the conditional value function when there are Δk investments is always strictly steeper than the value function associated with $\Delta k - 1$ investments, and the value functions are continuous (again inherited from properties of $u(\cdot)$) they can cross at most once.

Conditional value functions cross at least once

To see that the conditional value functions do have at least one crossing, I show the limits of their difference as aggregate income falls and rises.

As aggregate income falls towards Δkd, the cost of making Δk investments, the value of making $\Delta k - 1$ investments today remains positive. However the value of making Δk investments goes to negative infinity as there are no resources left for consumption (since $u(\cdot)$ satisfies the Inada conditions). Hence the difference in value becomes infinite:

$$\lim_{Y \to d} V_{\Delta k-1} - V_{\Delta k} = \infty$$

Conversely, as aggregate income this period rises towards infinity, the difference in utility today between investing and not investing goes to zero. Hence the difference in the conditional value functions is just the difference in the value between having $k + \Delta k - 1$ or having $k + \Delta k$ investments (collectively), where $k \equiv |\boldsymbol{\kappa}_k|$ is the number of existing investments. Since the value is increasing in the number of investments, the difference in values is negative.

$$\lim_{Y \to \infty} V_{\Delta k-1} - V_{\Delta k} = \beta \mathbb{E} \left[V(\boldsymbol{\omega}', \boldsymbol{\kappa}_{k+\Delta k-1}) - V(\boldsymbol{\omega}', \boldsymbol{\kappa}_{k+\Delta k}) \right] < 0$$

Hence since the conditional value functions are continuous in Y, they must cross. Then, since I showed earlier that $\frac{\partial V_{\Delta k}}{\partial Y} > \frac{\partial V_{\Delta k-1}}{\partial Y} > 0$, there can be at most one crossing of the value functions.

3.7.4 Proof of Lemma 3.2: The Thresholds are Decreasing in Capital

I first show that under full commitment, the value function $V_{\Delta k}(\kappa_k)$ has *increasing differences* in $(\Delta k, k)$:

$$V_{\Delta k+1}(\kappa_k) - V_{\Delta k+1}(\kappa_{k-1}) > V_{\Delta k}(\kappa_k) - V_{\Delta k}(\kappa_{k-1})$$

To see this I expand the conditional value functions

$$V_{\Delta k+1}(\kappa_k) - V_{\Delta k+1}(\kappa_{k-1}) = u(c^1_{\Delta k+1,k}) - u(c^1_{\Delta k+1,k-1}) + \beta \mathbb{E} \left[V(\kappa_{k+\Delta k+1}) - V(\kappa_k) \right]$$

and

$$V_{\Delta k}(\kappa_k) - V_{\Delta k}(\kappa_{k-1}) = u(c^1_{\Delta k+1,k}) - u(c^1_{\Delta k+1,k-1}) + \beta \mathbb{E} \left[V(\kappa) - V(\kappa_{k+\Delta k-1}) \right]$$

Hence the double difference, $[V_{\Delta k+1}(\kappa_k) - V_{\Delta k+1}(\kappa_{k-1})] - [V_{\Delta k}(\kappa_k) - V_{\Delta k}(\kappa_{k-1})]$ gives

$$[u(c^1_{\Delta k+1,k}) - u(c^1_{\Delta k+1,k-1})] - [u(c^1_{\Delta k,k}) - u(c^1_{\Delta k,k-1})] + \beta\mathbb{E}\left[V(k+\Delta k+1) - V(k+\Delta k-1)\right]$$

Letting $C := [Y + kR - \Delta kd]$ denote aggregate consumption, I note that household consumption is proportion to aggregate consumption. The increase in aggregate consumption when initial capital increases from $k-1$ to k is independent of the number of investments made today, i.e. $C_{\Delta k+1,k} - C_{\Delta k+1,k-1} = C_{\Delta k,k} - C_{\Delta k,k-1} = R$. Then the difference in consumption is the same in both the first and second set of square brackets above, but by concavity of the utility function $u(\cdot)$ the gain in utility from this increase is higher at lower levels of consumption i.e. when investment is higher. Hence $u(c^1_{\Delta k+1,k}) - u(c^1_{\Delta k+1,k-1}) > u(c^1_{\Delta k,k}) - u(c^1_{\Delta k,k-1}) > 0$, so the first two terms are (together) strictly positive. Since value functions are increasing in the level of capital, the final term is also strictly positive, so the value function exhibits increasing differences in $(\Delta k, k)$. Then, since investment and capital are positive integers, $\Delta k, k \in \mathbb{Z}_+$, the set of possible values for $(\Delta k, k)$ form a *lattice*.[36] Finally, as $V_{\Delta k}(\kappa_k)$ has increasing differences in $(\Delta k, k)$, and the set of possible $(\Delta k, k)$ form a lattice, $V_{\Delta k}(\kappa_k)$ is supermodular in $(\Delta k, k)$. Hence by application of *Topkis' Theorem* (Topkis, 1978), the optimal choice of Δk is non-decreasing in k at any given income. This implies the threshold level of income needed for Δk investments to be optimal, $\widehat{Y}_{\Delta k}$, is weakly lower as k increases i.e. $\mathbf{D}_k \widehat{Y}^{\text{FC}}_{\Delta k} < 0$.

3.7.5 Proof of Proposition 3.2: Investment is an Inverted-U in Income Inequality

The proof is in two parts. First I show that, starting with a completely equal income distribution, in which no-one wants to invest in autarky, increasing income inequality leads to increased investment.[37] Next I show that if inequality rises too much, this leads to declining investment.

The key to the proof is to note that there exists a threshold level of income $\tilde{\tilde{y}}$ such that if individual income $y^i > \tilde{\tilde{y}}$, then it will be optimal for the planner to allow i to invest (when there is limited commitment), even though this would be suboptimal with full commitment. To see why such a threshold exists, note that for an uninvested household whose limited commitment constraint binds, the planner needs to provide $\Omega(y^i, 0)$ in the cheapest way. To meet this utility promise the planner can provide transfers today, or expected promised utility for the future. Providing investment is one way of providing utility in the future, because it raises the household's outside option, so increases the amount of utility the household can expect to receive. It also reduces the cost of providing this future utility, by increasing future income. Let $\tilde{\tilde{y}}(\mathbf{y}^{-i}, \boldsymbol{\kappa})$

[36] A lattice is a partially-ordered set where for any pair of elements in the set, the least upper bound and greatest lower bound of the elements are also in the set. For more details see Milgrom and Shannon (1994).

[37] Since incomes are equal, in autarky either all households do or don't want to invest. If all households already can invest, then there is no poverty trap. This circumstance is not relevant to my empirical context.

be defined implicitly as

$$\Omega(\tilde{\tilde{y}}, 0) = u(\tilde{\tilde{y}} - \tau_0^i(\mathbf{y}, \boldsymbol{\kappa})) + \beta\bar{\omega}_0^i(\mathbf{y}, \boldsymbol{\kappa}) = u(\tilde{\tilde{y}} - d - \tau_1^i(\mathbf{y}, \boldsymbol{\kappa})) + \beta\bar{\omega}_0^i(\mathbf{y}, \boldsymbol{\kappa})$$

where subscripts 0 and 1 denote the optimal decisions when investment is forced not to/forced to take place for i. As earlier, the marginal utility of consumption today is greater under investment than non-investment, so there is a single crossing point $\tilde{\tilde{y}}(\mathbf{y}^{-i}, \boldsymbol{\kappa})$ moving from the planner optimally choosing non-investment to optimally choosing investment as income for i rises, holding others' incomes constant.

Consider how this threshold changes as the income for some other household, y^j, is reduced. By reducing another household's income, the planner desires more transfers to take place to that household, but these come not only from i but also other households. Hence $\tilde{\tilde{y}}$ rises more slowly than y^j falls.

Next I consider what the existence of this threshold means for income inequality. Starting with an equal income distribution, consider performing a mean preserving spread, decreasing y^j by δy, and increasing y^i by the same amount. Since $\tilde{\tilde{y}}$ rises more slowly than y^i, at some point $y^i = \tilde{\tilde{y}}$, so that the planner now allows household i to invest. Hence investment is initially increasing in income inequality.

Now consider repeating this for other households. More of the households that have higher incomes may initially be taken over the investment threshold. But, doing this raises the threshold, which may reduce the effect of inequality on increasing investment. Eventually the lower bound on household income will mean that mean preserving spreads would be between households who are above the threshold, pushing some of them back below the threshold. In the limit where only one household has (almost) all the income, the total number of investments will fall back to only one.

3.7.6 Proof of Proposition 3.3: Investment is Increasing in Network Size

Increasing network size both lowers the threshold level of aggregate income needed for investment ($\widehat{Y}_1$), and increases the the probability that aggregate income in the network exceeds this threshold.

To see the first effect, I write individual income y^i as the sum of common and idiosyncratic components, η and ϵ^i. Then aggregate income is $Y = N\eta + \sum_i \epsilon^i$, and mean income is $\bar{y} = Y/N = \eta + N^{-1}\sum_i \epsilon^i$. The variance of mean income is $\sigma_\eta^2 + N^{-1}\sigma_\epsilon^2$, which is declining in N.[38]

[38]Note that the decomposition of individual income into common and idiosyncratic components makes use of the symmetry of individuals. Without this there might be some more complex correlation structure across incomes. The only essential point here is that as N increases, the variance of mean income declines. This would still be true as long as the income of 'new' households added to the network is not perfectly correlated with the sum of income of all existing households.

With full commitment, a household's consumption is a fixed share, α^j, of mean income, so the variance of consumption is proportional to the variance of mean income. Since the planner is risk-averse, reducing the variance of his consumption improves his utility i.e. $\mathbf{D}_N V > 0$, where $\mathbf{D}_N$ is the finite difference operator (the discrete analogue of the derivative) with respect to N. Also due to risk aversion, this effect is larger when consumption is lower, i.e. under $V_{\Delta k}$ rather than $V_{\Delta k-1}$. Hence $\mathbf{D}_N \Gamma = \mathbf{D}_N V_{\Delta k-1} - \mathbf{D}_N V_{\Delta 1} < 0$. Then by the implicit function theorem, since we already saw $\partial\Gamma/\partial Y < 0$, $\operatorname{sgn}(\mathbf{D}_N \widehat{Y}_1) = \operatorname{sgn}(\mathbf{D}_N \Gamma) < 0$. This means that the threshold level of aggregate income needed is declining as the number of households increases.

With limited commitment, a household's consumption share is not fixed, as it is adjusted when any households' limited commitment constraint binds. To see that consumption still becomes less volatile as group size increases, consider combining two groups of size N which receive the same common shock. The planner could always decide to make no transfers across the two groups, as though they remained separate. However, in general it will be beneficial to make some cross group transfers, as this will allow additional smoothing i.e. as group size increases, consumption will vary less for given income realisations. Combined with the above result that aggregate income will vary less, this again implies that $\widehat{Y}_1$ will decline with N.

The second effect is immediate from the definition of aggregate income. Since $Y = N\eta + \sum_i \epsilon$, increases in N will shift upward the distribution of income, thus (weakly) increasing the probability that aggregate income is above the threshold.[39]

[39] This occurs only 'weakly' because if initially all the density for Y is far below the threshold, then shifting up the threshold will bring $Y^{\max}$ closer to $\widehat{Y}_1$, but if it does not cross the threshold then the probability of investment remains zero. Similarly, if N is large enough, all the density may be above the threshold (depending on the distribution of income shocks), in which case again there is no change in the probability.

3.8 Additional Tables and Figures

Table 3.8: *Share of eligible's links to other categories of household*

Sample: Census sample (35 villages), baseline

	Actual (1)	Random linking (2)
Share in of links in:		
Whole village	.94	.91
Low wealth	.70	.55
Other eligibles	.12	.06
Total links	578	590
Total households	197	197
Links per household	2.98	3.04

Notes: *These statistics are constructed using baseline (pre-program) data for all households in a 35 village subsample of the data. In these villages, the sample includes a census of all households, allowing the characteristics of a household's 'neighbours' to be observed. A pair of households are linked ('neighbours') if either reports (a) going to the other household for assistance in a crisis; (b) going to the other household to borrow food; or (c) receiving transfers from the other. Wealth classes are determined using a Participatory Rural Assessment, which aggregates classifies households into four or five wealth classes. Additional criteria are used to determine which households in the poorest wealth class are eligible for treatment. 'Village' includes all households within the village; 'low wealth' includes all households in the bottom two wealth classes (bottom three when five wealth ranks were used); 'other eligibles' includes only households who are also eligible for the program. Column 1 shows the share of all links from eligible households to households in these other 'wealth class' categories. Column 2 shows the share of all links from eligible households that would go to households in these other 'wealth class' categories if links were formed randomly. 'Total links' shows the total number of links observed (Col 1), or the number that would be observed under random linking (Col 2).*

Table 3.9: *Townsend Test: How does expenditure change with income changes*

Dependent Variable: Change in log expenditure
Sample: Control households (main sample), 0-2-4 year
Village level clustered standard errors

Group definition:	Eligibles (1)	All Poor (2)	Whole Village (3)
Change in log income	.048***	.041***	.043***
	(.015)	(.013)	(.012)
P-values:			
Same as Column (1)	—	.026	.128
Same as Column (2)	—	—	.403
Demographic controls	9,587	9,587	9,587
Households	9,587	9,587	9,587
Observations (household $\times$ wave)	19,174	19,174	19,174
Clusters	1,409	1,409	1,409

Notes: **** denotes significance at 1%, ** at 5%, and * at 10% level, when treated as a standalone regression. The sample comprises all households in control villages across the full sample. Observations are at the household level. Wealth classes are determined using a Participatory Rural Assessment, which classifies households into four or five wealth classes. Eligible households are those in the lowest wealth class who also meet additional financial and demographic criteria. All poor households are those belonging to the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Whole village includes all households. Data were collected using a stratified random sampling scheme, so weights are used to make the sample representative of the population. All financial variables are in 2007 USD terms. They are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation. They are then converted to dollars using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. The outcome measures the increase in log expenditure. Expenditure includes food (both purchased an produced), fuel, cosmetics, entertainment, transportation, utilities, clothing, footwear, utensils, textiles, dowries, eduation, charity, and legal expenses. Income includes total household income from all sources, but does not include transfers. Demographic controls include changes in household size by age category, and changes in education enrollment. All columns show regressions of the outcome on a constant, the change in log income, the change in demographic controls, interactions of all of these variables with an indicator for not being eligible, and a group dummy. The interaction with the indicator means the coefficients shown are those for the eligible households. Across the columns what changes is the definition of a group. In Column (1) the groups are eligibles or ineligibles within a village. In Column (2) the groups are all poor or not within a village. In Column (3) the groups are the villages.*

Table 3.10: *Methods used by households to cope with crises*

Methods of smoothing if household experienced crisis
Sample: Low wealth control households, 0-2-4 year

	(1) Hh member ill	(2) Crop Loss
Reduce consumption	.37	.33
Use savings	.47	.41
Borrowing/Transfers	.39	.51
Borrowing	*.25*	*.30*
Transfers	*.16*	*.23*
Observations	4,767	4,594

Notes: *Constructed using data from all three waves (2007, 2009, 2011) for low wealth households in control villages who report experiencing a crisis. Observations are at the household level. Data were collected using a stratified random sampling scheme, so weights are used throughout to make the sample representative of the population. All poor households are those in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Households were asked whether or not they suffered a crisis. Having a household member ill and suffering crop loss are the two biggest sources of crisis. Households who report having such crises are asked how they coped with the crisis. Multiple coping stratgies are permitted.*

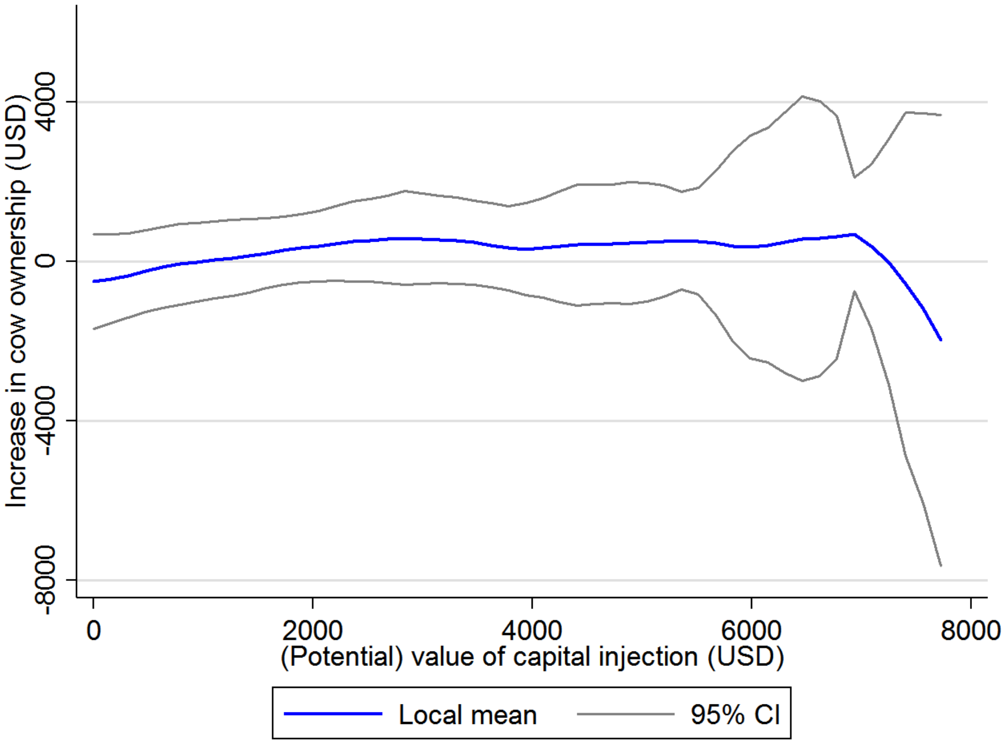
Table 3.11: *Other variables are smooth at the kink in investment*

Standard errors (in parentheses) clustered at district level

	Total Income (1)	Total Savings (2)	Business Assets (3)	Total Assets (4)	Village Size (5)	Group Size (6)	Total Migrants (7)	Migrant Households (8)
Slope below threshold for Treated	2.85	-.021	.307	.661	-.002**	-.002**	-.035	-.035
	(3.71)	(.308)	(.486)	(.787)	(.001)	(.010)	(.010)	(.010)
Change in slope at threshold for Treated	-1.11	.137	-.149	-.150	.001	-.024	-.024	-.024
	(2.12)	(.195)	(.247)	(.442)	(.001)	(.015)	(.015)	(.015)
Treated	-7900	-594	-2110**	-4150*	4.30*	6.35**	.200**	.200**
	(7950)	(679)	(910)	(1560)	(2.05)	(.029)	(.029)	(.029)
Mean Baseline Level among Controls	60,000	3,300	4,890	11,200	.766	.766	.766	.766
Observations (Clusters)	1360 (13)	1360 (13)	1360 (13)	1360 (13)	1360 (13)	1360 (13)	1360 (13)	1360 (13)

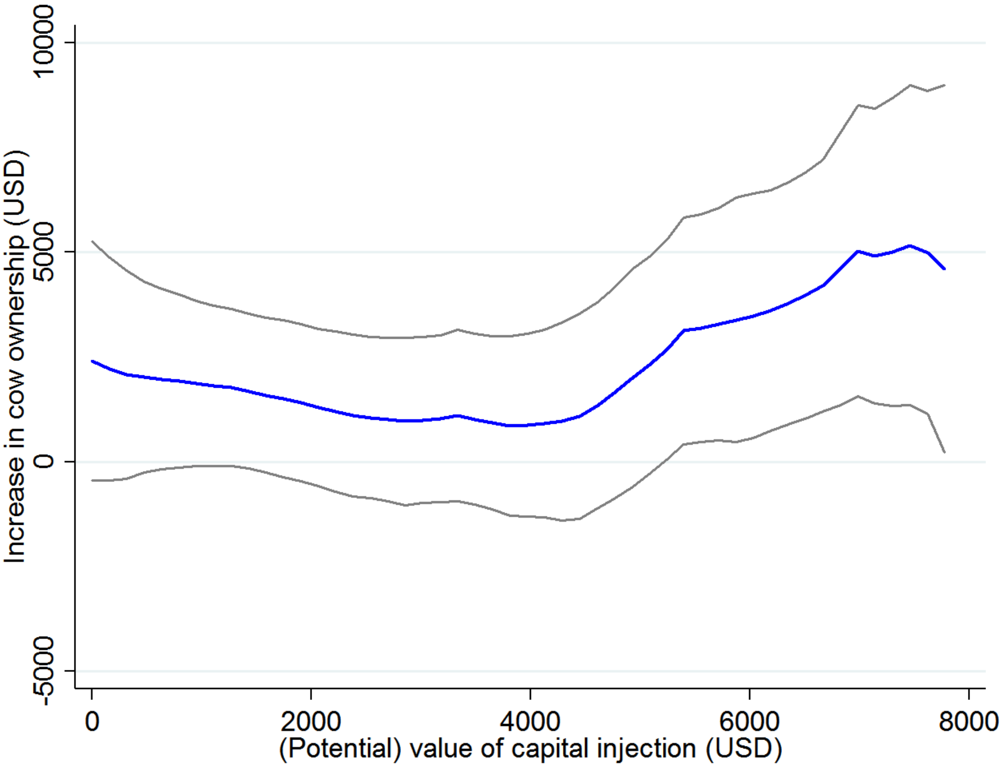
Notes: **** denotes significance at 1%, ** at 5%, and * at 10% level. The sample comprises all households in all villages across the full sample in 2009. Data were collected using a stratified random sampling scheme, so weights are used to aggregate household data to the risk-sharing network level. All financial variables are in 2007 USD terms. Where necessary, they are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation. They are then converted to dollars using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. In Column (1) the outcome measures the change in the price of milk. The price of milk is calculated by dividing for each household their total expenditure on milk by the total quantity they report purchasing, and then averaging this across households in the village. In Column (2) the outcome measures the change in the price of cows. The price of cows is calculated by dividing for each village the total reported value of cows by the total reported number of cows. In Column (3) the outcome measures the change in the wage. The wage is calculated by dividing for each household the total reported income by the total reported number of wage hours worked. In Column (4) outcome measures the change in the share of all poor households who are without cows and report aspiring to own cows. The table shows regression of the outcome in 2009 on the aggregate value of capital provided to the risk sharing network in 2007. Financial controls are the total income, total savings, and total value of cows in 2007.*

Figure 3.7: *Placebo test – impact of future capital injection on investment*



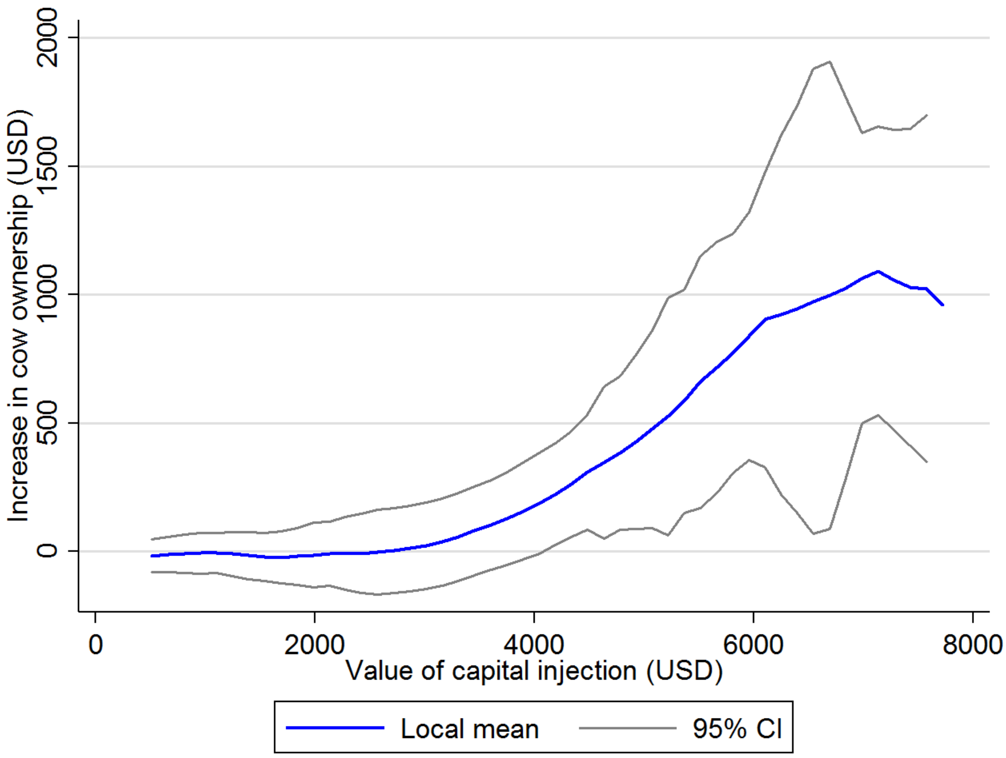
Notes: *Constructed using data on all poor households in control villages across the full sample. Observations at the risk-sharing network level. Within a village, a risk-sharing network is the set of households in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Data were collected using a stratified random sampling scheme, so weights are used throughout to aggregate household data to the risk-sharing network level. The potential value of the capital injection is the value of the assets that would have been transfered to an eligible household (515 USD PPP) if they had been in a treated village, multiplied by the number of eligible households in the risk-sharing network. The potential value of transfer to an eligible household is calculated by converting the value of the assets in Bangladeshi Taka (9,500TK) to 2007 USD terms, using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. Increase in cow ownership is measured as the increase in the value of cows between 2009 and 2011 owned by all households in the risk-sharing network. The values in Bangladeshi Taka are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation, and then converted to USD PPP. A thin tail of networks receive more than $8,000 worth of assets. Since the density on this part of the support is low (fewer than five networks for any value of asset transfers), I trim these networks (2% of the sample). The figure shows non-parametrically the relationship between the increase in the aggregate value of cows in a risk-sharing network between two years and four years after transfers would have been made, and the value of capital that would have been provided to the network by the program had the network been in a treatment village. Investment is consistently flat at zero. The kernel-weighted local mean, estimated using an Epanechnikov kernel with bandwidth $800. The outer region provides the 95% confidence interval.*

Figure 3.8: *Impact of capital injection on investment – whole village*



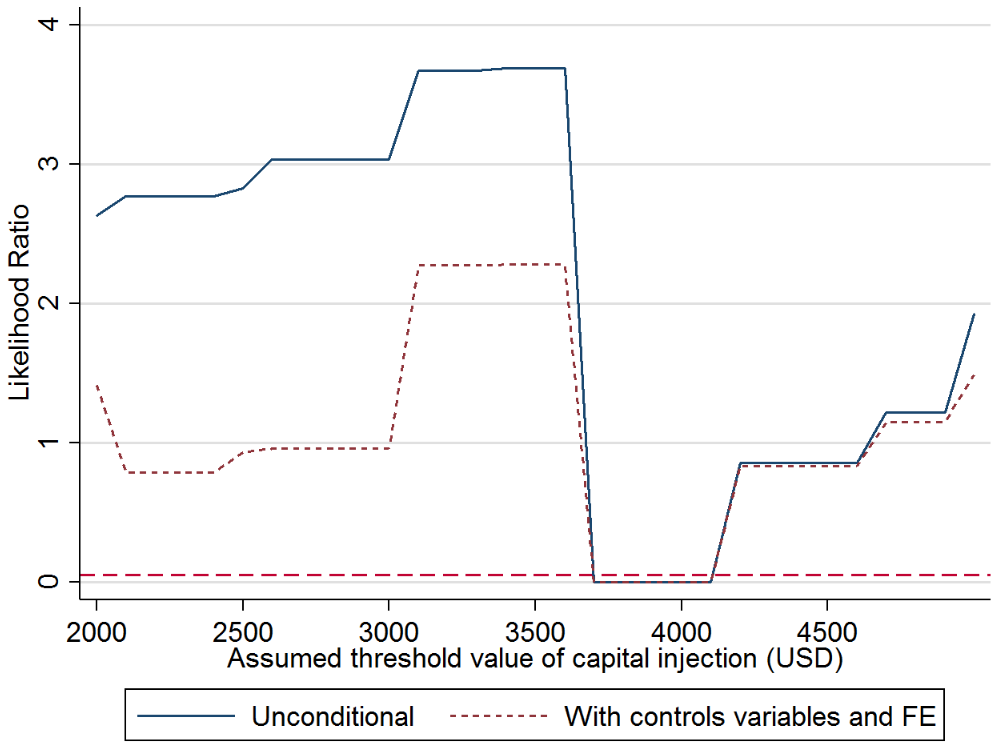
Notes: *Constructed using data on all households in treated villages across the full sample. Observations at the village level. Data were collected using a stratified random sampling scheme, so weights are used throughout to aggregate household data to the village level. The value of the capital injection is the value of the assets transfered to an eligible household (*515 *USD PPP) multiplied by the number of eligible households in the village. The value of transfer to an eligible household is calculated by converting the value of the assets in Bangladeshi Taka (*$9,500$*TK) to 2007 USD terms, using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. Increase in cow ownership is measured as the increase in the value of cows between 2009 and 2011 owned by all households in the village. The values in Bangladeshi Taka are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation, and then converted to USD PPP. A thin tail of villages receive more than* $8,000 *worth of assets. Since the density on this part of the support is low (fewer than five networks for any value of asset transfers), I trim these villages (2% of the sample). The graph shows the kernel-weighted local mean, estimated using an Epanechnikov kernel with bandwidth $800. The outer region provides the 95% confidence interval.*

Figure 3.9: *Impact of capital injection on investment – only eligible households*



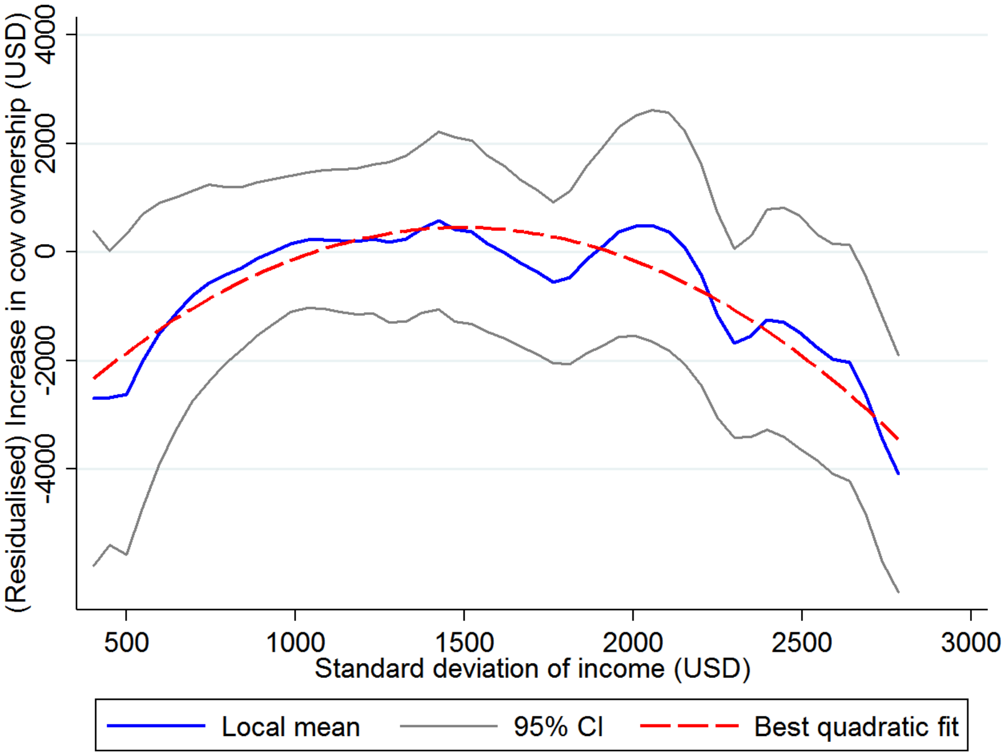
Notes: *Constructed using data on only eligible households in treated villages across the full sample. Observations at aggregated across eligible households to the village level. The value of the capital injection is the value of the assets transfered to an eligible household (515 USD PPP) multiplied by the number of eligible households in the village. The value of transfer to an eligible household is calculated by converting the value of the assets in Bangladeshi Taka ($9,500TK$) to 2007 USD terms, using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. Increase in cow ownership is measured as the increase in the value of cows between 2009 and 2011 owned by all eligible households in the village. The values in Bangladeshi Taka are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation, and then converted to USD PPP. A thin tail of villages receive more than $8,000 worth of assets. Since the density on this part of the support is low (fewer than five networks for any value of asset transfers), I trim these villages (2% of the sample). The graph shows the kernel-weighted local mean, estimated using an Epanechnikov kernel with bandwidth $800. The outer region provides the 95% confidence interval.*

Figure 3.10: *Hansen Test for threshold location*



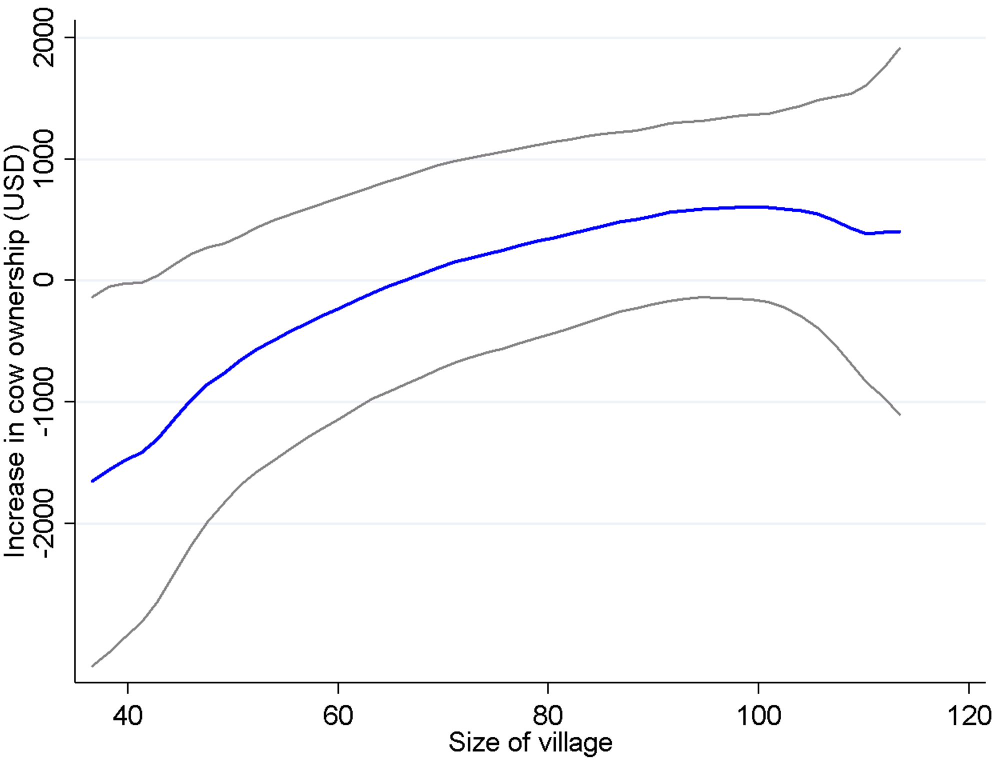
Notes: *Constructed using data on all poor households in treated villages across the full sample. Observations at the risk-sharing network level. Within a village, a risk-sharing network is the set of households in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Data were collected using a stratified random sampling scheme, so weights are used throughout to aggregate household data to the risk-sharing network level. The value of the capital injection is the value of the assets transfered to an eligible household (515 USD PPP) multiplied by the number of eligible households in the risk-sharing network. The value of transfer to an eligible household is calculated by converting the value of the assets in Bangladeshi Taka ($9,500$TK) to 2007 USD terms, using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. I sequentially run the specification in Equation 3.34 at different values of the threshold, varying the threshold between \$2000 and \$5000, at intervals of \$100. The figure shows, for each assumed threshold value of capital injection, the likelihood ratio (LR) statistic. This statistic is the difference in residual sum of squares (RSS) from the assumed threshold regression, relative to the RSS from the regression for which the lowest RSS was achieved, divided by that minimum RSS, and multiplied by the sample size. Any possible thresholds for which the LR is below .05 cannot be rejected as possible values for the threshold. The graph show the range of LR statistics both for the unconditional case and when additional controls (lagged income and asset variables, and network size), and district level fixed effects are included. In both cases it is clear that a threshold value of $\$3,700 - \$4,100$ is by far the most likely, and all other thresholds can be rejected.*

Figure 3.11: *Investment is an inverted-U in income inequality, as measured by standard deviation*



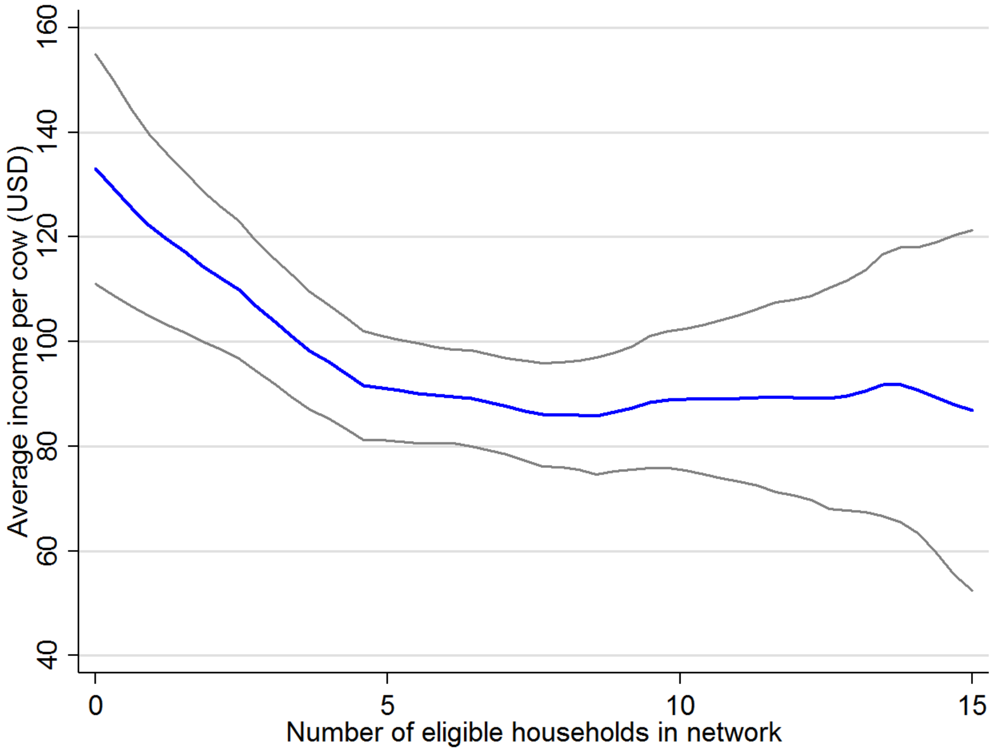
Notes: *Constructed using data on all poor households in control villages across the full sample. Observations at the risk-sharing network level. Within a village, a risk-sharing network is the set of households in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Data were collected using a stratified random sampling scheme, so weights are used throughout to aggregate household data to the risk-sharing network level. The standard deviation of income for a network is the standard deviation of the cross-sectional income distribution in 2009. It is converted to 2007 USD terms, using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. Increase in cow ownership is measured as the increase in the value of cows between 2009 and 2011 owned by all households in the risk-sharing network. The values in Bangladeshi Taka are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation, and then converted to USD PPP. Residualised increase in cow ownership is the residuals from first regressing increase in cow ownership on total income, total saving, and the value of cows, all in 2009 and 2007, and also network size. The kernel-weighted local mean, estimated using an Epanechnikov kernel with bandwidth $100. The outer region provides the 95% confidence interval.*

Figure 3.12: *Investment is increasing in network size*



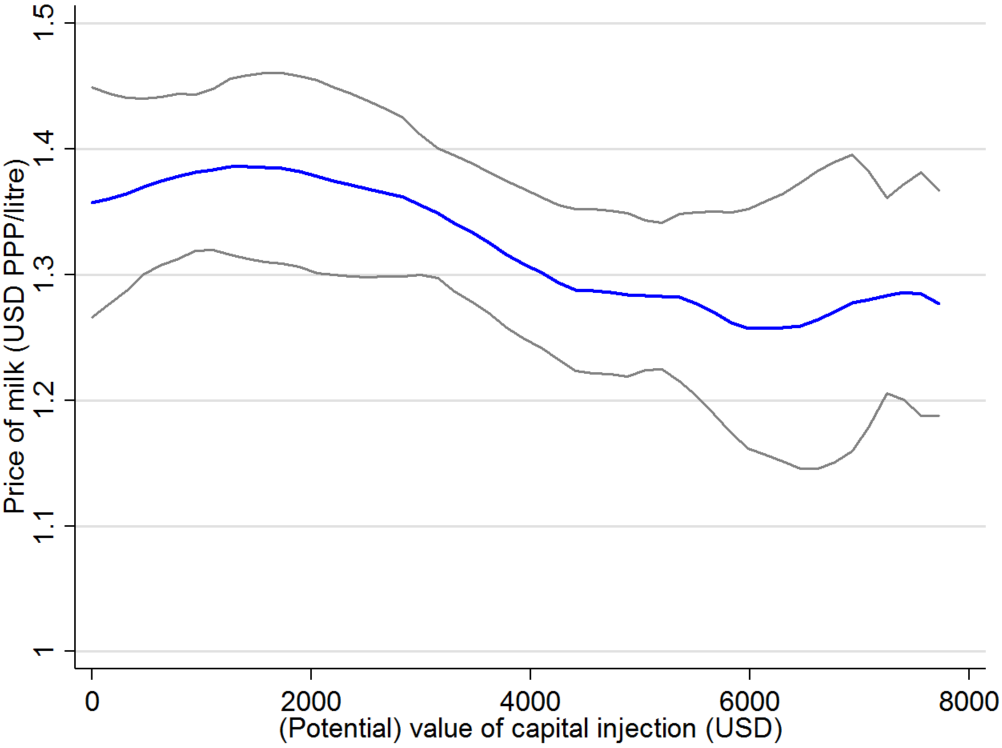
Notes: *Constructed using data on all poor households in control villages across the full sample. Observations at the risk-sharing network level. Within a village, a risk-sharing network is the set of households in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Data were collected using a stratified random sampling scheme, so weights are used throughout to aggregate household data to the risk-sharing network level. The network size is measured by the number of households in the network. Increase in cow ownership is measured as the increase in the value of cows between 2009 and 2011 owned by all households in the risk-sharing network. The values in Bangladeshi Taka are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation, and then converted to USD PPP. Residualised increase in cow ownership is the residuals from first regressing increase in cow ownership on total income, total saving, and the value of cows, all in 2009 and 2007, and also network size. The kernel-weighted local mean, estimated using an Epanechnikov kernel with bandwidth 10. The outer region provides the 95% confidence interval.*

Figure 3.13: *Average return on cows is declining in number of transfer recipients*



Notes: *Constructed using data on all poor households in treated villages across the full sample in 2009. Observations at the risk-sharing network level. Within a village, a risk-sharing network is the set of households in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Data were collected using a stratified random sampling scheme, so weights are used throughout to aggregate household data to the risk-sharing network level. Average income per cow is the mean income per cow across cow-owning households in the network in 2009. The values in Bangladeshi Taka are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation, and then converted to USD PPP, using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. A thin tail of networks have more than* 15 *eligible households. Since the density on this part of the support is low (fewer than five networks for any number of eligible households), I trim these networks (2% of the sample). The graph shows the kernel-weighted local mean, estimated using an Epanechnikov kernel with bandwidth 1.8. The outer region provides the 95% confidence interval.*

Figure 3.14: *Program doesn't affect the price of milk*



Notes: *Constructed using data on all poor households in treated villages across the full sample. Observations at the risk-sharing network level. Within a village, a risk-sharing network is the set of households in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Data were collected using a stratified random sampling scheme, so weights are used throughout to aggregate household data to the risk-sharing network level. The value of the capital injection is the value of the assets transfered to an eligible household (515 USD PPP) multiplied by the number of eligible households in the risk-sharing network. The value of transfer to an eligible household is calculated by converting the value of the assets in Bangladeshi Taka (9,500 TK) to 2007 USD terms, using purchasing power parity (PPP) exchange rates, where 1 USD = 18.46TK in 2007. Price of milk is constructed by taking the ratio of household expenditure on milk with household consumption of milk in 2009. Household level prices are winsorised, replacing prices below the 1st (above the 99th) percentile with the price at the 1st (99th) percentile. These are then averaged over households in the entire village, to give an estimated price of milk in each village. These prices in Bangladeshi Taka are next deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation, and then converted to USD PPP. A thin tail of networks receive more than $8,000 worth of assets. Since the density on this part of the support is low (fewer than five networks for any value of asset transfers), I trim these networks (2% of the sample). The graph shows the kernel-weighted local mean, estimated using an Epanechnikov kernel with bandwidth $800. The outer region provides the 95% confidence interval.*

Figure 3.15: *No impact on further investment from share of households treated*

Notes: *Constructed using data on all poor households in treated villages across the full sample. Observations at the risk-sharing network level. Within a village, a risk-sharing network is the set of households in the lowest two (three) wealth classes, when there are four (five) possible wealth classes. Wealth classes are determined using a Participatory Rural Assessment. Data were collected using a stratified random sampling scheme, so weights are used throughout to aggregate household data to the risk-sharing network level. The share of poor households receiving transfers is the proportion of households in the poorest wealth class who receive transfers. Increase in cow ownership is measured as the increase in the value of cows between 2009 and 2011 owned by all households in the risk-sharing network. The values in Bangladeshi Taka are first deflated to 2007 terms using the Bangladesh central bank CPI measure of inflation, and then converted to USD PPP, where 1 USD = 18.46TK in 2007. A thin tail of networks receive more than* $8,000 *worth of assets. Since the density on this part of the support is low (fewer than five networks for any value of asset transfers), I trim these networks (2% of the sample). The graph shows the kernel-weighted local mean, estimated using an Epanechnikov kernel with bandwidth $800. The outer region provides the 95% confidence interval.*

Chapter 4

Melting Pot or Salad Bowl: The Formation of Heterogeneous Communities[1]

4.1 Introduction

Ethnic, linguistic, and religious heterogeneity are associated with a variety of politico-economic problems, including low growth, low provision of public goods, and conflict. Yet the effects of diversity are not uniform: some heterogenous populations manage to avoid societal conflict and are economically successful. In some cases, diversity can even have positive effects on growth, productivity, and innovation.[2] Seeking to explain this puzzle, recent work finds that social cohesion plays an important role in how well a population 'deals' with its heterogeneity. For example, more segregated populations are found to have lower quality of governance, lower trust between citizens, and worse education and employment outcomes, compared to similarly heterogenous populations that are better integrated.[3]

[1]This chapter is cco-authored with Bryony Reich. We are extremely grateful to Alberto Alesina, Sanjeev Goyal, and Imran Rasul. For very helpful comments we would like to thank Ana Babus, Tim Besley, Antonio Cabrales, Partha Dasgupta, Christian Dustmann, Andrea Galeotti, Raffaella Giacomini, Sriya Iyer, Matthew Jackson, Terri Kneeland, Edward Lazear, Friederike Mengel, Paul Milgrom, Jonathan Newton, Marco van der Leij, Chris Parmeter, Ilya Segal, Jorgen Weibull, and Peyton Young, as well as audiences at ASREC Washington DC, Barcelona Summer Forum, Cambridge, CTN Warwick, Oxford, PET Lisbon, SAET Paris, and Stanford. Advani gratefully acknowledges financial support from the ERC (GA313234). Reich gratefully acknowledges financial support from the NET Institute (www.NETinst.org) and the UK Economic and Social Research Council (grant number ES/K001396/1).

[2]Easterly and Levine (1997), Alesina et al. (1999), Montalvo and Reynal-Querol (2005), Ashraf and Galor (2013), and Rasul and Rogger (2015), amongst others.

[3]See for example Alesina and Zhuravskaya (2011) and Cutler and Glaeser (1997). Firm and team studies suggest that environments that bring together heterogeneous skills and ideas, at the same time as fostering cooperation, are associated with positive effects of diversity on innovation and productivity. At the national level Ashraf and Galor (2013) show a hump-shaped relationship between genetic diversity and growth, consistent with their hypothesis that diversity is beneficial for innovation and production but can also be costly when it reduces

Despite its clear importance, relatively little is known about what determines whether a heterogenous population ends up in a cooperative or divisive situation. In this chapter we propose a model that allows us to study this question. We ask, first, what social structures can arise when society is composed of heterogeneous groups, and second, what features influence which social structure arises in a particular environment. Understanding this is crucial in an increasingly mobile world in which immigration and social cohesion are frequently at the forefront of political agendas.[4]

In our model, individuals from two distinct groups living side-by-side, make choices over activities to engage in, and who in the community to interact with. We will use the example of natives and immigrants, although other interpretations are possible. Some activities, such as language choice, are 'cultural actions', in the sense that each group will start with an *ex ante* preferred language, and there is a cost to switching to any other choice. Other activities are 'non-cultural': there is no group-specific reason why one activity is more costly to engage in than another.[5] For example, there may be no group-specific reason why one sport should be more costly to engage in than another.[6] Interaction provides opportunities for economic exchange, and so individuals can benefit from forming social ties with others in the population. Since interaction requires some degree of commonality of actions, we assume that the benefit of a link is increasing in the number of shared activities, with a fixed (per-link) cost of formation. The fundamental trade-off is that an individual prefers to maintain his cultural practices but doing so can hinder opportunities for interaction and exchange with those who adopt different practices.[7]

We find that only three classes of social structure are possible in (Nash) equilibrium: assimilation, segregation, and multiculturalism. In assimilation equilibria, all individuals engage in the same activities (cultural and non-cultural), and interact with everyone. In segregation equilibria, all individuals play their type-specific cultural activities, and form social ties only with those of the same type. In multicultural equilibria, individuals play their type-specific cultural activities, but all coordinate on common non-cultural activities, and this allows individuals to

cooperation and increases disarray within a society. See Alesina and La Ferrara (2005) and Alesina et al. (2013) for further discussion and references.

[4]Europe provides an example of a variety of policies designed to integrate immigrants and perhaps also influence voters. For example, in France, in 2010, the senate voted $246-1$ to ban the full Islamic veil, with the French President arguing it was not consistent with French identity. In the UK, in 2013, the government announced plans to increase English language requirements to ensure migrants were 'able to integrate into British society'.

[5]In the model presented in the main body of the chapter an individual chooses only two actions, a cultural action and a non-cultural action. In Subsection 4.6.2 we allow for multiple cultural and non-cultural actions.

[6]Of course some groups have ties to particular sports and so which activities are non-cultural and cultural will depend on the groups. A word of caution: a result of this chapter is that non-cultural activities can in equilibrium become associated with a particular group. Thus many of the examples that spring to mind that associate sport or other activities with a particular group may in fact be an equilibrium outcome and not a result of some *ex ante* preferences.

[7]In an example of this, Algan et al. (2013) find an economically significant trade-off faced by Arabic parents in France between attachment to their own culture (in their study, the desire to pass on an Arabic name) with the future economic performance of their children in the form of work-related penalties to having an Arabic name. This is an important trade-off and is present in varying forms in Kuran and Sandholm (2008), Lazear (1999), Bisin et al. (2011b), Carvalho (2013b), and Carvalho (2013a), amongst others.

form ties with all others, regardless of type. The existence of shared non-cultural practices allows interaction between disparate groups who maintain distinct cultures.

What kind of environments sustain these three outcomes? The key pattern we identify is that social structure in heterogeneous populations changes discretely in the share of the minority. When the minority (immigrant) group is small relative to the majority, assimilation occurs. Intuitively, restricting social interaction within such a small group is not desirable and an immigrant does better by paying the cost of switching culture and being absorbed into a much larger group. Multiculturalism and segregation are equilibria only once the share of the minority group in the population exceeds particular (different) thresholds. Above a certain threshold, there is a large enough 'critical mass' of immigrants that if the group maintains its distinct culture then, for any immigrant, the cost of switching culture outweighs the benefits of increased interaction. This threshold result is due to complementarities: maintaining a distinct minority culture is only 'worth it' if others do so, therefore either no one maintains minority culture or a large group of immigrants does.[8]

The location of these thresholds, above which segregation and multiculturalism are equilibria, depends on three parameters: (*i*) cultural distance; (*ii*) the importance of culture in everyday life; and (*iii*) the cost of forming a social tie. When the cultural distance between immigrants and natives is larger, the cost of switching cultural actions is relatively higher. This reduces the threshold immigrant share at which multiculturalism and segregation – equilibria where immigrants retain their own culture – become possible. In contrast, a higher importance of culture in daily life makes having a common culture more important for interaction, and so it is harder to sustain multiculturalism (the threshold on multiculturalism rises). To illustrate, if social interaction involving alcohol is ubiquitous, and an immigrant's culture prohibits alcohol, then this makes it difficult for the immigrant to both maintain his cultural practices and integrate with natives, and thus he is forced to choose. Finally, higher costs of forming a link lower the threshold on segregation. Intuitively, since the costs of assimilation come from switching culture, while the benefits come from improved interaction with natives, a higher cost of link formation lowers the relative benefits of assimilation and makes segregation easier to sustain.

We test two key predictions of our model – that a threshold should exist in community behaviour, and that the location of this threshold should depend on cultural distance – using census data on immigrant populations in the United States at the beginning of the 20th Century. During this period large numbers of immigrants arrived in America, radically altering the national make-up of the country. This is exactly the kind of situation our framework seeks to understand. Indeed, the three distinct forms of social structure we find are evident in scholarly discussion

[8]As we discuss below, the importance of thresholds as a natural feature of community behaviour is increasingly being recognised, for example in Card et al. (2008) and Chay and Munshi (2013).

of communities in the United States following mass migration (Gordon, 1964). Early scholars argued that a single culture would prevail: the so-called 'melting pot'. As it became clear that not all communities assimilated, but some instead 'retained distinctive economic, polical and cultural patterns long after arriving in the United States', segregation became a major concern (Bisin and Verdier, 2000). However, there was also discussion of a 'third way' – 'the salad bowl' – where immigrants could become 'American' and integrate whilst maintaining some cultural distinction. This period in history has the added benefit of furnishing us with data on a large number of heterogeneous communities with varying immigrant group sizes, vital to any analysis.

If our framework captures the key forces driving community formation, then we should expect behaviour in heterogeneous populations to exhibit the threshold patterns predicted by (and central to) the model. These predictions also allow us to separate our model from explanations based on selection, which do not suggest the discontinuities between group size and community outcomes that are critical to our model. We test our predictions on the decision of immigrants to acquire the English language or not (a cultural action), and on the decision to in-marry (a partial measure of interaction). We find sharp and significant thresholds in behaviour when an immigrant group forms around one-third of the community: above this threshold English acquisition in the immigrant group falls by almost a half, from 90% to 50%, and in-marriage rises by a third from 55% to 75%. Using data on linguistic distance between the language of the immigrant group and English, we additionally show that, as predicted, when cultural (linguistic) distance is higher, the estimated threshold for segregation and multicultural outcomes is lower.

The literature examining choice of culture highlights the variety of outcomes that can arise in heterogenous populations. In seminal work in economics on cultural transmission, Bisin and Verdier (2000) examine the persistence of different traits in a mixed community even in the long run.[9] Iannaccone (1992), Berman (2000), and Carvalho (2013b) study the stability of costly and restrictive cultural practices within religious groups in heterogenous societies.[10] Kuran and Sandholm (2008) examine the convergence of cultural practices when diverse groups interact. The question we ask in this chapter is different. We want to understand *selection* between various outcomes.[11] That is, when does one of these social structures emerge rather than another?

[9]See also Bisin et al. (2004) for empirical work on this topic. A key finding in Bisin and Verdier (2000), that smaller groups may exert more effort in passing on their culture to their children, sounds at odds with our finding that smaller groups are more likely to assimilate. This is not the case. When groups continue to maintain different practices in equilibrium we find that smaller groups must put in more 'effort' in the form of maintaining higher diversity of practices in order to sustain segregation. See the model of action choice along more than two dimensions in Subsection 4.6.2. Bisin and Verdier (2000) also point out that effort does not necessarily relate monotonically to outcomes: a small group could put in lots of effort to pass on their culture but it may still die out, depending on the cost function.

[10]Berman (2000) and Iannaccone (1992) model religion as a club good with 'extreme' cultural practices as a means of taxing other goods to increase contribution to the religious good. This is slightly different to the framework presented in this chapter and the other papers cited, which draw on the Akerlof and Kranton (2000) model of identity.

[11]We are reassured that our framework permits, as equilibria, outcomes consistent with those that are studied in detail in these papers.

It is important to highlight the two features of the theoretical framework that make it rich enough to produce the different social structures. The first is the novel introduction of non-cultural actions, which play an important role in uniting or dividing communities. Adoption of a common non-cultural action is necessary for the existence of multicultural equilibria. At the other end of the spectrum, rather than bridging the gap between groups, non-cultural actions can also be used to divide them. We find a subclass of segregation equilibria in which immigrants not only retain distinct cultural practices but also create 'new diversity' by adopting deliberately different non-cultural activities from natives. Such polarisation of non-cultural practices occurs in order to maintain segregation by raising the cost of interacting with the other group. This extreme form of segregation occurs when culture is relatively unimportant in everyday life.

The second feature of the framework to highlight, is the interaction between choice of behaviour and choice of interaction. In pioneering work, Lazear (1999) studies the choice of whether or not to adopt the same language or culture.[12] In Lazear (1999) and Carvalho (2013a), agents choose a cultural practice, but take interaction as given. A second literature models link formation in order to study the important concept of homophily - the tendency for similar individuals to be linked. In this case individuals choose interaction but take behaviour as given (Currarini et al., 2009; Bramoullé et al., 2012; Currarini and Vega-Redondo, 2011). We take a novel approach which, importantly, addresses both these choices (choice of practices and choice of interaction) within a single tractable framework.[13] Our alternative approach allows payoffs from links formed to depend not only on *ex ante* heterogeneity but also on the action choices of both partners.[14] This results in a broader range of social structures and comparative statics.

Our theoretical approach contributes to a literature on network formation where individuals play coordination games across links, in particular Jackson and Watts (2002) and Goyal and Vega-Redondo (2005). They show that in homogenous populations, under equilibrium refinement, total integration and social conformism always prevail. We show that this is not true for heterogenous populations in which different individuals have preferences for coordinating on different activities. Our analysis of network formation in heterogeneous populations enables us to study the important question of why different social structures prevail and when.

Card et al. (2008) and Chay and Munshi (2013), like us, test empirically for the presence of a threshold in community behaviour that depends on the size of the minority population, where the location of this threshold is *a priori* unknown. Card et al. (2008) look at local migration

[12]See also Eguia (2013) who examines this from the perspective of discrimination.

[13]In a different framework, Bisin and Verdier (2000) highlight the importance of choice of social interaction and make segregation effort a choice. However, they do not analyse resulting levels of segregation since their focus is whether diverse cultural traits persist. Also Carvalho (2013b) examines a choice of segregation in the decision to veil or not.

[14]Or, to put it the other way around, payoffs from actions taken to depend not only on *ex ante* heterogeneity but also on who one linked with and their choice of action.

and Chay and Munshi (2013) at local migration and voting behaviour. In contrast, our interest is in social interaction and convergence (or not) of behaviours within the community. Together with Card et al. (2008) and Chay and Munshi (2013), our findings highlight the importance of looking for these kind of threshold patterns when considering community-related behaviour. Our findings are also in line with recent work on the US by Abramitzky et al. (2014), Fouka (2014), and Fulford et al. (2015), which suggests that environmental features were significant in determining how culture evolved in 19th and early 20th century United States.

In Section 4.2 we present the basic model. Section 4.3 characterises the Nash equilibria and provides comparative statics. Section 4.4 provides empirical evidence from communities in the United States in the age of mass migration supporting the predictions of the model. The final section discusses implications for government policy and welfare and concludes.

4.2 Theoretical Framework

First we present the framework for choice of action, then introduce social interaction, and finally we summarise the payoffs.

4.2.1 Culture

A population (or community) consists of a set of individuals $i \in \{1, 2, \ldots, n\}$. Each individual i is endowed with a type, $k \in \{M, m\}$, which is common knowledge. We refer to the majority M types as 'natives', and the minority m as 'immigrants' (although many other interpretations are clearly possible). There are $n_M \in \mathbb{N}$ individuals of type M and $n_m \in \mathbb{N}$ individuals of type m, where $n_M, n_m \geq 2$, $n = n_M + n_m$, and we assume $n_M \geq n_m$.[15]

To illustrate the model with an example, consider the population to be a neighbourhood. An immigrant group has moved into the neighbourhood from a different country and has come with different cultural practices to the native group. We denote the cultural practices associated with the native group by the action x^M and the cultural practices associated with immigrant group by the action x^m.[16] Cultural practices could be activities, such as type of food eaten or language spoken, or behaviours, for example concerning education, gender, or marriage. Since cultural practices are, in themselves, simply activities and behaviours, this leads to an observation: when immigrants move to a new country they can, if they wish, adopt the cultural practices of the native group (or vice versa). We refer to this as switching culture. While it is possible to switch

[15] We use the terms 'native group' and 'immigrant group' as an illustration. Of course we need not always consider the native group as the majority group, for example Aboriginal populations of Australia and Native American populations of the United States.

[16] We model the practices that define a group's culture by a single action. We enrich this model of culture by allowing for multiple dimensions of culture in Subsection 4.6.2. We do the same for non-cultural actions (introduced below). The main results remain.

culture, there is a cost c from doing so. This is how we define culture: an individual chooses an action x_i from the set $\{x^M, x^m\}$, where it is possible to adopt the action associated with the other group, but at a cost.[17]

Berry (1997) describes cultural changes that occur when groups with different practices share the same environment as ranging 'from relatively superficial changes in what is eaten and worn, to deeper ones involving language shifts, religious conversions, and fundamental alterations to value systems'. Regarding religion, Iannaccone (1992) highlights that 'people can and often do change religions or levels of participation over time'. The cost to switching culture can arise for a variety of reasons. There may be fixed costs involved, such as learning a new language, or participating in unfamiliar activities. Alternatively cultural practices may be considered valuable in their own right.[18] Culture can be so deeply entrenched that individuals find it psychologically costly to adhere to behaviours or attitudes that differ from the culture one has grown up with.[19] We reduce these different possibilities to a single cost, c, in line with previous work on culture choice (Akerlof and Kranton, 2000; Bisin and Verdier, 2000). The magnitude of c is interpreted as a measure of cultural distance.

It is rather extreme to assume that all activities going on in this neighbourhood are necessarily associated with type or culture. We therefore also consider non-cultural practices: activities which can facilitate social interaction, but where there is no differential cost by group membership. Non-cultural actions are modeled by an individual i's choice of action $y_i \in \{y^A, y^B\}$ for which there is no associated type-specific cost.[20] For example, while religion often prescribes or prohibits certain activities, religions rarely proscribe what sports should be played or what music can be listened to. In some cases sports provide a common ground that all groups can share. However, we may also see communities divided by such activities, such as in Northern Ireland, where Catholic and Protestant communities typically play different sports. A result that we will highlight later is that activities which *ex ante* have nothing to do with type and culture, such as sports, can in equilibrium become associated with a particular group. Thus many of the examples that spring to mind that associate sport or other activities with a particular group may in fact be an equilibrium outcome and not a result of some *ex ante* preferences.

[17] The modeling of culture and a cultural group presented here is consistent with the introduction of identity into economic modeling by Akerlof and Kranton (2000). Note that in our framework an individual chooses one action or the other. For some practices such as language, however, an individual might continue to speak his group's language, but also learn the language of the other group. The results will hold provided that the relative benefit of learning the native language increases the more members of the immigrant group there are that learn it. This would be the case, for example, if interaction and exchange among individuals and groups occurs more and more often in the native language as more immigrant group members learn it.

[18] Algan et al. (2013) estimate that the utility an Arabic parent in France gets from passing an Arabic name to their child is equivalent to a 3% rise in lifetime income of the child. See also Bisin and Verdier (2000) and Kuran and Sandholm (2008) for examples.

[19] See Berry (1997) for a review and further references on psychological and sociocultural costs. This framework does not incorporate group penalties, but such an assumption could be incorporated.

[20] Of course there may be multiple activities and behaviours associated with religion and multiple non-cultural activities. We extend the cultural and non-cultural activities that can be chosen by the population to multiple dimensions in Subsection 4.6.2.

To summarise the modeling of culture, each individual has a given type, M or m. Each individual i of type $k \in \{M, m\}$ chooses a pair of actions (x_i, y_i), where $x_i \in \{x^M, x^m\}$, $y_i \in \{y^A, y^B\}$. There is no type-specific cost associated with the non-cultural action y_i whilst the cost of cultural action x_i is

$$c_k(x_i) = \begin{cases} 0 & \text{if } x_i = x^k \\ c & \text{if } x_i \neq x^k \end{cases}$$

4.2.2 Social Interaction

Were this the end of the model, individuals would never choose to pay the cost and switch culture. However, there is another issue at stake. Social interaction within the population is valuable, providing opportunities for economic exchange of varying types. Social ties (or 'links') allow for exchange of valuable information. For example, personal contacts play a large role in information about job opportunities and referrals, suggesting an important effect of social interaction on employment outcomes and wages (see Jackson, 2009, for a review of the literature). Social interaction can also provide other economic opportunities, including trade, favour exchange, or economic support such as risk sharing and other valuable joint endeavours (Angelucci et al., 2015). We assume all individuals provide the same opportunities for economic exchange.[21]

Crucially, personal connections and social interaction require commonality of some degree. If two individuals do not speak the same language it limits exchange of information, discussion and agreement on trade, and any other activities that involve verbal communication (Lazear, 1999). Diversity more generally has been found to reduce communication and interaction within organisations (see Williams and O'Reilly, 1998, for a review). Communication difficulties aside, if two individuals take part in completely different activities then not only do they rarely meet (thus reducing opportunity for exchange), but even when they do meet they may not have relevant information to exchange.[22] There are two sides to this story: on the one hand a lack of commonality can make forming a tie more costly or difficult, and, on the other, it will make a tie less valuable if there are complementarities in information exchange and activities.

We model this formally as follows. Individual i chooses whether or not to form a social tie with the other $n-1$ individuals in the population. If individual i forms a social tie with individual j, we denote this by $g_{ij} = 1$, if not $g_{ij} = 0$. There is a fixed cost to forming a social tie, L. Player

[21] There are two ways to relax this assumption in the current framework. The first is for one group to have greater opportunities, so that, all else equal, members of that group would make a more valuable partner. The other is to include some degree of 'love of diversity' to allow for benefits of getting different information and opportunities from different types. To introduce this one might assume that the first tie with someone of a different type is highly valuable, while the value may decline the more contacts of that type. The key trade-off of the model remains in place provided coordination remains important to interaction.

[22] For further discussion on the need for coordination in interaction see Lazear (1999) and Kuran and Sandholm (2008).

i's choice of social ties can be represented by a vector of 0's and 1's, $g_i = (g_{i1}, g_{i2}, \ldots, g_{in})$, where $g_{ii} = 0$. If i forms a social tie with j, the value of that social tie is increasing the more activities the two individuals have in common. The value i receives from a social tie with j is

$$\alpha\pi_1(x_i, x_j) + (1-\alpha)\pi_2(y_i, y_j) - L,$$

where

$$\pi_1(x_i, x_j) = \begin{cases} 1 & \text{if } x_i = x_j \\ 0 & \text{if } x_i \neq x_j \end{cases}$$

$$\pi_2(y_i, y_j) = \begin{cases} 1 & \text{if } y_i = y_j \\ 0 & \text{if } y_i \neq y_j \end{cases}$$

and $0 < L < 1$.

The benefit from a tie is increasing in commonality of actions, while the cost is fixed. Note that the parameter α measures the relevance of cultural versus non-cultural actions in economic exchange. The framework can be interpreted in one of two ways. Either, there is a fixed value normalised to 1 of having a personal connection with another individual in the population, where the cost or difficulty of forming that tie depends on how much the two individuals have in common. Alternatively, we can think of a fixed cost L to forming a personal connection, where the possibility (or value) of economic exchange is increasing the more individuals have in common.

4.2.3 Payoffs

Each individual $i \in \{1, .., n\}$ chooses a cultural action $x_i \in \{x^M, x^m\}$, a non-cultural action $y_i \in \{y^A, y^B\}$, and social ties $g_i = (g_{i1}, g_{i2}, \ldots, g_{in})$. An individual's strategy is thus represented by the vector

$$s_i = (x_i, y_i, g_i) \in S_i. \tag{4.1}$$

The utility of individual i of type $k \in \{M, m\}$ is given by

$$u_k(s_i, s_{-i}) = \sum_j (\alpha\pi_1(x_i, x_j) + (1-\alpha)\pi_2(y_i, y_j) - L)g_{ij} - c_k(x_i). \tag{4.2}$$

The model presented describes an n-player game where each player $i \in \{1, \ldots, n\}$ chooses a strategy $s_i \in S_i$ and receives a payoff $u_k(s_i, s_{-i})$. The strategy profile $(s_1^*, s_2^*, \ldots, s_n^*)$ is a Nash equilibrium if $u_k(s_i^*, s_{-i}^*) \geq u_k(s_i, s_{-i}^*)$ for all $i \in N$ and for all $s_i \in S_i$. We refer to a strategy profile $(s_1, \ldots, s_n)$ as a 'state'. We examine pure strategy equilibria.

The set-up is akin to a version of a battle-of-the-sexes game between groups. An individual gets a higher payoff the better he coordinates with his social ties, allowing for greater economic exchange. But, he would rather coordinate on his own cultural practices. The framework has the additional twist that each individual chooses his social ties. An individual also gets a higher payoff the more social ties he has with whom he is coordinating, again allowing for greater economic exchange. The assumption here, that more contacts are better, has support from the literature on social interaction (Currarini et al., 2009; Currarini and Vega-Redondo, 2011). It implies that individuals might trade-off cultural costs against the benefits of more contacts.[23]

Note two things. Individual i cannot differentiate his choice of action by social tie. Intuitively, this requires that an individual be 'consistent' in his behaviour within the given population.[24] For example, it is often not possible to adopt two different religions. Religious activities might take place at the same time, the different practices might be contradictory, it could be too time-consuming to do both, or it might be made impossible because of hostility from others. Second, social ties are one-sided. If i forms a social tie with j then the value of the social tie accrues to individual i but not to individual j unless individual j forms a social tie with i. In equilibrium, however, if i forms a social tie with j then j will also form a social tie with i. The model of link formation presented here is not the only way to model social interactions and is chosen to simplify the exposition (see Subsection 4.6.3 for further discussion).[25]

4.3 Analysis of the Model

In this section we characterise the Nash equilibria of the game and provide comparative statics.[26] We first introduce two assumptions made to simplify the exposition. Assumption 4.1 rules out the uninteresting case where a single individual prefers to maintain his cultural action even if the rest of the population all adopt a common, different cultural action.

Assumption 4.1. *(No Man is an Island)*

$$(1 - L)(n - 1) - c > \max\{0, (1 - \alpha - L)(n - 1)\}.$$

[23] We discuss alternative specifications of this assumption in Subsection 4.6.3.

[24] A population consists of n individuals who each have the opportunity to interact with all the others. Thus a population refers to a workplace, a neighbourhood, a school, etc. An individual may interact in multiple populations and play different strategies in different populations. This framework then does not prohibit different behaviours at work or school from those in the neighbourhood, for example.

[25] The Nash equilibria in this chapter also all satisfy the definition of *pairwise stability*, an important measure of stability in network formation when link formation is reciprocal (Jackson and Wolinsky, 1996), although these are not the only pairwise stable outcomes. Alternatively, a two-sided link formation model, related to pairwise stability, where an individual could delete any number of links and form any number of agreed upon links would produce the same outcomes as the link formation model we use (see Subsection 4.6.3 for more details of this issue also).

[26] In Section 4.7 we consider a multiple generation framework, and study which of these Nash equilibria are persistent when we introduce cultural transmission, and what influence this has on the comparative statics.

Second, we assume that when an individual is indifferent between forming a tie or not, the tie is formed.[27]

4.3.1 Characterisation of Nash Equilibria

Proposition 4.1 describes the Nash equilibria of the game. Nash equilibria take one of three contrasting forms: (*i*) assimilation, (*ii*) segregation, or (*iii*) multiculturalism. Nash equilibria are characterised by thresholds on the share of immigrants in the population: if the immigrant group is small, assimilation states are the only equilibria; if the immigrant group forms a large enough share of a population, different social structures can emerge in equilibrium. For convenience we first introduce our formal definition of assimilation, segregation, and multiculturalism.

Definition 1. *A state $s = (s_1, \ldots, s_n)$ is defined as:*

> ***Assimilation*** *if all individuals adopt the same actions, $x_i = x_j$, $y_i = y_j$ $\forall i, j \in N$, and each individual forms a social tie with all other individuals.*
>
> ***Segregation*** *if type M adopt action x^M, type m adopt action x^m, and each individual forms a social tie to all other individuals of the same type as him but does not form a social tie to individuals of a different type.*
>
> ***Multiculturalism*** *if type M adopt action x^M, type m adopt action x^m, both types adopt the same non-cultural action $y_i = y_j$ $\forall i, j \in N$, and each individual forms a social tie to all other individuals.*

Proposition 4.1. *States which satisfy the definition of assimilation, segregation, and multiculturalism are the only possible Nash equilibria of the game. Further,*

(i) Any assimilation state is always a Nash equilibrium;

(ii) There exists a Nash equilibrium which satisfies the definition of segregation if and only if the share of the minority group in the population weakly exceeds δ;

(iii) There exists a Nash equilibrium which satisfies the definition of multiculturalism if and only if the share of the minority group in the population weakly exceeds η;

where

$$\delta = \max\left\{\frac{1}{2} - \frac{c+L-1}{2n(1-L)}, \frac{1-\alpha-L}{2(1-\alpha)-L} + \frac{1-\alpha}{(2(1-\alpha)-L)n}\right\}, \qquad \eta = \begin{cases} \frac{1}{2} - \frac{c-\alpha}{2n\alpha} & \text{if } 1-\alpha \geq L \\ 1 & \text{if } 1-\alpha < L \end{cases}$$

[27] Both can be relaxed.

The proof is found in Section 4.6. Only three types of social structure are Nash equilibria. Assimilation: one group pays the cost of switching cultural action to facilitate interaction with the other group. Segregation: groups do not pay the cost of switching cultural action but restrict interaction within their own group. And a third structure, multiculturalism, in which groups each maintain their respective cultural action but adopt a common non-cultural action. Crucially, coordination on these non-cultural practices enables interaction across groups.[28]

In our framework, an individual's payoff from taking an action is weakly increasing in the number of others that take that same action. It is important to highlight the role that such strategic complementarities play in our findings. First, this interaction between the individual's payoff and what the rest of the community are doing gives rise to multiple equilibria, allowing similar populations to end up in contrasting states. To see this, suppose individual i's group assimilates, so everyone in the population adopts the other group's cultural practices. Maintaining his own cultural practices then becomes very costly to i since it hampers interaction with everyone else, thus he can do no better than assimilate also. Hence assimilation is clearly an equilibrium. Now consider the other extreme: suppose everyone in i's group segregates and maintains their own cultural practices. Maintaining his own cultural practices is now much less costly to i, since he can still interact fruitfully with everyone in his group. Thus segregation may also be sustainable as a Nash equilibrium. Second, strategic complementarities create a multiplier effect, driving members of a group towards doing the same thing.[29] Even if we add some within group heterogeneity – see Subsection 4.6.4 for this addition to the model – because of the multiplier effect, members of the same group will still have a tendency to do the same thing (even if not everyone acts identically).

The two features discussed above lead to the key finding that equilibria are characterised by thresholds on the share of the minority group. When the minority group is small, assimilation states are the only equilibria; segregation and multiculturalism are equilibria only in populations where the share of the minority group is above the thresholds given in Proposition 4.1. An easy way to see this is to observe that maintaining a distinct culture is only 'worth it' if enough others do so as well. Segregation is an equilibrium when no individual in the minority group is willing to pay to switch culture and interact with the larger majority group, which is true only when the share of the minority group is large enough. Thus only if the share of the minority group reaches

[28]These three distinct forms of social structure are evident in scholarly discussion of communities in the United States following mass migration (Gordon, 1964). Assimilation among immigrants in certain communities persuaded early scholars that a single 'American culture' would prevail. As it became clear that not all communities assimilated, but some 'retained distinctive economic, polical and cultural patterns long after arriving in the United States', scholars accepted that assimilation was not the only possibility for diverse communities (Bisin and Verdier, 2000). Even early on, however, there was discussion of a 'third way' where immigrants could become 'American' and integrate but also maintain some cultural distinction.

[29]By multiplier effect we mean the following. Once one individual adopts a particular action this raises the relative payoff to that action, which may then induce other agents to adopt the action, which further raises the payoff to that action, which may induce further agents to adopt, and so on.

this 'critical mass' or 'threshold' can segregation be sustained. An analogous intuition holds for multiculturalism.[30] At these critical masses, a small change in the share of the minority group can result in a large change in equilibrium social structure. This is illustrated in Figure 4.1.

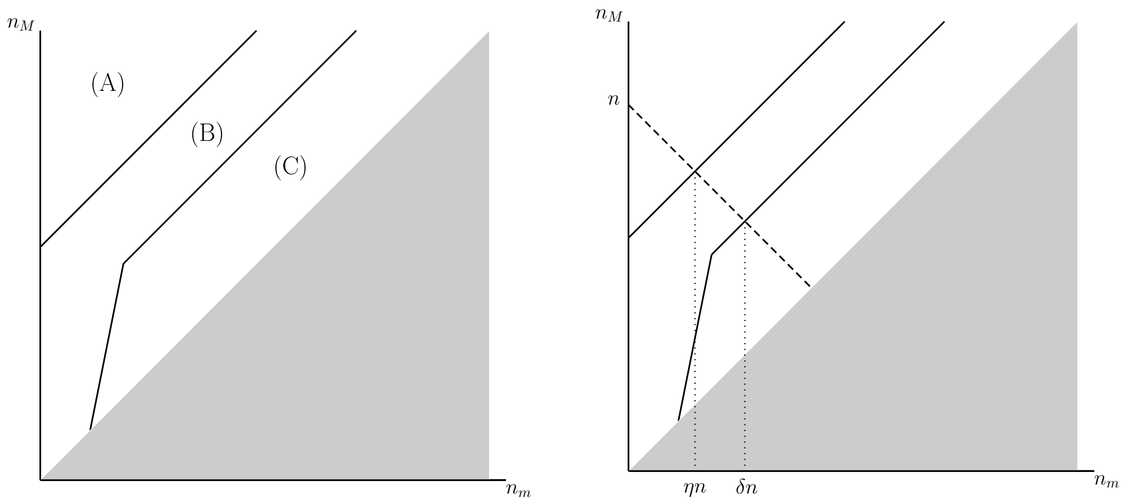


Figure 4.1: *Nash equilibria and minority group size*

The Nash equilibria are illustrated for parameter values $1 - \alpha \geq L$. The axes measure the size of each group. Group M is only in the majority above the 45° line, so the area below the 45° line is greyed out. Assimilation is a Nash equilibrium in areas (A), (B), and (C) in the left hand panel of Figure 4.1. Multiculturalism is an equilibrium in areas (B) and (C). Segregation is an equilibrium in area (C) only. The right hand panel of Figure 4.1 illustrates the same graph, highlighting the results for a population of fixed size n. The dashed line shows all possible shares for the minority group, from $n_m/n = 0$ to $n_m/n = 1/2$, for a population of a given size n. The dotted lines illustrate, for this population of size n, the size of the minority group above which multiculturalism and respectively segregation are equilibria. Note, the size of a minority group that can sustain multiculturalism is smaller than the size of the minority group that can sustain segregation because the cost to the minority group of multiculturalism is less than the cost of segregation.

4.3.2 Comparative Statics

The parameters c, α, and L shift the location of the thresholds described in Proposition 4.1. That is, c, α, and L determine the size of the critical mass of immigrants that is necessary to sustain segregation or multiculturalism. The parameter L is the cost of forming a link. Parameters c and α both measure 'culture' but have distinct interpretations. The magnitude of c is the cost of adopting the other group's cultural action. For example, it might be less costly to switch from speaking German to English, which is a closely related language, than it would be to switch from Italian. Contrast this with the parameter α, which measures the importance of cultural activities relative to non-cultural activities in everyday interaction and economic exchange. A high α environment is one in which cultural practices are frequently relevant to interaction. For example, a environment in which most social activities taking place are organised by churches

[30]Under multiculturalism the minority group will interact with the majority group, but the benefit of interaction is lower than if they were to assimilate, and therefore the same intuition holds.

and other religious organisations. In this example, a low α environment is one in which social activities related to religion comprise only a small part of daily life.

Corollary 4.1. *The threshold share of the minority group above which segregation states are Nash equilibria, δ, is decreasing in cultural distance, c; decreasing in the importance of culture, α; and decreasing in the cost of forming a link, L.*

The threshold share of the minority group above which multiculturalism states are Nash equilibria, η, is decreasing in cultural distance, c; but increasing in the importance of culture, α; and increasing in the cost of forming a link L.

When cultural distance between groups, c, is higher, switching culture is more costly, and so smaller minority groups are more willing to maintain their own culture. This makes it easier to sustain both segregation and multiculturalism and the respective thresholds both fall.

Counterintuitively, an increase in the importance of culture, α, makes groups less willing to maintain their own culture under multiculturalism (the threshold rises). This is because, under multiculturalism, a higher α makes having a common culture more important to interaction, so maintaining one's own culture is more costly in terms of lost opportunities for exchange. Indeed, when culture 'dominates' everyday life (precisely $\alpha > 1 - L$), social interaction based on common non-cultural actions is not enough to sustain a tie and multiculturalism breaks down. To see this, consider again an environment where most social activities are related to religion. It is then difficult for individuals to maintain their own religious practices and integrate with those who adopt different practices. The set of equilibria may be reduced to just segregation and assimilation. An increase in α has the opposite effect on segregation making it a less costly option. This is illustrated in Figures 4.2 and 4.3.

A higher cost of forming a link, L, makes it easier to sustain segregation (lowers the threshold), by making assimilation a less attractive option for minority members. The costs of assimilation come from switching culture, while the benefits come from improved interaction with natives; a higher cost of link formation lowers the relative benefits of interaction with natives. In contrast, a higher cost of forming a link makes it harder to sustain multiculturalism. When the costs of forming a link are high, these can outweigh the benefits of interaction with the other group under multiculturalism.

To summarise, we find that one of only three structures – assimilation, segregation or multiculturalism – can arise in heterogeneous communities in equilibrium. Equilibria are characterised by a threshold, or 'critical mass'. When the minority group is small, assimilation is the only equilibrium. The size of the critical mass of immigrants necessary to sustain segregation or multiculturalism depends on cultural distance, the importance of culture in everyday life, and the cost of link formation. In Section 4.4 we find evidence of such thresholds in heterogeneous communities in the United States in the age of mass migration: if immigrant groups hit a critical mass in the local community (the magnitude of which we estimate) they are much more likely to maintain their own practices and segregate.

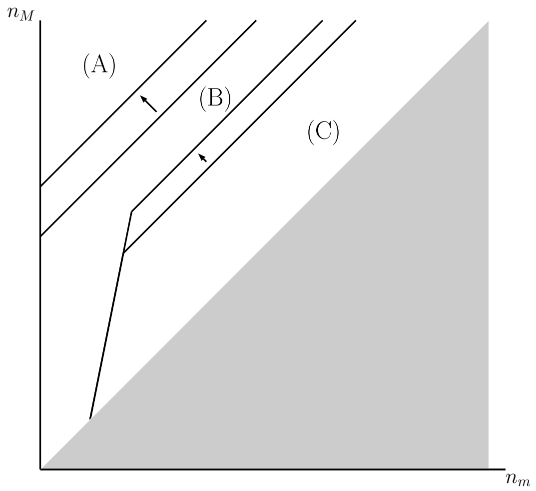


Figure 4.2: *An increase in c*

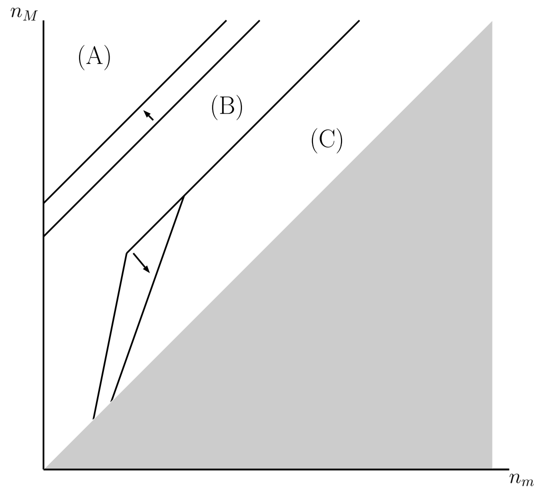


Figure 4.3: *A reduction in α*

Figure 4.2 replicates Figure 4.1 and illustrates an increase in c. The arrows show that areas (B) and (C) expand with an increase in c. That is, the parameter ranges under which segregation and multiculturalism are equilibria expand with an increase in c. Figure 4.3 replicates Figure 4.1 and illustrates a reduction in α. The arrows show that area (B) expands and area (C) shrinks with a reduction in α. That is, the parameter range under which multiculturalism is an equilibrium expands with a reduction in α and the parameter range under which segregation is an equilibrium shrinks.

A low importance of culture, α, appears at first to be a positive force for cross-group interaction since it makes it easier to sustain multiculturalism. However, our second comparative static result shows that a lower importance of culture leads to more polarising behaviour in segregation equilibria.

Corollary 4.2. *When $\alpha \leq 1 - L$ segregation is a Nash equilibrium when $n_m/n \geq \delta$ if and only if groups adopt different cultural actions* and *different non-cultural actions.*

When the importance of culture is low, $\alpha \leq 1 - L$, the segregation Nash equilibria require that the two groups adopt different cultural actions *and* different non-cultural actions. Groups differentiate their practices above and beyond *ex ante* cultural differences. This is not simply a result of coordination on different non-cultural actions. Instead the minority group must adopt a different non-cultural action y from the majority, in order to raise the cost of interaction with the majority group such that segregation can be sustained. This outcome is an equilibrium, despite the reduction in economic exchange that it entails. We refer to this type of segregation equilibrium as 'extreme segregation'.

Non-cultural actions, that is activities and practices with no type-specific costs, play an important role in the social structures that emerge in equilibrium. It has long been recognised that when groups with distinct cultures interact, individuals do not simply choose between which of these cultures to practice. Instead, interaction can produce new 'emergent cultures', where actions that were previously culturally inconsequential take on an important role in uniting or dividing the community. This result can be seen clearly in the equilibria produced in our model.

In the multiculturalism equilibrium, groups maintain their different cultural actions, but the two groups adopt a common non-cultural action, either y_A or y_B. Construction of common practices is necessary to sustain integration when groups also retain some diversity of practices. This equilibrium can be related to the frequently advanced ideal that populations can maintain, for example, their diverse religions while at the same time having a common national identity and culture.[31] We show that this 'ideal' is feasible when culture is not 'all encompassing' in society; groups must be able to find enough common ground to be able to interact.

At the other end of the spectrum, under the extreme segregation equilibria, practices can become more polarised: groups differentiate their behaviours and practices beyond different cultural actions. Harris (2009) writes that some minority groups hold 'secondary cultural differences ... that emerge after the two groups have been in continuous contact.' We show that when α is low, segregation can be maintained only when each group coordinates on distinct non-cultural actions. Rather than bridging the gap between groups, the non-cultural action is used to emphasise differences between them. For example, groups practising two different religions *ex ante* will also differentiate other behaviours that are un-related to religious requirements (for example, dress code, increased food restrictions, and even sports).[32] This appears to be a novel explanation of this type of behaviour and we find it to be a positive signal of a model purporting to explain cross-cultural interaction that there is an equilibrium admitting this possibility.[33]

[31] As early as 1915, at a time of fervent discussion of integration of immigrants in the United States, Harvard philosopher Horance Kallen described the possibility of the United States being 'a democracy of nationalities, cooperating voluntarily and autonomously through common institutions' where 'the common language ... would be English, but each nationality would have ... its own peculiar dialect or speech, its own individual and inevitable esthetic and intellectual forms.' This is sometimes now referred to as the 'salad bowl'.

[32] Berman (2000) describes the birth of Ultra-Orthodox Judaism in the late 18th and 19th Centuries which followed emancipation and the possibility of greater integration with the local European populations. The Ultra-Orthodox 'were not only conservative about rejecting new forms of consumption ... but amplified existing restrictions', such as introducing new dietary restrictions, and 'changed existing customs (dress codes, speaking Yiddish) into religious acts'.

[33] A number of papers examine this type of outcome in detail including Berman (2000) on Ultra-Orthodox Judaism; Austen-Smith and Fryer (2005) on why some black students deride working hard at school as 'acting white' and act in opposition to this; and Akerlof and Kranton (2002) and Bisin et al. (2011b) on the emergence of oppositional identities, whereby groups increase their identification with their own culture in order to reduce the psychological cost of interacting with those who adopt a different culture.

4.4 Evidence from the Age of Mass Migration

The 'Age of Mass Migration' in the late 19th and early 20th centuries is one of the most important episodes of migration of the modern era. From 1830 to 1930, 38 million people arrived in the United States. They joined a population that in 1830 consisted of only 13 million people, and which Alba (1985) describes as having both 'culture and institutions, [that] derived largely from the English models'. Immigrants came to America from around the world, including virtually every country in Europe.[34]

We look at heterogeneous communities throughout the United States in this era and ask how this heterogeneity manifested itself. Did immigrants adopt the behaviours and practices of the local population? Did they integrate or form segregated communities? The setting provides a natural environment in which to examine some of the implications of our model. Importantly, it provides us with a large number of heterogeneous communities with different immigrant groups of varying sizes.

We test two key predictions of our model: that group behaviour changes sharply once the share of the immigrant group in the community reaches some (*ex ante* unknown) critical threshold, and that the location of this threshold varies in the predicted way with the cultural distance between immigrants and natives. In particular, we consider two choices facing immigrants: speaking English, and in-marriage (homogamy).[35]

Testing these predictions of our model is important for two reasons. First, if our framework captures an important trade-off, then we should expect to observe such thresholds in data. Second, evidence of such thresholds, and hence of the forces which drive our model, has many implications for policy, some of which we discuss in the concluding section.

4.4.1 Data Description

Data Sources

We use data from the 1900, 1910, 1920, and 1930 census samples provided by IPUMS. These are 5% (1900 and 1930) or 1% (1910 and 1920) samples. These censuses asked individuals whether or not they speak English. They also allow us to link (almost all) married individuals to their spouses, so we can measure whether individuals marry endogamously (i.e. within the same nationality).[36] Important to our question, in these particular years individuals were asked their

[34] Not all immigrants remained in the US permanently (Bandiera et al., 2013). We will discuss later the implications of this for our results.

[35] In contrast to our baseline modeling assumption, immigrants could continue to speak both their native language and learn English. As mentioned previously, our results will hold provided that the relative benefit of learning English increases the greater the number of members of the immigrant group who learn it.

[36] Individuals whose spouses do not live in the same household cannot be matched. They make up less than 3.1% of married individuals in our sample.

country of origin, and – unlike earlier and later censuses – their year of arrival in the US, allowing us to control for how long individuals have lived in the US. Focusing on adult household heads, we have a repeated cross-section with 611,000 migrants.

The census samples provide information on an individual's place of residence down to the county level. Counties are small administrative units, with on average 63 counties per state, and an average population of 5300 households per county in 1900, of which 1300 were headed by immigrants.[37] We also use data from Spolaore and Wacziarg (2009), collected by Dyen et al. (1992) and Fearon (2003), on the linguistic distance between pairs of languages.

Data Construction

We treat counties as the relevant population for our analysis. That is, individuals are presumed to make their decisions about what behaviours to adopt and with whom to form ties, taking those in the county as the pool of people they could potentially interact with.[38]

Within each county, we define an immigrant group as consisting of all household heads of the same nationality (country of origin), with the exception that we group Germany, Austria, and Switzerland into a single group; Sweden, Denmark, and Norway into a single group; and the Netherlands and Belgium into a single group.[39, 40] Our qualitative results are unchanged in the absence of this aggregation. Henceforth, by 'nationality' we refer to these groups. In each county, the 'nationality share' is the proportion of sampled adult household heads in that county who have that nationality: this is the empirical analogue of n_m/n in our model.[41]

Since our model predictions are at the group level, we aggregate the behaviour of individuals into 'cohorts'. Cohorts are defined by nationality, d; year of arrival (in 10 year bands), a; time

[37] Note, the difference between these figures and Table 4.1 comes both from changes over time in county size (the table pools all censuses), and from the fact that larger counties have more observed cohorts, so receive more weight in the table.

[38] Counties are the finest population partition available to us. It seems possible that decisions are made based on a more local population of people, such as the town/village. In that case we observe only a noisy measure of the population share and population decisions, likely attenuating our estimated effect and reducing power.

[39] We focus on household heads since at the time they were likely to have made choices for the whole household. In the case of in-marriage it also avoids the double counting of homogamous relationships. Finally, household heads are less likely to have received education in the US which might influence the outcomes we observe.

[40] We made these groupings due to use of a common language and because accounts of communities in the US suggest these amalgamations are appropriate. Austrians spoke, in main part, a form of German, and the Swiss who emigrated were largely German speaking. At the time we are considering, Danish remained the official language of Norway, and also Swedish, Danish, and Norwegian are considered mutually intelligible. The Belgians who emigrated to the US were largely of Flemish descent, and so spoke Dutch.

[41] Note that this treats all individuals who are not members of that nationality group as though they were members of the majority group. In all counties, the largest other group that immigrants could consider joining is the native group. Hence, if an immigrant group is considering switching to another culture, this is likely to be the most profitable. Although in some cases it may be lower cost to join another immigrant group due to their lower cultural distance, this is likely mitigated both by our amalgamation of individuals from certain countries and the fact that these groups are typically much smaller than the natives. In our analysis sample there are only 33 counties where more than one group exceeds 10% of the county population, 4 counties where more than one group exceeds 15%, and none at a threshold of 20%. An alternative empirical specification would be to define the denominator for nationality share as the sum of the number of immigrants of that nationality and natives, thus excluding other immigrants. Our results are robust to this redefinition.

since immigration (also grouped, to the nearest 10 years), t; and county, o.[42] Splitting the observed behaviour of immigrant groups into these cohorts allows us to control for arrival year and tenure effects that might be important in explaining behaviour and might vary systematically by nationality. For example, both ability to speak English and out-marriage are likely to be positively related to tenure in the United States. Additionally, varying conditions and policies over the decades we study, such as use of schooling to influence assimilation (Bandiera et al., 2017), may have affected these decisions.

We measure English acquisition by the proportion of a cohort that reported to the enumerator that they spoke English.[43] For this analysis we exclude British, Irish, and Canadian immigrants, since their mother tongue is likely to be English anyway. To proxy for social segregation, we construct a measure of the extent to which each cohort marries people of the same nationality. We calculate the proportion of married members of the cohort that are married to someone of the same nationality.[44] Although marriage is a partial description of social connections, it is clearly defined and measured for this period, and marriage decisions are considered to be strongly revealing of the dimensions that divide a society (for a review see McPherson et al., 2001).

Table 4.1 shows the mean and variance for these outcomes, as well as nationality size (n_m in our model), county size (n), and nationality share (n_m/n), across the analysis sample. We also provide the within-nationality variance, and the share of total variance that is within variance. From this it can be seen that there is significant within nationality variation that we can exploit in our analysis, allowing us to avoid concerns that the effects we capture come only from differences across nationalities.

To measure the cost of learning English, c in our model, we use the data on linguistic distance. The data from Dyen et al. (1992) define linguistic similarity between a pair of languages based on the proportion of frequently used words that share a common root.[45] We assume that immigrants' mother tongue is the majority language spoken in their country of origin. We also assume that the cost of switching language to English is greater for those immigrants whose mother tongue is more dissimilar to English. We then split cohorts into 'near' or 'far' from

[42] In particular, arrival years are grouped as 1886-1895, 1896-1905, and so on. Tenure (time since immigration) is then measured as the difference between the midpoint of the arrival year band (1890, 1900, etc) and the census year.

[43] Enumerators were instructed 'Where difficulty is encountered in making the head of the family understand what is wanted, you should call upon some other member of the family who is able to speak English . . . if no member of the family can aid you in your work, then the assistance of some neighbour of the same nationality and able to speak English should be obtained, whenever possible.' Where this was also not possible, paid interpreters were used. Hence this question is likely to have been able to meaningfully capture whether individuals were able to speak English. https://www.census.gov/history/pdf/1900instructions.pdf

[44] The Coleman index provides an alternative popular measure of segregation (see Currarini et al., 2009). It is not suitable for our context as it requires knowledge of the potential marriage market for a particular cohort at the time marriage decisions were taken, something we do not observe. Our preferred measure is therefore the simple ratio of in-marriage to total marriage for a cohort, which has the additional benefit of being easy to interpret.

[45] This measure of distance is only available for Indo-European languages. We impute some of the missing data using a variable from Fearon (2003) which measures linguistic proximity using a 'language family tree'. Our results are qualitatively robust to instead dropping these observations with missing data.

Table 4.1: *Descriptive statistics for key variables*

	(1) Mean	(2) Total Variance	(3) Within Var	(4) Within/Total
Share speaking English	.901	.023	.011	.489
Share of in-marriages	.609	.055	.022	.391
Nationality Size ('000s)	.459	.391	.374	.957
County Size ('000s)	8.79	106.9	96.5	.903
Nationality Share	.085	.007	.005	.732
Observations	1746	1746	1746	1746

Notes. *Observations are at the cohort level. Cohorts are defined by nationality, county, tenure in the US grouped to the nearest 10 years, and year of arrival in 10 year bands. We include only cohorts with at least 30 individuals, to ensure that cohort averages are well measured. Share speaking English measures the proportion of a cohort that reports speaking English. Share of in-marriages measures the proportion of married individuals in a cohort who are married to someone of the same nationality. Nationality size measures the number of people, in thousands, of the same nationality (potentially of different cohorts) living in the same county. County size measures the number of people, in thousands, living in the same county. Nationality share is the ratio of nationality size to county size. Columns (1) and (2) provide the mean and variance for these variables. Column (3) shows the remaining variance after removing nationality fixed effects, and Column (4) provides the share of total variance that remains once nationality fixed effects are removed.*

English, defining a language as far from English if less that 40% of the words are mutually intelligible with English.[46]

For our analysis we also only consider cohorts with at least 30 observations, so that the average group behaviour is well estimated. This means that we exclude counties where the number of immigrants was very small. There is a trade-off between choosing a low minimum threshold for immigrant group size, to minimise the bias that comes from effectively ignoring very small community shares, and choosing a high threshold to reduce noise in estimated group behaviour. We test the sensitivity of our results to this by also using cut off points of 20 or of 40, and find no qualitative difference. As an alternative sensitivity test, we also performed our analysis with weighted regressions. In particular, the weight given to an observation was equal to the number of observations with the same nationality share in the full sample, divided by the number with that nationality share in the estimation sample.[47] This over-weights observations with small nationality shares, since these are the ones excluded by our group size thresholds. Again we find similar results. These results are presented in Section 4.8.

In total we have 1746 cohorts, when we impose a minimum cohort size of 30, of which Germans and Italians are the biggest groups. Table 4.7 in Section 4.8 shows the number of immigrant cohorts we observe, and splits out some of the larger nationalities. It also shows how this varies depending on the minimum size we require for a cohort to be included in our sample.

[46] We take this cut-off from Advani and Rasul (2015), who choose it because there are no languages in the range $30 - 50\%$ mutually intelligible with English, so it forms a natural break. According to this metric, Spanish and Greek are considered far from English, whilst German and Danish are considered close.

[47] For constructing weights we split nationality share into deciles, with all observations in the decile being given equal weight.

4.4.2 Empirical Framework

To formally test for the existence of a threshold in community behaviour we use a linear regression of the form

$$Y_{dato} = \beta_0 + \beta_1 \mathbb{1}\{z_{do} > \tau\} + \gamma_0 z_{do} + \gamma_1 z_{do} \mathbb{1}\{z_{do} > \tau\} + \kappa_d + \lambda_a + \nu_t + \varepsilon_{dato} \tag{4.3}$$

where Y_{dato} is the share of immigrants of nationality d, arriving in year a, with tenure t, in county o, who speak English (respectively, in-marry); z_{do} is the nationality share of immigrants of nationality d in county o; and τ is the proposed threshold immigrant share. κ, λ, and ν are respectively the nationality, arrival year, and tenure fixed effects. The coefficient of interest is β_1: the coefficient on the indicator function showing whether there is a 'break' in the level of speaking English (respectively, in-marriage) when the nationality share z_{do} crosses the threshold τ.

Since we do not know theoretically where the threshold τ should be, we use an iterative regression procedure, and test for significance using a Quandt Likelihood Ratio (QLR) test (Quandt, 1960; Andrews, 1993). We perform a sequence of regressions, testing a prespecified range of values for τ. For each regression we calculate the F-statistic, comparing the model with the threshold to the same model but without a threshold. We then select from among these regressions, the one with the highest F-statistic. The corresponding break point in that regression is then taken as the estimated location of the threshold, denoted τ^*. To test whether this threshold value is 'significant', we compare the F-statistic to the limiting distribution for this statistic under the null (Andrews, 1993), thus correcting for the multiple testing.[48, 49]

We also estimate Equation 4.3 separately for cohorts that are 'far from' and 'near to' English, and compare the point estimates, to see whether $\tau^*_{far} < \tau^*_{near}$, for both speaking English and in-marriage. However, we are unaware of any formal way to test for the statistical significance of any difference in location.

4.4.3 Results

Threshold in proportion speaking English

Figure 4.4 shows graphically the evidence for a threshold in the share of the immigrant group that speaks English, as its share of the community rises. Each dot shows the mean proportion that report speaking English for a two percentage point range of nationality share. That is, the

[48] For robustness, we also use an entirely non-parametric approach to finding the threshold location, as proposed by Henderson et al. (2015). The intuition of this approach is similar to the parametric approach, searching over a sequence of thresholds, but without any functional form restrictions away from the threshold. The location of the thresholds we find are the same as for the parametric approach.

[49] An alternative method, followed by Card et al. (2008), is to split the sample into a test sample, which one can use to select between multiple potential thresholds, and an analysis sample where the selected threshold can be used. This allows conventional testing approaches with more standard distributions, but at the cost of reduced test power.

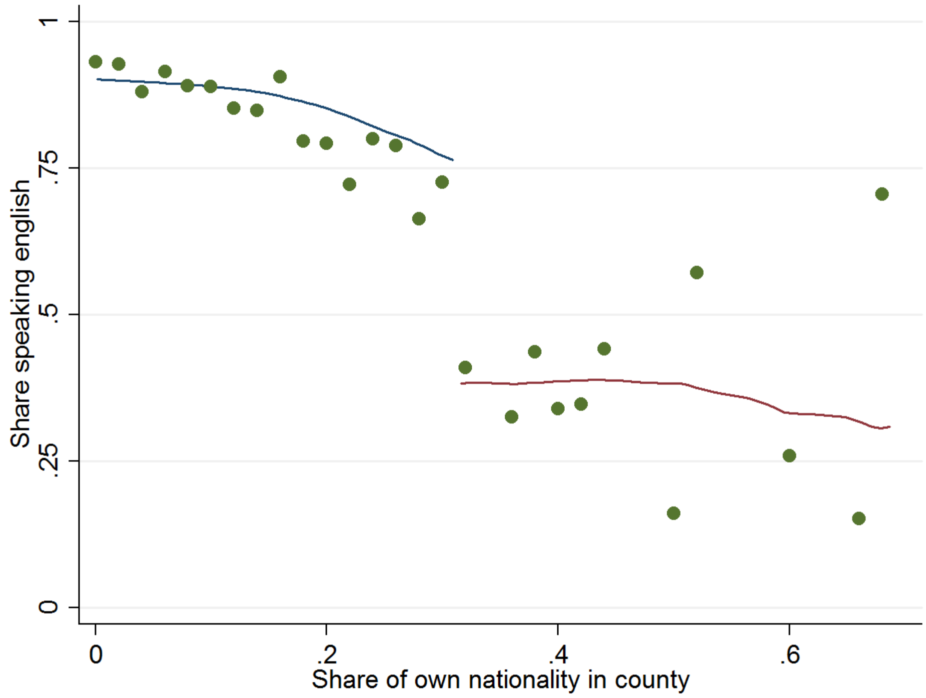


Figure 4.4: *Share of people in the cohort that speak English*

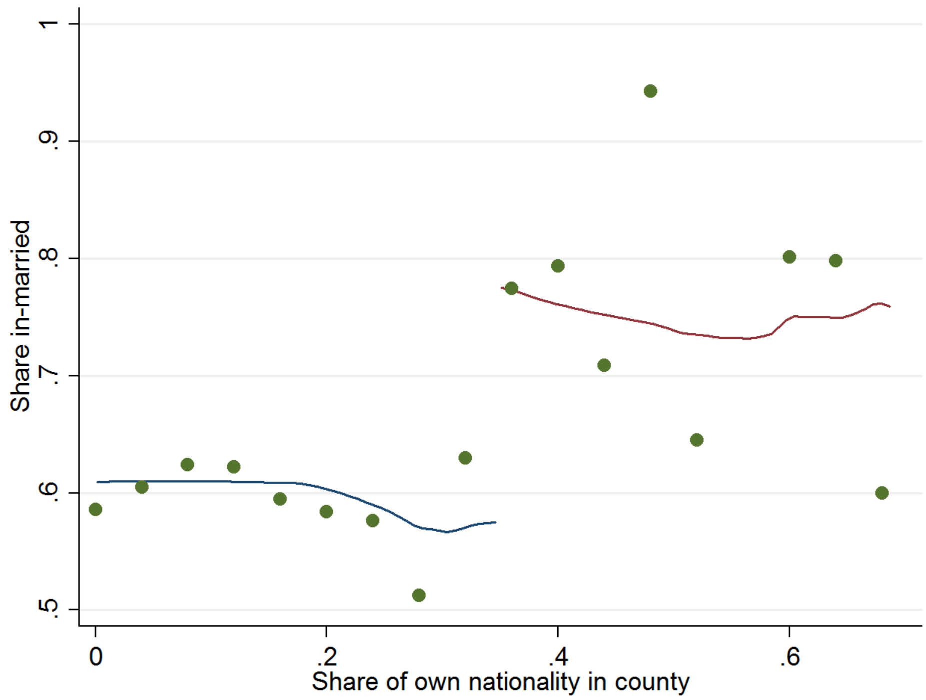


Figure 4.5: *Share of married people in the cohort that are married to someone of the same nationality*

Notes. *The figures show the relationship between the share of a cohort that speak English (Fig 4.4) or marries within its nationality (Fig 4.5), and the proportion of adult household heads in their county that are of the same nationality. Dots show the mean share speaking English/in-married, grouping nationality share into bins two percentage points wide. The line shows the kernel-weighted local mean, estimated using an Epanechnikov kernel with bandwidth of 7% fit through the whole (i.e. unbinned) data. In each figure the local mean is estimated either side of an estimated threshold: .31 for Fig 4.4 and .35 for Fig 4.5. These points are chosen using the Quandt Likelihood Ratio procedure descibed in Subsection 4.4.2. Cohorts are defined by nationality, 10 year grouped arrival year, 10 year grouped tenure in the US, and county of residence. A minimum cohort size of 30 is used.*

Table 4.2: *Testing for a threshold effect in share speaking English*

Dependent Variable: Proportion of people in the cohort that speak English
Standard Errors in Parentheses
Cohorts defined by Nationality, County, Grouped Year of Arrival, Grouped Tenure

	(1) Unconditional	(2) With Fixed Effects	(3) With Slopes
Optimal threshold ($\beta_1 \vert \tau = \tau^*$)	-.509***	-.257***	-.375***
	(.025)	(.017)	(.062)
Constant (β_0)	.885***	.878***	.913***
	(.004)	(.002)	(.004)
Optimal Threshold Level (τ^*)	.31	.31	.31
F-statistic	423	241	37
1% critical value for F-statistic	6.6	6.6	6.6
Cohort Fixed Effects	No	Yes	Yes
Slopes in Nationality Share	No	No	Yes
Observations	1272	1272	1272

Notes. **** denotes significance at 0.1%, ** at 1%, and * at 5% level, when treated as a standalone regression. The outcome measures the proportion of the cohort that speaks English. Cohorts with English, Canadian, and Irish nationalities are excluded from the sample, since English is likely to be their mother tongue. Minimum cohort size of 30. Cohort fixed effects are composed of nationality, arrival year (grouped), and tenure (grouped) fixed effects. Column (1) is a regression of share speaking English on a constant, and a dummy for whether nationality share in the county exceeds a threshold. Column (2) allows for cohort fixed effects. Column (3) allows the share speaking English to also have a slope in nationality share, and allows this slope to vary either side of the threshold. All covariates except the threshold are demeaned, so the constant (β_0) can be interpreted as an estimate of the mean share speaking English at nationality shares below the threshold. These specifications are run sequentially at different values of the threshold, varying the threshold between .20 and .40, at intervals of .01. We provide the results for the threshold among these which produced a regression with the highest F-statistic when tested against the null of no threshold. The value for this threshold, the estimated F-statistic, and the 1% critical value for this statistic (which corrects for the repeated testing, see Andrews, 1993) are provided at the bottom of the table.*

leftmost dot shows the mean proportion of immigrants speaking English for all immigrants in groups that constitute 0-2% of their community. We also plot a kernel-weighted local mean, fitting it separately either side of a nationality share, z_{do}, of .31 (below we describe how this point was chosen). From this figure we can immediately see three things: (*i*) when immigrants make up only a small share of the population, the proportion of the cohort that speaks English is high, at around 90%; (*ii*) there is a sharp drop in this mean, to less than 40%, when the immigrant group reaches approximately 1/3 of the population; (*iii*) there is more variation in the mean after the threshold. These three findings are consistent with our model. The first result shows clear evidence that small groups of migrants are assimilated into American culture, almost all learning the language. The visually striking threshold is clearly suggestive of the type of strategic complementarity notion at the heart of our model. And the final result is consistent with the existence of multiple equilibria above the threshold, so that in some communities we may see assimilation outcomes even at these high nationality shares.

The choice of .31 for the break in our local mean plot comes from performing a formal test for the presence and location of a significant break. We use the QLR test described above, searching a grid between [.20, .40] with increments of .01, and test for the most likely value of the threshold, τ, in a sequence of increasingly flexible models. We consistently find a significant break in English acquisition, and find the most likely value for the threshold is when the immigrant group constitutes 31% of the population. Table 4.2 shows the results of the tests.

The first specification includes only a constant and a threshold term, so that any systematic variation in the share of the cohort speaking English can only be picked up as some kind of threshold effect. This amounts to imposing that the slope terms are zero ($\gamma_0 = \gamma_1 = 0$) and ruling out nationality, year of arrival, and tenure fixed effects ($\kappa_d = \lambda_a = \nu_t = 0$) in Equation 4.3. We find a strong support for the presence of a threshold, with an F-statistic of 423 (compared with a 1% critical value of 6.6, taken from Andrews, 1993), and the most likely location for the threshold is $\tau^* = .31$.[50]

Our next specification allows for nationality, arrival year, and tenure fixed effects, so we only impose $\gamma_0 = \gamma_1 = 0$. This allows for the possibility that, for example, Germans might live in groups with systematically smaller nationality shares than Italians and also find it easier to learn English, although even this sort of story is unlikely to give rise to so clear a threshold as seen in Figure 4.4. Similarly we now allow for variation in settlement patterns and language acquisition by groups arriving in different years, and who have been present in the US for different tenures. Again we can strongly reject that no threshold exists (F-statistic of 241), with the same location. We also now see that the point estimate for $\beta_1|\tau = \tau^*$ is reduced, showing that some of the reduction in English speaking is explained by the fixed effects.

Finally, in Column (3) we allow also for slopes either side of the potential threshold. Although our model as written does not have such effects, simple extensions of the model, such as allowing some heterogeneity in costs or benefits, might allow some sort of slope either side of the threshold. Even allowing for such effects, we find continued support for a threshold, again when immigrants make up 31% of the local population, with an F-statistic of 37.

Threshold in proportion in-married

We now repeat the above analysis using proportion in-married as the outcome variable. Figure 4.5 is constructed in the same way as Figure 4.4, and tells a similar story to that seen with speaking English. We see that: (*i*) when immigrants make up only a small share of the population, the assimilation type outcome is apparent, with relatively less in-marriage; (*ii*) in-marriage rises sharply after the threshold; and (*iii*) we see increased variation after the threshold. The picture is less sharp here, since the choice of whether to in-marry only captures a single, limited dimension of interaction between different groups. Additionally, since the timing of the marriages are unknown, some of the in-marriage might reflect relationships formed prior to immigration,

[50] These results are robust to minimum cohort size chosen. Table 4.8 replicates Column (1) using different sample size thresholds, and also using the weighting method discussed earlier. The location and significance of the threshold τ^* remains unchanged. The same threshold is also found when we use the completely non-parametric search procedure suggested by Henderson et al. (2015).

Table 4.3: *Testing for a threshold effect on share in-married*

Dependent Variable: Proportion of the married people in the cohort that are married to someone of the same nationality
Standard Errors in Parentheses
Cohorts defined by Nationality, County, Grouped Year of Arrival, Grouped Tenure

	(1) Unconditional	(2) With Fixed Effects	(3) With Slopes
Optimal threshold ($\beta_1 \vert \tau = \tau^*$)	.151***	.065***	.166***
	(.042)	(.017)	(.049)
Constant (β_0)	.606***	.606***	.562***
	(.006)	(.003)	(.006)
Optimal Threshold Level (τ^*)	.35	.23	.23
F-statistic	13.0	15.0	11.5
1% critical value for F-statistic	6.6	6.6	6.6
Cohort Fixed Effects	No	Yes	Yes
Slopes in Nationality Share	No	No	Yes
Observations	1746	1746	1746

Notes. **** denotes significance at 0.1%, ** at 1%, and * at 5% level, when treated as a standalone regression. The outcome measures the married proportion of the cohort that is 'in-married' i.e. married to someone of the same nationality. Minimum cohort size of 30. Cohort fixed effects are composed of nationality, arrival year (grouped), and tenure (grouped) fixed effects.*

Column (1) is a regression of share in-married on a constant, and a dummy for whether nationality share in the county exceeds a threshold. Column (2) allows for cohort fixed effects. Column (3) allows the share in-married to also have a slope in nationality share, and allows this slope to vary either side of the threshold. All covariates except the threshold are demeaned, so the constant (β_0) can be interpreted as an estimate of the mean share in-married at nationality shares below the threshold. These specifications are run sequentially at different values of the threshold, varying the threshold between .20 and .40, at intervals of .01. We provide the results for the threshold among these which produced a regression with the highest F-statistic when tested against the null of no threshold. The value for this threshold, the estimated F-statistic, and the 1% critical value for this statistic (which corrects for the repeated testing, see Andrews, 1993) are provided at the bottom of the table.

giving a relatively high level of in-marriage even when nationality shares are very low. Despite these limitations, at the threshold we see a jump of one-fifth, from less than 60% to more than 75% in-married.

Table 4.3 shows our main formal results on in-marriage, again testing for the presence and location of a significant break using the QLR procedure described above (searching a grid between [.20, .40] with increments of .01). Column (1) estimates our empirical specification without fixed effects ($\kappa_d = \lambda_a = \nu_t = 0$) or slopes ($\gamma_0 = \gamma_1 = 0$). We find a significant break (F-statistic of 13.0 against a 1% critical value of 6.6) in in-marriage when the immigrant group constitutes 35% of the population. Levels of in-marriage rise from 60% to 75% at this threshold.[51] Column (2) includes fixed effects, and Column (3) additionally allows slopes to vary. We continue to find a significant break in in-marriage (F-statistics of 15.0 and 11.5 respectively), although it now occurs earlier, when the immigrant group constitutes 23% of the population.

In summary, we show evidence of stable levels of in-marriage in communities where the immigrant group is below approximately a quarter of the population (once we account for fixed

[51] These results are robust to minimum cohort size chosen. Table 4.9 in Section 4.8 replicates Column (1) using different sample size thresholds, and also using a weighting method, and significance of the threshold τ^* remains unchanged. The same threshold is also found when we use the completely non-parametric search procedure suggested by Henderson et al. (2015).

Table 4.4: *Comparing threshold locations for linguistically far and near cohorts*

Dependent Variable: (A) Proportion of people in the cohort that speak English;
(B) Proportion of cohort in-married
Standard Errors in Parentheses
Cohorts defined by Nationality, County, Grouped Year of Arrival, Grouped Tenure

	(A) Speaking English		(B) In-Married	
	(1) Ling. Far	(2) Ling. Near	(1) Ling. Far	(2) Ling. Near
Optimal threshold ($\beta_1 \| \tau = \tau^*$)	-.118***	-.115**	-.031	.182*
	(.029)	(.035)	(.034)	(.079)
Constant (β_0)	.876***	.977***	.761***	.402***
	(.005)	(.007)	(.006)	(.011)
Optimal Threshold Level (τ^*)	.31	.39	.39	.23
F-statistic	17.0	10.7	.80	5.3
5% critical value for F-statistic	3.8	3.8	3.8	3.8
1% critical value for F-statistic	6.6	6.6	6.6	6.6
Cohort Fixed Effects	Yes	Yes	Yes	Yes
Slopes in Nationality Share	Yes	Yes	Yes	Yes
Observations	810	462	810	936

Notes. **** denotes significance at 0.1%, ** at 1%, and * at 5% level, when treated as a standalone regression. In Panel A the outcome measures the proportion of the cohort that speaks English, and cohorts with English, Canadian, and Irish nationalities are excluded from the sample. In Panel B the outcome measures the married proportion of the cohort that is 'in-married' i.e. married to someone of the same nationality. Minimum cohort size of 30. Cohort fixed effects are composed of nationality, arrival year (grouped), and tenure (grouped) fixed effects.*
All columns present regressions of the outcome variable on a constant, an indicator for whether nationality share is above a threshold, cohort fixed effects, and nationality share itself (allowing for different slopes either side of the threshold). In both panels, Column (1) includes only the subset of cohorts from our main sample that come from countries whose main language is deemed far from English (less than 40% mutual intelligibility), whilst Column (2) uses those who are linguistically near. All covariates except the threshold are demeaned, so the constant (β_0) can be interpreted as an estimate of the mean share speaking English at nationality shares below the threshold. In each case we vary the threshold between .20 and .40, at intervals of .01. We provide the results for the threshold among these which produced a regression with the highest F-statistic when tested against the null of no threshold. The value for this threshold, the estimated F-statistic, and the 5% and 1% critical values for this statistic (Andrews, 1993) are provided at the bottom of the table.

effects). When the immigrant group reaches this nationality share we find a significant jump upwards in rates of in-marriage. This discontinuous increase in segregation, even when using a partial measure of interaction, provides strongly supportive evidence of the mechanisms driving our model.

Higher thresholds when culturally closer

Our model predicts not only the presence of a threshold, but also that the threshold should vary with c. In Table 4.4 we now estimate the location and significance of the threshold separately for groups who are from countries with a language 'near to' or 'far from' English. Splitting the sample into these two cases, we estimate the unrestricted version of Equation 4.3, analagous to the third columns of Tables 4.2 and 4.3. Figures 4.6 and 4.7 show the results graphically.

For speaking English (Panel A) we find a significant threshold at the 1% level for both far and near nationalities. As predicted, for near nationalities the threshold occurs at a larger nationality share ($\tau^* = .39$) than for far nationalities ($\tau^* = .31$). This is consistent with our theoretical

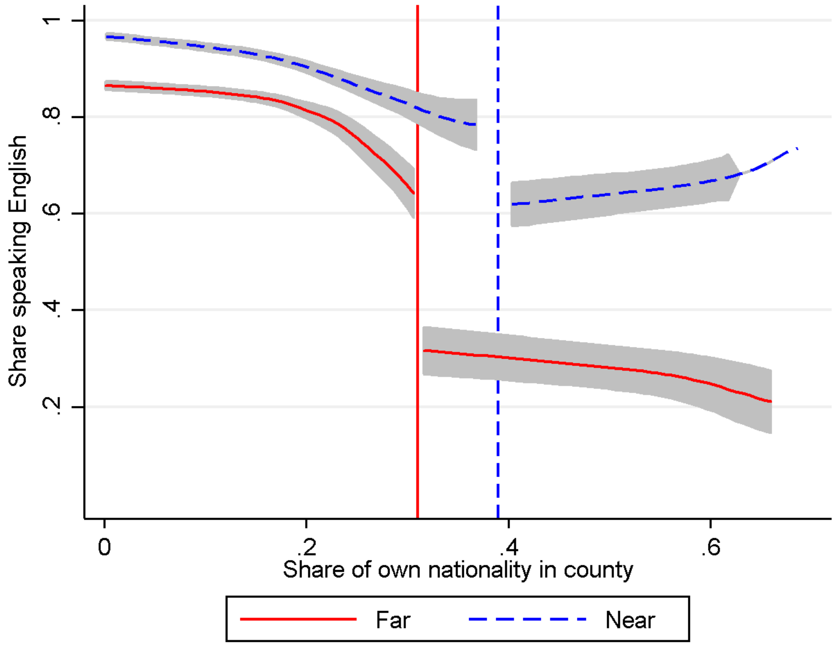


Figure 4.6: *Share of people in the cohort that speak English, for linguistically near and far cohorts*

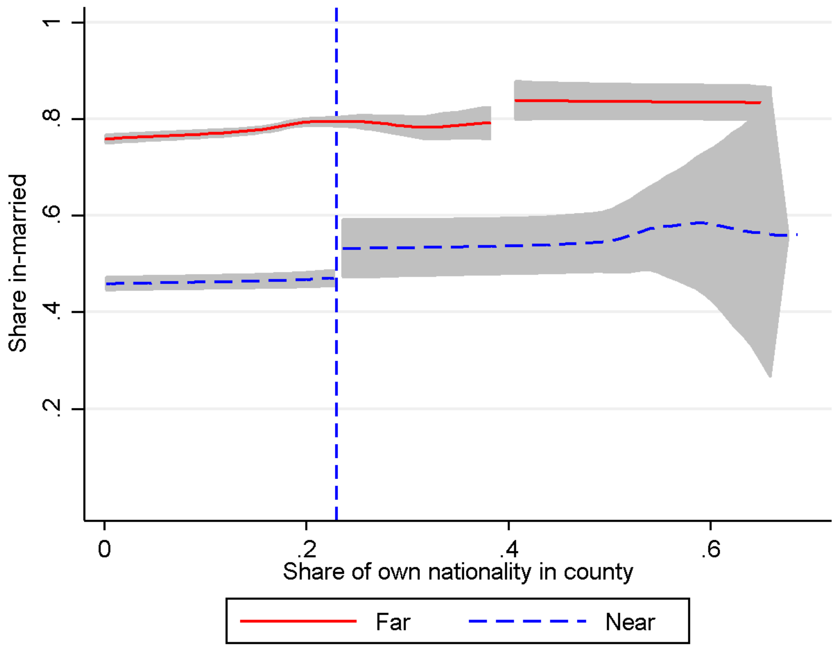


Figure 4.7: *Share of married people in the cohort that are married to someone of the same nationality, for linguistically near and far cohorts*

Notes. *The figures show the relationship between the share of a cohort that speak English (Fig 4.4) or marries within its nationality (Fig 4.5), and the proportion of adult household heads in their county that are of the same nationality, splitting cohorts into those which are linguistically near and those which are linguistically far. The line shows the kernel-weighted local mean, estimated using an Epanechnikov kernel. The shaded areas denote the 95% confidence intervals. In each figure the local mean for the near and far cohorts is estimated separately either side of an estimated threshold, chosen using the Quandt Likelihood Ratio procedure descibed in Subsection 4.4.2. A vertical line is shown in the cases where the threshold was found to be significant in the parametric testing. Cohorts are defined by nationality, 10 year grouped arrival year, 10 year grouped tenure in the US, and county of residence. A minimum cohort size of 25 is used.*

finding that nationalities that are culturally closer to English need to make up a relatively larger proportion of the local population before they choose to retain their own language.

For in-marriage (Panel B), we see a significant threshold still for the near group, at 23% (significant at 5% level), but we do not find a significant threshold for the far group. In part this is likely due again to our measure of interactions being only a partial one.

4.4.4 Discussion of results and limitations

In interpreting our results, we have so far abstracted from the important issue of location choice and the influence that selection might have on our results. We argue that the presence of such selection does not change our ability to draw conclusions about the key trade-offs inherent in our model.

The first question of selection is whether migrants choose where to live based on their individual costs and benefits of interaction, and of the actions they take. The worry might be that any heterogeneity in the cost of switching cultural practices (for example, different costs of learning English or different preferences for maintaining religious practices) will manifest as sorting into different areas. Immigrants with higher costs of switching cultural practices would have a higher relative return from locating in areas where their group continues to maintain their own practices. Such selection would only be a problem for us if we were to imagine that there were discontinuities in the distribution of heterogeneity. Without this, the discontinuity we observe in behaviour must be driven by the structure of the game played by individuals in these communities. Sorting by individuals may still mean that migrants with different costs choose to locate in systematically different communities but, as argued by Lazear (1999), this is simply 'a question of timing'. If individuals sort, knowing that after choosing their location their action choice and payoff will depend on those around them, then this is still supportive of the mechanism of our model. The policy implications will differ, however, if some of the difference between the assimilated and segregated communities comes from differences in the costs of assimilation.

The second role for selection relates to who stays in the US. We know from Bandiera et al. (2013) that there was significant churn, with 60-75% of migrants leaving the US. If migrants knew *ex ante* that they were planning to leave the US and go back to their native country, then they would discount the benefits of adopting the practices of the new community.[52] The benefits would now only be felt for a more limited period, whilst the costs of adoption would remain unchanged. This is equivalent to these individuals having the same benefits but higher costs of learning English, and hence the same logic as selection into location applies.

[52] If the individuals do not know that they might leave, then it cannot influence their decision.

Clearly we do not capture all the richness of the theoretical model. For example, it may be that multiculturalism manifests as the whole community speaking English to one another and sharing some other norms of behaviour, but at the same time different immigrant groups maintaining different religious practices, different cuisines, or other different activities. By looking only at the choice to learn English or not we cannot discern such an outcome. Similarly we may miss evidence of 'extreme segregation' because we do not have a rich enough set of outcome variables. An extension of the empirical analysis that looks at choices along more dimensions might allow us to better 'pull apart' the different equilibria. However, the data requirements for such an analysis are strong, including information on social interaction and various action choices among multiple heterogenous populations. Although we can only pick up some of the detail of the theoretical model in our empirical analysis, the variables we do consider show clear evidence that the shape of the relationship is consistent with our predictions.

4.5 Discussion and Concluding Remarks

This chapter builds a tractable framework to answer two questions: what social structures can arise in heterogeneous populations, and what environmental features determine which social structure arises. At the center of our analysis is the idea that group behaviour and practices, as well as social cohesion, are endogenous. Heterogeneity itself, as well as social relations in heterogeneous populations, greatly depend on the environment.

Understanding how the environment shapes heterogeneous populations is critical for policy design. Integration of immigrant and minority groups receives a lot of political attention, and governments frequently propose and implement policies intended to influence the way minority groups integrate. Our findings paint a nuanced picture of both the relevance and effects of such policies. We briefly consider what our framework implies for two important policies.

First, our model suggests that secularisation policies, rather than restraining religion, can actually make groups more likely to maintain diverse religious practices. Secularism prescribes some degree of restriction on religion in public life. In France in 2004 the principle of secularism was applied to ban all conspicuous religious symbols (including veils of any kind) from public schools. Our framework shows that policies which reduce the relevance of cultural practices in everyday life, such as removing religion from state institutions, can enable multiculturalism. That is, it implies that maintaining diverse religions is easier in a more secular society!

Second, our framework implies that policies which reduce barriers to interaction across groups (e.g. school bussing or desegregation) may result in a response whereby groups not only remain segregated but amplify their differences. Reduced costs of interaction require minority groups to distinguish themselves *even more* if they are to remain segregated. Thus increased opportunities for interaction can lead to 'secondary' differences where minority groups begin to create new distinguishing features and emphasise previously insignificant behaviours as culturally important.[53]

Our framework does not rule out a role for policy however. The key welfare tension in the model comes from the inability of group members to coordinate, which can lead to 'inefficient segregation'.[54] A failure to coordinate among immigrants allows segregation to arise in situations where a collective move to assimilation or multiculturalism would be a Pareto improvement. This is because when groups are segregated under Nash equilibrium, a minority individual does worse by switching culture and joining the majority group, even though the minority group as a whole might do better by adopting majority culture. Policy can influence this by reducing the costs of switching culture for at least some minority individuals, thus breaking down the segregation equilibrium. Schooling policy provides an example of this, where the teaching of national values to the children of immigrants makes it easier for them to integrate with natives, thus making it harder for their parents to remain segregated (see Alesina and Reich (2015), Bandiera et al. (2017) and references therein).

Of course, the costs of switching culture and the ubiquity of cultural and religious practices in daily life could be influenced not only by governments by also by individuals and groups. 'Group leaders' might be able to influence these parameters to achieve their individual aims. For example religious leaders might care most about preserving religious adherence, and attempt to adjust the cost of switching practices or increase the ubiquity of religious activities, even at the cost of lowering the utility of group members. Alternatively, the costs of switching culture and the relevance of cultural and religious practices could emerge endogenously in a decentralised model. This chapter takes a key step in understanding endogeneity of heterogeneity.[55] However, there remain many important avenues for further work in this direction.

A second important direction for future research would be to allow for not just more than two types of individual but different dimensions of types. For example, we may have two different nationality groups with *ex ante* different practices, but each nationality group may itself be composed of a mix of people of two different religions which have their own associated practices. An illustration is provided by the Wars of the Three Kingdoms, which saw Protestants in England, Scotland, and Ireland (before the formation of the United Kingdom) fighting together against Catholics from their own countries. Adapting our framework with such an addition may provide answers to questions about when and why one divide (for example religion) becomes more important than another (for example ethnicity) in different societies.

[53]This result is also present in Bisin and Verdier (2000) and Bisin and Verdier (2000), but through a different mechanism

[54]See Subsection 4.6.5 for an expanded welfare analysis.

[55]For work on the endogeneity of national culture see Alesina and Reich (2015).

4.6 Theoretical Appendix

4.6.1 Proof of Proposition 4.1

In a Nash equilibrium all players of the same type must play the same pair of actions. Suppose not, and that two players of the same type, i and j, play different pairs of actions. One player, without loss of generality player i, must do weakly worse than the other. Then i can do strictly better by mimicking j's strategy and also forming a tie with j. Forming a tie with j ensures that player i either has an additional tie compared to j's previous strategy, or that the value i receives from the tie between himself and j is higher than the value j would have been receiving from this tie, since they now coordinate on more actions. Hence in equilibrium i and j must play the same actions. In equilibrium each player must also link with all others who adopt the same action pair since the value of such a tie is strictly positive. Thus all individuals in the same group form a tie in equilibrium.

Suppose $1-\alpha < L$. Suppose there is (at least) one tie between different types in equilibrium. If the two types do not have culture in common then the payoff from this tie is at most $1-\alpha-L < 0$, and the individual can do strictly better by dropping the tie. In equilibrium the two groups must have culture in common for there to be a tie between groups. Suppose the two groups have culture but do not have non-cultural practices in common. Then a player from the weakly smaller group can do strictly better by adopting the non-cultural action of the weakly larger group and linking accordingly. Therefore, if there is a tie between groups in equilibrium, all individuals must play the same action pair and thus all individuals must be linked.

Suppose there are no ties between different types in equilibrium. If the two types adopt the same cultural practices but do not have non-cultural practices in common, then a player from the weakly smaller group can do strictly better by adopting the non-cultural action of the weakly larger group and instead linking with that group. If the two types adopt the same action pair, then they would have strictly higher utility by linking across groups. Thus if there are no ties between different types, in equilibrium the two types must adopt different cultural actions. Further, each type must adopt its own cultural action since, if not, a type M must be doing weakly better by playing x_m and linking with his group than playing x_M and linking with the other group. But then a type m must do strictly better by playing x_m, linking with the other group and not paying the cost of switching culture. A contradiction.

We have ruled out all states apart from assimilation and segregation states as possible equilibria. Assimilation is an equilibrium under all parameter values since we assumed $(1-L)(n-1)-c > \max\{0, (1-\alpha-L)(n-1)\}$, and so all deviations by an individual do worse. Segregation is an equilibrium if $(1-L)(n_m-1) \geq (1-L)(n-n_m)-c$. It is straightforward to see that under this condition all deviations by an individual do worse.

Suppose $1-\alpha \geq L$. Suppose there are no ties between different types in equilibrium. If the two types adopt the same cultural practices but do not have non-cultural practices in common, then a player from the weakly smaller group can do strictly better by adopting the non-cultural action of the weakly larger group and instead linking with that group. If the two types adopt the same action pair, then they would have strictly higher utility by linking across groups. The remaining possibility is that the two types adopt different cultural actions. If the two types adopt the same non-cultural action the value of a tie with the other type is $1-\alpha-L \geq 0$ and so this is not an equilibrium. Thus different types must adopt different cultural and non-cultural actions. Each type must also adopt their own cultural action, by the above.

Suppose there is (at least) one tie between different types in equilibrium. If the two types have cultural practices in common then, by the proof above, they must also have non-cultural actions in common and thus all individuals must be linked. If the two types do not have cultural practices in common, then suppose they also do not have non-cultural practices in common. Then the value of a tie between types is strictly negative and the individual with a tie to a different type does strictly better by dropping it. If the two types do not have cultural practices in common, then the remaining possibility is that they do have non-cultural practices in common. Then the value of a tie with the other type is weakly positive and all individuals must be linked.

We have ruled out all states apart from assimilation, segregation, and multiculturalism as possible equilibria. Assimilation is an equilibrium under all parameter values. Given the assumption $(1-L)(n-1)-c > (1-\alpha-L)(n-1)$, any deviation from assimilation by a lone individual results in a strictly lower payoff.

Segregation is an equilibrium when no player wants to deviate. No player wishes to deviate, play the action pair of the other group, and link only with the other group if and only if

$$(1-L)(n_m-1) \geq (1-L)(n-n_m)-c$$

or deviate, switch non-cultural action, and link with the other group as well as their own type if and only if

$$(1-L)(n_m-1) \geq (\alpha-L)(n_m-1)+(1-\alpha-L)(n-n_m).$$

It is straightforward to see that under these conditions all other deviations by an individual away from segregation do worse. Multiculturalism is an equilibrium when no player wishes to mimic the strategy of the other group if and only if

$$(1-L)(n_m-1)+(1-\alpha-L)(n-n_m)\geq(1-L)(n-n_m)+(1-\alpha-L)(n_m-1)-c.$$

It is straightforward to see that under these conditions all deviations by an individual do worse.

4.6.2 Extending the model to action choices along more than two dimensions

We have modeled action choice along two dimensions. In reality, individuals make many more than two decisions. The minority group may initially speak one language and the majority group another, the minority group may participate in certain religious practices and the majority group others, the culture of the minority group may permit alcohol while that of the majority group forbids it, and so on. The baseline model condenses these many choices into just two: a choice of cultural action, which represents choice over practices for which there is type-specific cost, and a choice of non-cultural action, which represents choice over practices which have no type specific cost. Very little intuition is lost through this abstraction. Here we relax this assumption and highlight the novel features which arise.

Let x_{i1} denote i's choice of language, which is chosen from the set $\{x_1^m, x_1^M\}$, where x_1^m is the language of the minority group and x_1^M is the language of the majority. If two individuals speak the same language then the benefit from interaction increases by α_1. Suppose the cost to switching language is c_1. Let x_{i2} denote i's choice to drink alcohol or not, chosen from the set $\{x_2^m, x_2^M\}$, where group m permits alcohol and group M does not. Suppose α_2 is the importance of having this in common in terms of enabling interaction and economic exchange. Let c_2 be the cost of switching. Generally let $(x_{i1},\ldots,x_{iQ})$ denote the Q 'cultural choices' of individual i. The set $(x_1^m,\ldots,x_Q^m)$ contains the *ex ante* cultural practices of group m, and the set $(x_1^M,\ldots,x_Q^M)$ the *ex ante* cultural practices of group M. Some cultural practices may be very costly to switch away from and others less so, represented by the cost of switching for each practice, c_r, $r\in\{1,\ldots,Q\}$. Some practices may be very relevant to interaction and others less so, represented by α_r, $r\in\{1,\ldots,Q\}$. Non-cultural actions are modeled similarly. Individual i chooses R non-cultural actions, $(y_{iQ+1},\ldots,y_{iQ+R})$, where for each action there is a choice, $\{y_{Q+1}^A, y_{Q+1}^B\},\ldots,\{y_{Q+R}^A, y_{Q+R}^B\}$, where the choice is denoted by the superscripts A and B. Any non-cultural choice has an associated cost of zero, $c_r=0$. The importance of each practice in terms of social interaction is given by α_r, $r\in\{Q+1,\ldots,Q+R\}$. We normalise these to sum to one.

$$\sum_{r=1}^{Q+R}\alpha_r=1.$$

Each individual chooses social ties as before, but now chooses a larger set of actions $(x_{i1}, \ldots, x_{iQ}, y_{iQ+1}, \ldots, y_{iQ+R})$. The utility function is now

$$u_k(s_i, s_{-i}) = \sum_{j=1}^{n} \left[\sum_{r \in \{1,\ldots,Q\}} \alpha_r \pi_r(x_{ir}, x_{jr}) + \sum_{r \in \{Q+1,\ldots,Q+R\}} \alpha_r \pi_r(y_{ir}, y_{jr}) - L \right] g_{ij} - \sum_r c_k(x_i r) \quad (4.4)$$

where

$$\pi(x_{ir}, x_{jr}) = \begin{cases} 1 & \text{if } x_{ir} = x_{jr} \\ 0 & \text{if } x_{ir} \neq x_{jr} \end{cases}$$

$$\pi(y_{ir}, y_{jr}) = \begin{cases} 1 & \text{if } y_{ir} = y_{jr} \\ 0 & \text{if } y_{ir} \neq y_{jr} \end{cases}$$

and for individual i of type $k \in \{M, m\}$

$$c_k(x_i r) = \begin{cases} 0 & \text{if } x_{ir} = x_r^k \\ c_r & \text{if } x_{ir} \neq x_r^k \end{cases}$$

The Nash equilibria and comparative statics with multiple dimensions are analogous to those in the main model presented. We highlight only the notable richness that adding more dimensions brings. Under similar conditions to the main model, assimilation outcomes are Nash equilibria, where all individuals form a tie, all adopt the same practices where one group adopts all the cultural practices of the other group. There now exist what we refer to as 'melting pot' equilibria, a type of assimilation equilibrium where all individuals form a tie and all adopt the same practices (both cultural and non-cultural), but the action choices could be a mix of the *ex ante* cultural practices of the two groups. Thus the *ex post* culture that emerges is a mix of the *ex ante* different cultures (Kuran and Sandholm, 2008). There are also multiculturalism equilibria, similar to main model, whereby all individuals form a tie, have all non-cultural practices in common, but the different groups maintain at least some of their different cultural practices. There is some added richness here which does not appear when there are only two action choices: the smaller the minority group the fewer minority traits they can sustain in a multiculturalism equilibrium. There is a sequence of multiculturalism equilibria, from the case where each group maintains all their own cultural actions, to fewer and fewer of one group's traits being retained, moving towards assimilation. The only other Nash equilibria are segregation equilibria: individuals form ties only within their group and not to the other group, and adopt sufficiently different action sets (this includes versions of the 'extreme segregation' outcome).[56] There is also added richness

[56]The segregation equilibria have the feature that, although a group will mainly adopt its own cultural practices, it may be that a group adopts a few cultural practices of the other group. This is a feature of examining Nash equilibria in a coordination game.

here: the smaller the minority group, the more they must differentiate their practices in order for segregation to be a Nash equilibrium. This is consistent with empirical findings (Bisin et al., 2013) and theoretical findings (Bisin and Verdier, 2000) that smaller minority groups must exert more effort to retain their diverse traits.

4.6.3 Discussion of alternative models of link formation

As mentioned, the model of link formation presented here is not the only way to model social interactions. For example, the Nash equilibria in this chapter all satisfy the definition of *pairwise stability*, an important measure of stability in network formation (Jackson and Wolinsky, 1996). In fact, a two-sided link formation model, related to pairwise stability, where an individual could delete any number of links and form any number of mutually agreed upon links would produce the same outcomes as the link formation model we use. There are many ways to model how individuals form ties and in any framework a choice must be made. The key consideration in choosing the model of link formation here was to avoid a multitude of equilibria (often a feature of social interaction models) that add complication without giving additional intuition with respect to the questions we seek to answer.

Another way of modeling social interaction would be to add a cap on the number of ties an individual can form, or to assume decreasing returns in the number of ties. With such assumptions, the qualitative results would remain in smaller communities and we would see more segregation in larger populations where the cap on the number of contacts becomes relevant to decision making.

4.6.4 Heterogeneous costs of switching culture.

Suppose for members of the immigrant group that the costs of switching culture are given by

$$c_{1m} \leq c_{2m} \leq \ldots \leq c_{n_m m}$$

and similarly for the native group

$$c_{1M} \leq c_{2M} \leq \ldots \leq c_{n_M M}.$$

Let us consider the case where $1 - \alpha < L$. Then individuals will not form a tie if they have only non-cultural actions in common.

Assimilation (to immigrant or native culture) is always an equilibrium provided that the following condition, analogous to the previous *no man is an island* assumption, holds:

$$(1-L)(n-1) - c_h > 0 \quad \forall h \in \{1m, 2m, \ldots, n_m m, 1M, 2M, \ldots, n_M M\}.$$

Suppose all individuals in the population adopt the same cultural action, then by the same argument as the proof of Proposition 1, in equilibrium all must adopt the same non-cultural action and form a link. Thus assimilation states are Nash equilibria under all parameters and they are the only Nash equilibria where all individuals adopt the same cultural action.

Suppose some individuals in the population adopt different cultural actions. Let n_1 be the size of the group adopting x_m and n_2 be the size of the group adopting x_M. In equilibrium those playing the same cultural action must play the same non-cultural action and link otherwise there are profitable deviations.

Suppose $n_1 \geq n_2$. Then any minority individuals playing x^M do strictly better by switching to play x_m and so in any Nash equilibrium all minority individuals must be in group n_1. No minority individual wishes to deviate. A majority individual in group n_1 with cost c_h does not want to deviate if

$$(1-L)(n_1 - 1) - c_h \geq (1-L)n_2$$

and a majority individual in group n_2 with cost c'_h does not want to deviate if

$$(1-L)(n_2 - 1) \geq (1-L)n_1 - c'_h.$$

Note that the final two conditions hold only if all majority individuals in group n_1 have a lower cost of switching, c_h, than all individuals in group n_2.

Instead suppose $n_2 \geq n_1$. In any Nash equilibrium all majority individuals must be in group n_2. No minority individual with cost c_h wishes to deviate from group n_2 if

$$(1-L)(n_2 - 1) - c_h \geq (1-L)n_1$$

No minority individual with cost c'_h wishes to deviate from group n_2 if

$$(1-L)(n_1 - 1) \geq (1-L)n_2 - c'_h.$$

This is the same as above, although it is not guaranteed that it will hold even if the whole minority group is part of n_1. As with the homogeneous cost case segregation only holds when the minority group is large enough.

4.6.5 Welfare results and discussion

We will consider which states are Pareto efficient and which would be chosen by a social planner, where the social planner maximises the sum of utilities in the population

$$\sum_i u_k(s_i, s_{-i}).$$

We highlight three findings.

First, multiculturalism is Pareto efficient under the same parameters that multiculturalism will persist in the long run. Similarly segregation is Pareto efficient exactly when it will persist in the long run.[57] That is, segregation and multiculturalism are likely to emerge in the long-run exactly when this outcome is in the minority group's best interest.[58]

Second, assimilation is always Pareto efficient. A group can do no better than when the other group assimilates, since this maximises interaction and the other group pays the cost of switching culture. This suggests that when a group is able to do so, it will put pressure on other groups to do the assimilating. Calls for immigrants to adopt local languages and cultural practices by the public or by politicians are thus unsurprising. In contrast, a social planner will not always choose assimilation.

The third point to highlight is the result of Proposition 4.2 (below). This reveals a tension between when segregation is a Nash equilibrium, when it is optimal for the minority group, and when it is optimal for the social planner. Proposition 4.2 says that segregation is a Nash equilibrium under a larger set of parameters than those for which it is Pareto optimal, and segregation is Pareto optimal under a larger set of parameters than those for which the social planner would choose to implement segregation. The reason for this tension is that segregation is a Nash equilibrium when, given the population is segregated, no individual from the minority group does better by switching culture and joining the majority group. However, even if a minority individual in this situation prefers to stick with minority culture, it could be that the minority group as a whole would do better by adopting majority culture. The social planner chooses segregation under a yet smaller set of parameters, since the social planner also incorporates the costs of segregation to the majority group. This same reasoning applies to multiculturalism under the parameters $1 - \alpha \geq L$.

[57] The only difference is that we do not need to account for discreteness as in Proposition 4.3.

[58] Excluding the case where the majority group assimilate to minority culture, which the minority group would prefer but cannot influence.

Let δ^W denote the threshold size of the minority group above which segregation is chosen by the social planner, let δ^P denote the threshold above which segregation is Pareto efficient, and δ is the threshold above which segregation is a Nash equilibrium, given in Proposition 4.1. Analogously for η^W, η^P, and η for multiculturalism. Observe in Proposition 4.2 that a social planner would choose multiculturalism when $1-\alpha \geq L$ and the minority group is large, again lending support to the proposed ideal of a multicultural society.

Proposition 4.2. *When* $1-\alpha < L$*:*

$$\delta^W > \delta^P > \delta.$$

When $1-\alpha \geq L$*:*

$$\eta^W > \eta^P > \eta.$$

Proof of Proposition 4.2.

Any assimilation state is always Pareto efficient since any individual whose cultural action is played attains his maximum payoff, doing strictly worse under any other state.

Segregation is Pareto efficient when (4.5) holds and $1-\alpha < L$. Each individual in the minority group gets a strictly higher payoff from segregation than assimilation to majority culture when

$$(1-L)(n_m-1) > (1-L)(n-1) - c$$

rewritten

$$n_m/n > 1 - c/n(1-L). \tag{4.5}$$

It is immediate to see that individuals in the minority group get a higher payoff from segregation than any other state where only the majority cultural action is played. Symmetrically for the majority group. It remains to show that some individual does strictly worse in any state, other than segregation, where at least two individuals adopt different cultural actions. Since $1-\alpha < L$, any individual who links with someone adopting a different cultural action does strictly worse. If fewer individuals play the minority cultural action, then those remaining playing the minority action must be strictly worse off (symmetrically for the majority cultural action). The remaining possibility is that the size of each group playing either cultural action remains the same, but some majority and minority individuals switch strategy. Clearly some individuals do strictly worse.

Multiculturalism is Pareto efficient when $1 - \alpha \geq L$ and the minority group does strictly worse by assimilating

$$(1-L)(n_m - 1) + (1-\alpha-L)n_M > (1-L)(n-1) - c$$

rewritten

$$n_m/n > 1 - \frac{c}{\alpha n}.$$

The proof follows from above, also noting that under $1-\alpha \geq L$ individuals do weakly better by all adopting the same non-cultural action and linking.

Assimilation by the minority gives higher welfare than any other state in which a single cultural action is played. When $1-\alpha < L$, in any state that maximises welfare with two cultural actions being played, any two individuals playing the same cultural action must also play the same non-cultural action and must link, and any two individuals playing different cultural actions must not link. Segregation gives weakly higher total welfare than assimilation by the minority when

$$n_m(1-L)(n_m-1) + n_M(1-L)(n_M-1) \geq n_m[(1-L)(n-1) - c] + n_M(1-L)(n-1)$$

rewritten

$$n_m/n \geq 1 - c/2n(1-L).$$

It remains to show that total welfare is weakly lower when some other combination of individuals play different cultural actions. Suppose a smaller group of minority individuals, $n_m - n_1$, play the minority cultural action and the rest, n_1, play the majority cultural action, for $n_1 \in \{1, \ldots, n_m - 1\}$. Total welfare is higher than under segregation if

$$\begin{aligned}(n_m - n_1)(1-L)(n_m - n_1 - 1) + (n_M + n_1)(1-L)(n_M + n_1 - 1) - n_1 c \\ \geq n_m(1-L)(n_m-1) + n_M(1-L)(n_M-1)\end{aligned}$$

rewritten

$$(n_M + n_1)/n - c/2n(1-L) \geq n_m/n.$$

A contradiction. The same applies if a smaller group of majority individuals play the majority cultural action. Finally it is clear that each type playing their respective cultural action maximises welfare in this situation.

Multiculturalism maximises the social planner's problem when $1-\alpha \geq L$ and it gives higher total welfare than assimilation by the minority, that is

$$n_m(1-L)(n_m-1)+n_M(1-L)(n_M-1)+n_m(1-\alpha-L)n_M$$
$$+n_M(1-\alpha-L)n_m \geq n_m[(1-L)(n-1)-c]+n_M(1-L)(n-1)$$

rewritten

$$n_m/n \geq 1-c/2\alpha n.$$

The result follows similarly to above. □

4.7 Extending the Model: Multiple Generations with Cultural Transmission

Section 4.3 characterises the Nash equilibria of the game. In this Appendix we incorporate the framework into a model of cultural transmission over generations. Considering multiple generations is interesting for two reasons. First, it may be that some outcomes that are Nash equilibria in a single generation are likely to die out in the long run. Second, since the static model has multiple equilibria for a given configuration of parameters, it is useful to consider whether some Nash equilibria are more 'likely' than others, thus the multiple generation framework acts as an equilibrium refinement.

We incorporate dynamic features into our baseline model. First, we assume a multiple generation framework where parents pass on their type to their children (a simplified version of Bisin and Verdier, 2000). Second, we do not assume that the population automatically starts off at an equilibrium. Children, once born, choose their actions and social ties optimally as above, but taking as given the existing choices of the population they are born into. This means that all else equal, a child born into a more assimilated group may behave differently to one born into a more segregated group, again capturing the importance of the behaviour of the rest of the community. Finally, we allow for a few members of a new generation to 'experiment' or 'make mistakes'. We can also think of these as temporary shocks which happen infrequently. For example, suppose the population is in a state where members of the minority group maintain their minority culture. In the next generation, however, a few children in the minority group 'experiment' and try out majority culture. This can set in motion a move to a different Nash equilibrium: if enough people experiment it may be optimal for future minority group children to also adopt majority culture. These shocks allow populations to move away from 'less stable' Nash equilibria and towards 'more stable' ones.[59]

The dynamic is modeled formally as follows. Suppose at (discrete) time t the population of size n is in some state which is described by $(s_1^t, s_2^t, \ldots, s_n^t)$, where s_i^t is the choice of actions and social ties of individual i at time t. A strategy at time t, s_i^t, is given by (4.1), as in the single generation (Nash) case. The state need not be a Nash equilibrium. At time $t+1$ there is an independent probability $q \in (0,1)$ for each individual i (parent) that they die and are replaced by a child, also denoted i. The child is the same type as the parent, M or m, and chooses an action which maximises his utility given the state in period t:

$$s_i^{t+1} \in \arg\max_{S_i} u_k(s_i^{t+1}, s_{-i}^t)$$

where $u_k(s_i^{t+1}, s_{-i}^t)$ is given by (4.2). Under this dynamic the population moves to a Nash equilibrium, and once it reaches an equilibrium it will stay there for all future generations. However, we also allow that with probability ϵ the child instead adopts some other strategy randomly. This is how we model experimentation by children.

We examine the case where the probability of experimenting, ϵ, is small. The process detailed above defines an aperiodic and irreducible Markov chain. The following proposition tells us that, for any given parameter values, a single outcome will be observed almost all of the time.[60] Which outcome is observed in the long-run depends on the parameters of the population in a similar way to the single generation case. We make the following assumption, which ensures cultural distance is sufficiently important that the trade-off between maintaining one's own cultural practices and forming social ties is economically relevant.[61]

Assumption 4.2. $c \geq (1-L)(n_m+1)$

Under Assumptions 1 and 2 we get the following result.

Proposition 4.3. *When the probability of experimenting is small, $\epsilon \to 0$, and $t \to \infty$, the following outcome is observed with probability tending to 1:*

(i) When $1-\alpha < L$ there exists a threshold δ^* such that

Assimilation by the minority group is observed if $n_m/n \leq \delta^* - 1/n$;

Segregation is observed if $n_m/n \geq \delta^* + 1/n$.

[59] As described in Young (1993), the process we model 'selects the equilibrium that is easiest to flow into from all other states combined, including both equilibrium and non-equilibrium states'.

[60] There is a unique stationary distribution which puts all weight on a single state.

[61] It is made for ease of exposition. If Assumption 4.2 does not hold, then the cost of switching culture is low, and both assimilation to M or assimilation to m could persist in the long run. See the proof of Proposition 4.3 for details.

(ii) When $1 - \alpha \geq L + (\frac{1}{n_M - 1})$ *there exists a threshold* η^* *such that*

Assimilation by the minority group is observed if $n_m/n \leq \eta^* - 1/n$*;*

Multiculturalism is observed if $n_m/n \geq \eta^* + 1/n$*.*

In the long-run, a unique outcome emerges (or is more likely to emerge). Proposition 4.3 tells us that if the minority group is small, assimilation occurs and it is the minority group that assimilates. It also shows that multiculturalism can be sustained as a stable long-run outcome in a diverse population. This occurs when culture is not too important in everyday life (α low) and the minority group is large. If culture is very important (α high) and the minority group is large, then we will see long-run segregation of groups.

Analogous to the single generation model, higher cultural costs, c, allow populations with a smaller proportion of minority individuals to sustain segregation or multiculturalism in the long run. A lower importance of culture, α, can ensure multiculturalism is a long-run outcome instead of segregation, and allows multiculturalism to be sustained in populations with a smaller proportion of minority individuals. The result that different cultural practices can persist, under certain parameters, even in the long-run, is consistent with the findings and discussion in Bisin and Verdier (2000).

The proof of Proposition 4.3 involves assessing the minimum number of shocks needed to induce a transition from one Nash equilibrium to another, and then aggregating to determine which of the equilibria is most likely to occur in the long-run. We simplify this by using a 'tree pruning' argument, the details of which can be found below. The $+1/n$ and $-1/n$ terms added to the thresholds and the small positive term $+(\frac{1}{n_M - 1})$ in Proposition 4.3 are there to avoid having to make a more complex statement about what happens at the boundary parameters between one long-run outcome and another. Details and precise values for thresholds can also be found below.

The forces driving the results of Proposition 4.3 are intuitive. First, why does the minority group assimilate and not the majority group? Suppose in generation t the minority group and majority group adopt their respective cultural practices. In generation $t+1$ some members of the minority group experiment by adopting majority culture and interacting with the majority group. If enough members experiment, then future minority group members will choose to switch and also adopt majority culture. Since the minority group is smaller, less experimentation is required to induce further minority group members to optimally choose to assimilate to majority culture than if we repeat the same scenario with the majority group. For a similar reason, assimilation occurs only when minority individuals constitute a small share of the population: the smaller the minority group, the less experimentation it takes to induce movement in the minority population towards assimilation.

A natural extension of the multiple generation framework would allow the costs of practices to change over generations. If a parent assimilates, the cost to the child of switching culture may be less than the cost the parent faced. Alternatively, where a parental generation segregates and further differentiates non-cultural actions, the child may face type-specific costs over the actions for which the parents faced no cost.[62] This would be an interesting avenue for future research.[63]

Proof of Proposition 4.3

The process described defines an aperiodic and irreducible Markov Chain; it therefore has a unique stationary distribution. We briefly describe our notation and method of computing the stationary distribution as $\epsilon \to 0$. We refer to the strict Nash equilibria of the game, denoted here by σ, as *nodes*. There is a minimum number of players required to make a mistake in their strategy for there to be a positive probability of transiting from one strict Nash equilibrium, call it σ', to another strict Nash equilibrium, say σ, by the best response dynamics. We refer to this minimum mistake transition from node σ' to node σ as the *link* from σ' to σ and the minimum number of players required to make a mistake as its *weight*. A path from node σ' to σ is a sequence of directed links starting at σ' and connecting to σ. A $\sigma - tree$ is graph with no cycles such that for each strict Nash equilibrium $\sigma' \neq \sigma$ a unique path connects it (directly or indirectly) to σ. The *stochastic potential* of σ is the total weight of the $\sigma - tree$ whose links sum to the lowest total weight. A *stochastically stable state* is a state with positive probability in the stationary distribution when $\epsilon \to 0$. The stochastically stable states are the strict Nash equilibria with minimum stochastic potential. This result and details can be found in Young (1993).[64]

To find the $\sigma - tree$ for each strict Nash equilibrium σ would involve enumerating a lot of possibilities. We use a tree-pruning argument to simplify things. We first give the proof for the case where $1 - \alpha < L$, and then for the complementary parameter range.

[62] Observe, it is not the case that cultural costs necessarily fall over generations. Suppose one individual from the parent generation experiments and assimilates while the rest of his group segregates. Then the cost to his child of speaking the other group's language falls somewhat. However that child may still find it preferable to choose to segregate, in which case the costs of speaking the other group's language may go up again for the grandchild.

[63] One way to do this would be to combine the current framework with features of Kuran and Sandholm (2008), which allows costs of taking an action to change as actions taken change.

[64] This result also requires that the process without experimenting converges almost surely to a strict Nash equilibria. This is straightforward and, for completeness, is found at the end of this Appendix.

Proof for the case when culture is relatively important

The proof here is given for the parameter range $1 - \alpha < L$. The result for the reverse inequality is given below.

Lemma 4.1. *Partition the strict Nash equilibria into three sets: segregation; assimilation to M ('M-assimilation'); and assimilation to m ('m-assimilation'). Such a set is denoted Σ_{z_1}, and $z_1 \in \{s, M, m\}$ respectively indexes the set. All equilibria in a set, $\sigma \in \Sigma_{z_1}$, have the same stochastic potential.*

Proof. Equilibria in a given set are symmetric so, for any $\sigma - tree$, an equivalent $\sigma - tree$ can be constructed with the same total weight for any other equilibrium in the same set.

Lemma 4.2. *A strict Nash equilibrium $\hat{\sigma} \in \Sigma_{z_1}$ has lower stochastic potential than a strict Nash equilibrium $\tilde{\sigma} \in \Sigma_{z_2}$, $z_2 \neq z_1$, if the link from $\tilde{\sigma}$ to $\hat{\sigma}$ has lower weight than the link from $\hat{\sigma}$ to σ for all $\sigma \in \Sigma_{z_3}$, for all $z_3 \neq z_1$.*

Proof. Take a tree which defines the stochastic potential of node $\tilde{\sigma}$. Denote by $p_{\sigma_1\tilde{\sigma}}$ the path from some node denoted σ_1 to $\tilde{\sigma}$ in such a tree. Next find the link from $\tilde{\sigma}$ to $\hat{\sigma}$ and denote this link by $l_{\tilde{\sigma}\hat{\sigma}}$. Adding this link to the tree forms a graph that contains paths from all nodes to $\hat{\sigma}$, since the graph contains the path $p_{\sigma_1\tilde{\sigma}} + l_{\tilde{\sigma}\hat{\sigma}}$ for all σ_1. Observe that the tree to $\tilde{\sigma}$ must involve a link from the chosen $\hat{\sigma}$ equilibrium to some other node σ_2, denoted $l_{\hat{\sigma}\sigma_2}$. This link, $l_{\hat{\sigma}\sigma_2}$ is incorporated into a path from σ_1 as follows: $p_{\sigma_1\hat{\sigma}} + l_{\hat{\sigma}\sigma_2} + p_{\sigma_2\tilde{\sigma}} + l_{\tilde{\sigma}\hat{\sigma}}$ (where it can be that $\sigma_1 = \hat{\sigma}$, so the path starts at $\hat{\sigma}$, and/or $\sigma_2 = \tilde{\sigma}$ so the link from $\hat{\sigma}$ in the original tree is directly to $\tilde{\sigma}$). It is clear that by deleting the link $l_{\hat{\sigma}\sigma_2}$ there is still a path from any node to the node $\hat{\sigma}$. If the weight of the link added ($l_{\tilde{\sigma}\hat{\sigma}}$) is lower than the weight of the link deleted (some $l_{\hat{\sigma}\sigma_2}$) then the lowest weight tree to $\hat{\sigma}$ has lower total weight than the lowest weight tree to $\tilde{\sigma}$.

To show that one need only examine links from $\hat{\sigma}$ out of the set Σ_{z_1}, observe the following. If the link $l_{\hat{\sigma}\sigma_2}$ is to some $\sigma_2 \in \Sigma_{z_1}$, then the path $p_{\sigma_2\tilde{\sigma}}$ must include a link from a node in Σ_{z_1}, denoted σ_3 to a node, denoted σ_4, where $\sigma_4 \in \Sigma_{z_2}$, $z_2 \neq z_1$, (where it could be that $\sigma_3 = \sigma_2$). The path $l_{\hat{\sigma}\sigma_2} + p_{\sigma_2\tilde{\sigma}}$ must include a path from $\hat{\sigma}$ to σ_3 that links only nodes in Σ_{z_1}. Now delete the link from σ_3 to $\sigma_4 \in \Sigma_{z_2}$, $l_{\sigma_3\sigma_4}$, and reverse all links on the path from $\hat{\sigma}$ to σ_3. This changes the path $p_{\sigma_1\hat{\sigma}} + l_{\hat{\sigma}\sigma_2} + p_{\sigma_2\tilde{\sigma}} + l_{\tilde{\sigma}\hat{\sigma}}$ into the following paths $p_{\sigma_1\hat{\sigma}}$, $p_{\sigma_3\hat{\sigma}}$, and $p_{\sigma_4\tilde{\sigma}} + l_{\tilde{\sigma}\hat{\sigma}}$. By deleting the link $l_{\sigma_3\sigma_4}$ there is still a path from any node on the path $p_{\sigma_1\hat{\sigma}} + l_{\hat{\sigma}\sigma_2} + p_{\sigma_2\tilde{\sigma}} + l_{\tilde{\sigma}\hat{\sigma}}$ to the node $\hat{\sigma}$. Similarly if the path from any node j includes any node on the path $p_{\sigma_1\hat{\sigma}} + l_{\hat{\sigma}\sigma_2} + p_{\sigma_2\tilde{\sigma}} + l_{\tilde{\sigma}\hat{\sigma}}$ then there continues to be a link to $\hat{\sigma}$. Observe that reversing a link between two nodes in a given set does not change the weight of the link. Thus if the weight of link $l_{\sigma_3\sigma_4}$ is higher than that of link $l_{\tilde{\sigma}\hat{\sigma}}$ then the lowest weight tree to $\hat{\sigma}$ has lower total weight than the lowest weight tree to $\tilde{\sigma}$. To conclude the proof, note that equilibria are symmetric so the path $l_{\sigma_3\sigma_4}$ will have an analogous path from $\hat{\sigma}$. □

To simplify notation, we denote an equilibrium where both types play (x^M, y^A) by (M, A) and analogously for other assimilation equilibria. We denote an equilibrium where type M play (x^M, y^A) and type m play (x^m, y^B) by $(M, A; m, B)$, and analogously for other segregation equilibria. The lowest weight link between each group, denoted ϕ, is detailed in Table 4.5. For now we allow the weight of a link to be a real number, $\phi \in \mathbb{R}$, and address the integer nature of mistakes at the end of the proof. Given Lemmas 4.1 and 4.2, the following statements follow in a straightforward manner from Table 4.5.

M-assimilation has lower stochastic potential than segregation when $\phi_{sM} < \phi_{Ms}$.
Take the lowest weight tree to any segregation strict Nash equilibrium. The lowest weight link from this node to some M-assimilation equilibrium is ϕ_{sM}. It can be seen from Table 4.5 that the weight of any link from M-assimilation to m-assimilation, or to segregation, is at least ϕ_{Ms}. Thus the result follows from Lemmas 4.1 and 4.2.

Segregation has lower stochastic potential than M-assimilation when $\phi_{Ms} < \phi_{sM}$.
Take the lowest weight tree to any M-assimilation equilibrium. The lowest weight link from this node to some segregation equilibrium is ϕ_{Ms}. Any link from segregation to M-assimilation, or to m-assimilation, has weight at least ϕ_{sM}. By Lemma 4.1 and 4.2 the statement holds.

Segregation has lower stochastic potential than m-assimilation when $\phi_{Ms} < \phi_{sM}$.
The lowest weight link from any m-assimilation equilibrium to segregation is ϕ_{Ms}. The result follows as above.

M-assimilation has equal stochastic potential to m-assimilation when $\phi_{sM} < \phi_{sm} < \phi_{Ms}$.
Take the lowest weight tree to any M-assimilation equilibrium. The lowest weight link from this node to some m-assimilation equilibrium is ϕ_{Ms}, since $\phi_{sm} < \phi_{Ms}$. Any link from m-assimilation to M-assimilation, or to segregation, has weight at least ϕ_{Ms}, since $\phi_{sM} < \phi_{Ms}$. By Lemma 4.1 and Lemma 4.2, m-assimilation has weakly lower stochastic potential than M-assimilation. Starting from the lowest weight tree to any m-assimilation equilibrium repeat the process to show M-assimilation has weakly lower stochastic potential than m-assimilation.

M-assimilation has lower stochastic potential than m-assimilation when $\phi_{sM} < \phi_{Ms} < \phi_{sm}$. Take the lowest weight tree to any m-assimilation equilibrium. Suppose this lowest weight tree includes a link from one of the segregation equilibria to (m, A) or (m, B). The minimum weight link from some segregation equilibrium to some m-assimilation equilibrium is ϕ_{sm}. Delete this link and form a link from the same segregation equilibrium, by the lowest weight link, to either (M, A) or (M, B), which will have weight ϕ_{sM}. The new graph has strictly lower total weight. Now form a link from the m-assimilation equilibrium to the same node, either (M, A) or (M, B), which will be weight ϕ_{Ms}, and delete a link from the node either (M, A) or (M, B), which are all of weight at least ϕ_{Ms}. Thus we have a graph from all nodes to (M, A) or (M, B) with strictly lower total weight than the tree to m-assimilation.

Suppose instead the lowest weight tree to any m-assimilation equilibrium has no link from a segregation equilibria to (m, A) or (m, B). Then there must be a link from either (M, A) or (M, B) to either (m, A) or (m, B) in order to have a tree to m-assimilation. Now a link from either (M, A) or (M, B) has the same weight to either (m, A) or (m, B) so we can replace it with a link to the chosen m-assimilation equilibrium while maintaining the tree and without affecting the weight of the tree. Now we delete this link and form the reverse link from the chosen m-assimilation equilibrium to the node (M, A) or (M, B) whose link to m-assimilation was just deleted. The graph now has strictly lower weight and maintains a link to all nodes.

Above we suppose $\phi \in \mathbb{R}$. Since our population is discrete, the minimum weights are in fact the lowest integer greater than or equal to a given ϕ. Given the discrete nature of the population the equilibrium is guaranteed to be unique when $\phi_{Ms} < \phi_{sM}$ if $\phi_{sM} - \phi_{Ms} \geq 1$ and when $\phi_{Ms} > \phi_{sM}$ if $\phi_{Ms} - \phi_{sM} \geq 1$. Since $\phi_{sM} \gtreqless \phi_{Ms}$ implies $n_m \gtreqless n - c/(1 - L)$, $\phi_{sM} - \phi_{Ms} \geq 1$ is equivalent to $n_m \geq n - c/(1 - L) + 1 = \delta^* + 1$, and $\phi_{Ms} - \phi_{sM} \geq 1$ is equivalent to $n_m \leq n - c/(1 - L) - 1 = \delta^* - 1$. M-assimilation has lower stochastic potential than m-assimilation when $\phi_{sm} > \phi_{Ms}$, allowing for discreteness requires $\phi_{sm} - \phi_{Ms} \geq 1$ which implies $n_m \leq c/(1 - L) - 1$. This final inequality is Assumption 4.2.

Table 4.5: *The lowest weight link from a node in one set to a node in a different set, for* $1-\alpha < L$ *and* $\alpha \gtreqless L$

Sets linked	Minimum number of mistakes to transit between sets is min. ϕ satisfying:	Notes	Weight denoted
segregation to M-assimilation	$(1-L)(n_M+\phi)-c \geq (1-L)(n_m-\phi-1)$	e.g. from $(M,A;m,B)$ to (M,A). This involves type m making a mistake of (M,A).	ϕ_{sM}
segregation to m-assimilation	$(1-L)(n_m+\phi)-c \geq (1-L)(n_M-\phi-1)$	e.g. from $(M,A;m,B)$ to (m,B). This involves mistakes of (m,B).	ϕ_{sm}
M-assimilation to segregation	$(1-L)\phi \geq (1-L)(n-\phi-1)-c$	e.g. from (M,A) to $(M,A;m,B)$. This involves mistakes of (m,B).	ϕ_{Ms}
m-assimilation to segregation	$(1-L)\phi \geq (1-L)(n-\phi-1)-c$	e.g. from (m,B) to $(M,A;m,B)$. This involves mistakes of (M,A).	ϕ_{Ms}
m-assimilation to M-assimilation	$(1-L)\phi \geq (1-L)(n-k-1)-c$ and $(1-L)(n_M+\phi')-c \geq (1-L)(n_m-\phi'-1)$	e.g. from (m,A) to (M,A). Type M want to transition after at least ϕ mistakes of (M,A). Once all type M have moved, type m want to transition if at least ϕ' mistakes are made (by type m). The minimum number of mistakes must satisfy both inequalities.	$\max\{\phi_{Ms},\phi_{sM}\}$
M-assimilation to m-assimilation	$(1-L)\phi \geq (1-L)(n-k-1)-c$ and $(1-L)(n_m+\phi')-c \geq (1-L)(n_M-\phi'-1)$	e.g. from (M,A) to (m,A). Type m want to transition after at least ϕ mistakes of (m,A). Once all type m have moved, type M want to transition if at least ϕ' mistakes are made (by type M).	$\max\{\phi_{Ms},\phi_{sm}\}$

Proof for the case when culture is relatively unimportant

The proof here is given for the parameter range $1-\alpha > L$.

Lemma 4.1 becomes Lemma 4.3:

Lemma 4.3. *Partition the strict Nash equilibria into four sets: segregation; multiculturalism; M-assimilation; and m-assimilation. Such a set is denoted Σ_{z_1} and $z_1 \in \{s, u, M, m\}$ respectively indexes the set. All equilibria in a set, $\sigma \in \Sigma_{z_1}$, have the same stochastic potential.*

Lemma 4.2 continues to apply. Notation is simplified in the same way as the proof for $1-\alpha < L$. Table 4.6 finds the lowest weight link between sets i.e. from some node in set Σ_{z_1} to some node in another set Σ_{z_2}.

Multiculturalism has lower stochastic potential than M-assimilation when $\phi_{Mu} < \phi_{uM}$.
Take the lowest weight tree to an M-assimilation equilibrium. The lowest weight link from this node to some multiculturalism equilibrium is ϕ_{Mu}. It can be seen from Table 4.6 that the weight of any link from multiculturalism to M-assimilation, m-assimilation or segregation is at least the minimum of ϕ_{uM} and $\frac{n_m-1}{2} + \frac{1-\alpha-L}{2(1-L)}n_M$. Observe $\phi_{uM} + \phi_{Mu} = n_m - 1$ and so $\phi_{Mu} < \frac{n_m-1}{2}$. Thus the result follows from Lemmas 2 and 3.

Multiculturalism has lower stochastic potential than m-assimilation when $\phi_{Mu} < \phi_{uM}$.
The transit from m-assimilation to multiculturalism also has weight ϕ_{Mu} so the argument follows as above.

M-assimilation has lower stochastic potential than multiculturalism when $\phi_{uM} < \phi_{Mu}$.
Observe that ϕ_{uM} is the lowest weight transition from multiculturalism to M-assimilation when $\phi_{uM} < \phi_{Mu}$ since then $\phi_{uM} < \frac{n_m-1}{2}$. From Table 4.6, any transition from M-assimilation to multiculturalism, segregation or m-assimilation has weight at least ϕ_{Mu} (since $\phi_{Ms} > \phi_{Mu}$). The result follows from Lemmas 2 and 3.

M-assimilation has lower stochastic potential than segregation when $\phi_{uM} < \phi_{Mu}$.
The transit from segregation to M-assimilation has lower weight than ϕ_{uM} and so the result follows from above.

Multiculturalism has lower stochastic potential than segregation when $\phi_{Mu} < \phi_{uM}$.
The inequality $\phi_{Mu} < \phi_{uM}$ implies $\frac{n-1}{2} - \frac{c}{2\alpha} < \frac{n_m-n_M-1}{2} + \frac{c}{2\alpha}$ which implies $n_M < \frac{c}{\alpha}$. Thus the lowest weight link from segregation to some multiculturalism equilibrium requires

Table 4.6: *The lowest weight link from a node in one set to a node in a different set, for* $1-\alpha > L$ *and* $\alpha \underset{<}{\overset{>}{=}} L$

Sets linked	Minimum number of mistakes to transit between sets is min. ϕ satisfying:	Notes	Weight denoted
M-assimilation to multiculturalism	$(1-L)\phi+(1-\alpha-L)(n-\phi-1) \geq$ $(1-L)(n-\phi-1)+(1-\alpha-L)\phi-c$, which implies $\phi \geq \frac{n-1}{2}-\frac{c}{2\alpha}$	e.g. from (M, A) to (MA, mA).	ϕ_{Mu}
m-assimilation to multiculturalism	As above, $\phi \geq \frac{n-1}{2}-\frac{c}{2\alpha}$	e.g. from (m, A) to (MA, mA).	ϕ_{Mu}
M-assimilation to segregation	$(1-L)\phi \geq (1-L)(n-\phi-1)-c$, which implies $\phi \geq \frac{n-1}{2}-\frac{c}{2(1-L)}$	e.g. from (M, A) to (MA, mB).	ϕ_{Ms}
m-assimilation to segregation	As above, $\phi \geq \frac{n-1}{2}-\frac{c}{2(1-L)}$	e.g. from (m, A) to (MB, mA).	ϕ_{Ms}
multiculturalism to M-assimilation	$(1-L)(n_M+\phi)+(1-\alpha-L)(n_m-\phi-1)-c \geq$ $(1-L)(n_m-\phi-1)+(1-\alpha-L)(n_M+\phi)$, which implies $\phi \geq \frac{n_m-n_M-1}{2}+\frac{c}{2\alpha}$; or $\phi \geq \frac{n_M-1}{2}+\frac{1-\alpha-L}{2(1-L)}n_m$	e.g. from (MA, mA) to (M, A). Alternatively examine e.g. (MA, mA) to (M, B).	ϕ_{uM}
multiculturalism to m-assimilation	$(1-L)(n_m+\phi)+(1-\alpha-L)(n_M-\phi-1)-c \geq$ $(1-L)(n_M-\phi-1)+(1-\alpha-L)(n_m+\phi)$, which implies $\phi \geq \frac{n_M-n_m-1}{2}+\frac{c}{2\alpha}$; or $\phi \geq \frac{n_m-1}{2}+\frac{1-\alpha-L}{2(1-L)}n_M$	e.g. from (MA, mA) to (m, A). Alternatively examine e.g. (MA, mA) to (M, B).	ϕ_{um}
segregation to M-assimilation	$(1-L)(n_M+\phi)-c \geq (1-L)(n_m-\phi-1)$, which implies $\phi \geq \frac{n_m-n_M-1}{2}+\frac{c}{2(1-L)}$	e.g. from (MA, mB) to (M, A).	ϕ_{sM}
segregation to m-assimilation	$(1-L)(n_m+\phi)-c \geq (1-L)(n_M-\phi-1)$, which implies $\phi \geq \frac{n_M-n_m-1}{2}+\frac{c}{2(1-L)}$	e.g. from (MA, mB) to (m, B).	ϕ_{sm}

Table 4.6 continued

M-assimilation to m-assimilation	$(1-L)\phi+(1-\alpha-L)(n-\phi-1)\geq$ $(1-L)(n-\phi-1)+(1-\alpha-L)\phi-c$; and $(1-L)(n_m+\phi')+(1-\alpha-L)(n_M-\phi'-1)-c\geq$ $(1-L)(n_M-\phi'-1)+(1-\alpha-L)(n_m+\phi')$. $(1-L)\phi\geq(1-L)(n-\phi-1)-c$; and $(1-L)(n_m+\phi')-c\geq(1-L)(n_M-\phi'-1)$.	e.g. from (M,A) to (m,A). Require ϕ_{Mu} mistakes for m to switch first. Then M will switch if ϕ_{um} mistakes are made. Alternatively examine e.g. (M,A) to (m,B). Require ϕ_{Ms} mistakes for m to switch first. Then M switch if ϕ_{sm} errors are made.	This is minimum of $\max\{\phi_{Mu},\phi_{um}\}$ and $\max\{\phi_{Ms},\phi_{sm}\}$.
m-assimilation to M-assimilation	$(1-L)\phi+(1-\alpha-L)(n-\phi-1)\geq$ $(1-L)(n-\phi-1)+(1-\alpha-L)\phi-c$; and $(1-L)(n_M+\phi')+(1-\alpha-L)(n_m-\phi'-1)-c\geq$ $(1-L)(n_m-\phi'-1)+(1-\alpha-L)(n_M+\phi')$. $(1-L)\phi\geq(1-L)(n-\phi-1)-c$; and $(1-L)(n_M+\phi')-c\geq(1-L)(n_m-\phi'-1)$.	e.g. from (m,A) to (M,A). ϕ_{Mu} mistakes required for M to switch first. Then m will switch if ϕ_{uM} errors are made. Alternatively examine e.g. (m,A) to (M,B). ϕ_{Ms} mistakes required for M to switch first. Then m will switch if ϕ_{sM} errors are made.	This is minimum of $\max\{\phi_{Mu},\phi_{uM}\}$ and $\max\{\phi_{Ms},\phi_{sM}\}$.
segregation to multiculturalism	$(1-L)\phi+(1-\alpha-L)n_M\geq(1-L)(n_m-\phi-1)$, which implies $\phi\geq\frac{n_m-1}{2}-\frac{1-\alpha-L}{2(1-L)}n_M$; and $(1-L)\phi+(1-\alpha-L)n_M\geq(1-L)n_M+(1-\alpha-L)\phi-c$ which implies $\phi\geq n_M-\frac{c}{\alpha}$. $(1-L)\phi+(1-\alpha-L)n_m\geq(1-L)(n_M-\phi-1)$, which implies $\phi\geq\frac{n_M-1}{2}-\frac{1-\alpha-L}{2(1-L)}n_m$	e.g. from (MA,mB) to (MA,mA). The second inequality ensures m do not want to play (M,A). Alternatively examine e.g. from (MA,mB) to (MB,mB). Observe that M do not want to play (m,B).	
multiculturalism to segregation	$(1-L)\phi\geq(1-L)(n_m-\phi-1)+(1-\alpha-L)n_M$, which implies $\phi\geq\frac{n_m-1}{2}+\frac{1-\alpha-L}{2(1-L)}n_M$; and $(1-L)\phi\geq(1-L)n_M+(1-\alpha-L)(n_m-k-1)-c$, which implies $\phi\geq\frac{1-L}{2(1-L)-\alpha}n_M+\frac{1-\alpha-L}{2(1-L)-\alpha}(n_m-1)-\frac{c}{2(1-L)-\alpha}$. $(1-L)\phi\geq(1-L)(n_M-\phi-1)+(1-\alpha-L)n_m$, which implies $\phi\geq\frac{n_M-1}{2}+\frac{1-\alpha-L}{1-L}n_m$; and $(1-L)\phi\geq(1-L)n_m+(1-\alpha-L)(n_M-k-1)$, which implies $\phi\geq\frac{1-L}{2(1-L)-\alpha}n_m+\frac{1-\alpha-L}{2(1-L)-\alpha}(n_M-1)$.	e.g. from (MA,mA) to (MA,mB). The second inequality ensures m do not want to play (M,A). Alternatively examine e.g. from (MA,mA) to (MB,mA). The second inequality ensures M do not want to play (m,A).	

Notes: *When $\alpha > L$ for $\phi_{Mu} < \phi_{uM}$, the weight for a link from segregation to multiculturalism is lower than specified above. This does not change the results. Multiculturalism to segregation transits continue to require at least $\frac{n_m-1}{2}+\frac{1-\alpha-L}{2(1-L)}n_M$ mistakes.*

$\frac{n_m-1}{2} - \frac{1-\alpha-L}{2(1-L)}n_M$ mistakes. It can be seen from Table 4.6 that any link out of multiculturalism is of weight at least the minimum of ϕ_{uM} and $\frac{n_m-1}{2} + \frac{1-\alpha-L}{2(1-L)}n_M$.

M-assimilation has lower stochastic potential than m-assimilation when $\phi_{uM} < \phi_{Mu}$ and $\phi_{sm} > \phi_{Ms}$.

Observe $\phi_{um} > \phi_{sm} > \phi_{Ms} > \phi_{Mu} > \phi_{uM}$. Since $\phi_{Mu} > \phi_{uM}$ then $n_M > c/\alpha$ and so $\phi_{Ms} > \phi_{sM}$ but we know $\phi_{Mu} < \phi_{Ms}$ and so ϕ_{Mu} is the minimum weight transition from m-assimilation to M-assimilation.

Without loss of generality consider the tree to (m, A). Suppose the tree has a link from (M, A) to (m, A), which is of weight ϕ_{um}, or from (M, A) to (m, B), which is of weight ϕ_{sm}. Form a link from (m, A) to (M, A), of weight ϕ_{Mu}, and delete the link from (M, A) that goes to either (m, A) or (m, B), to form a tree to (M, A) with strictly lower total weight. Suppose the new tree has a link from (M, B) to (m, A), which is of weight ϕ_{sm}, or from (M, B) to (m, B), which is of weight ϕ_{um}. Delete this link and form (m, A) to (M, B), of weight ϕ_{Ms}, to form a tree to (M, B) with strictly lower total weight.

Suppose instead the tree to (m, A) has no direct links from any M-assimilation node to any m-assimilation node. Then there must be a direct link from segregation and/or multiculturalism to some m-assimilation node. Suppose the direct link is from $(MA; mA)$ to either (m, A) or (m, B). This link is of weight at least ϕ_{um} or $\frac{n_m-1}{2} + \frac{1-\alpha-L}{2(1-L)}n_M$. Delete such a link and from the same node send the link instead to (M, A) of weight $\phi_{uM} < \frac{n_m-1}{2}$.[65] Now form a link from (m, A) to (M, A) of weight ϕ_{Mu} and delete any link from (M, A) which are all of weight at least ϕ_{Mu}. Suppose the direct link is from $(MA; mB)$ to either (m, A) or (m, B). This link is of weight at least ϕ_{sm}. Delete such a link and from the same node send the link instead to (M, A) of weight ϕ_{sM}. Now form a link from (m, A) to (M, A) of weight ϕ_{Mu} and delete any link from (M, A) which are all of weight at least ϕ_{Mu}.

Finally suppose none of the other links are present in the tree to (m, A). Then there must be a link from $(mB; mA)$ or $(MB; mB)$ to either (m, A) or (m, B) or both, and no other direct links. Transfer all such links to (M, B) which, as above, produces a graph of strictly lower total weight. Every node now has a path to (M, B) apart from (m, A) (and possibly (m, B) if it links directly to (m, A)). So finally form a link from (m, A) to $(MA; mA)$ of weight ϕ_{Mu} and delete a link from (M, B) which is at least ϕ_{Mu}.

[65] Observe that the difference between ϕ_{uM} and ϕ_{um} is the same as the difference between ϕ_{sM} and ϕ_{sm}.

For $\phi \in \mathbb{R}$, under any parameters there is a unique set, Σ_{z_1}, with lowest stochastic potential. However, in a finite population ϕ is discrete. Because of this discreteness we can get intervals where the set is not unique. To avoid issues associated with discreteness we examine the parameter range $|\phi_{uM} - \phi_{Mu}| \geq 1$. Then $\phi_{uM} - \phi_{Mu} \geq 1$ implies $n_m \geq n - c/\alpha + 1$, and $\phi_{Mu} - \phi_{uM} \geq 1$ implies $n_m \leq n - c/\alpha - 1$. To avoid the same issues associated with discreteness close to parameter values $1-L = \alpha$ we also require $\frac{n_m - 1}{2} + \frac{1-\alpha-L}{2(1-L)} n_M - \min\{\phi_{Mu}, \phi_{uM}\} \geq 1$. Since $\phi_{uM} + \phi_{Mu} = n_m - 1$ combined with the assumption $\phi_{uM} - \phi_{Mu} \geq 1$ implies $\min\{\phi_{Mu}, \phi_{uM}\} \leq \frac{n_m - 1}{2} - \frac{1}{2}$, this implies $1 - L \geq \alpha + \frac{1}{n_M - 1}$.[66] Finally Assumption 2 continues to ensure $\phi_{sm} - \phi_{Ms} \geq 1$.

Proof that the Markov process converges almost surely to a strict Nash equilibrium

We show there exists no recurrence class other than the strict Nash equilibria. To do so it suffices to show that from any state there are a finite number of positive probability events which lead to a strict Nash equilibrium.

Suppose $n_1 \leq n$ players play (M, A), $n_2 \leq n$ players play (m, A), $n_3 \leq n$ players play (M, B) and $n_4 \leq n$ players play (m, B), where either type can be playing either action. With positive probability a given type M is selected to update his strategy. Suppose the type M does (at least weakly) best by playing (M, A), and chooses (M, A). Suppose next a different individual of type M is selected to update his strategy; again this occurs with positive probability. This individual must strictly prefer (M, A) since the payoff from playing (M, A) relative to other actions has strictly increased and it was weakly optimal for a type M in the previous period. Let this continue such that all type M are selected and no type m. All type M now play (M, A). Next a given type m is selected to update his strategy with positive probability. His optimal strategy may be either (M, A), (m, A), or (m, B). He chooses one of these actions. Let this continue until all type m have been selected to update their strategy. The action chosen by the first type m agent selected to update his strategy is strictly best for all type m that follow. The process arrives at a strict Nash equilibrium since all type M play (M, A) and all type m play one of either (M, A), (m, A), or (m, B). Thus we have either assimilation by type m, segregation, or multiculturalism. The same argument holds if instead the initial type M selected to update his strategy selects (M, B). If instead the initial type M selected to update his strategy selects (m, A), then all type M or m that follow must do strictly best by playing (m, A). The result follows. Similarly if instead the initial type M selected does best by playing (m, B).

[66] This also guarantees that the alternative link from segregation to mulitculturalism has weight at least one unit higher than the alternative link from multiculturalism to segregation: $\frac{n_m - 1}{2} + \frac{1-\alpha-L}{2(1-L)} n_M - [\frac{n_m - 1}{2} - \frac{1-\alpha-L}{2(1-L)} n_M] \geq 1$.

4.8 Additional empirical results

Table 4.7: *Number of observed cohorts by nationality*

	Minimum cohort size of		
	(1) 30	(2) 20	(3) 40
Germany	303	455	234
Italy	291	404	225
Poland	196	276	149
Canada	174	259	116
Britain	167	248	127
Russia	151	219	122
Sweden	139	242	96
Ireland	133	185	109
Mexico	59	102	38
Other	133	227	91
Total	1746	2617	1307

Notes. *Number of cohorts we observe, broken down by nationality. Cohorts are defined by nationality, county, tenure in the US grouped to the nearest 10 years, and year of arrival in 10 year bands. Column (1) shows the number of cohorts when we require cohorts to contain a minimum of 30 people. Columns (2) and (3) show how these results change when thresholds of 20 or 40 are used instead. Germany includes Germans and Austrians. Sweden includes Sweden, Denmark, and Norway. Other is composed (in order of size, at cohort size 30) of Romania, Japan, Netherlands, Greece, Finland, Portugal, West Indies, France, China, Spain, Belgium, Cuba, South America. These nationalities are treated as seperate from each other in our analysis, and are aggregated here only for brevity.*

Table 4.8: *Comparing the threshold effect in share speaking English, using different sample restrictions or weights*

Dependent Variable: Proportion of people in the cohort that speak English
Standard Errors in Parentheses
Cohorts defined by Nationality, County, Grouped Year of Arrival, Grouped Tenure

	Minimum cohort size			
	(1) 30	(2) 20	(3) 40	(4) Weighted regression
Optimal threshold ($\beta_1 \vert \tau = \tau^*$)	-.509***	-.348***	-.475***	-.510***
	(.025)	(.016)	(.024)	(.035)
Constant (β_0)	.885***	.885***	.890***	.886***
	(.004)	(.004)	(.004)	(.004)
Optimal Threshold Level (τ^*)	.31	.27	.30	.31
F-statistic	422	457	394	208
1% critical value for F-statistic	6.6	6.6	6.6	6.6
Cohort Fixed Effects	No	No	No	No
Slopes in Nationality Share	No	No	No	No
Observations	1272	1925	955	1272

Notes. **** denotes significance at 0.1%, ** at 1%, and * at 5% level, when treated as a standalone regression. The outcome measures the proportion of the cohort that speaks English. Cohorts with English, Canadian, and Irish nationalities are excluded from the sample, since English is likely to be their mother tongue. Cohort fixed effects are composed of nationality, arrival year (grouped), and tenure (grouped) fixed effects.*
All columns are regressions of share speaking English on a sequence of nationality share thresholds. We vary the threshold between .20 and .40, at intervals of .01. We provide the results for the threshold among these which produced a regression with the highest F-statistic when tested against the null of no threshold. The value for this threshold, the estimated F-statistic, and the 1% critical value for this statistic (Andrews, 1993) are provided at the bottom of the table.
Column (1) requires a minimum cohort size of 30, as in our main specifications. Columns (2) and (3) use minimum sizes of 20 and 40 respectively. Column (4) retains the minimum cohort size of 30, but first divides cohorts in the unrestricted sample into deciles based on nationality share, and then weights observations in each of these deciles in the analysis sample inversely proportionally to the size of that decile in the analysis sample.

Table 4.9: *Comparing the threshold effect in share in-married, using different sample restrictions or weights*

Dependent Variable: Proportion of married people in the cohort that are married to someone of the same nationality
Standard Errors in Parentheses
Cohorts defined by Nationality, County, Grouped Year of Arrival, Grouped Tenure

	Minimum cohort size			
	(1) 30	(2) 20	(3) 40	(4) Weighted regression
Optimal threshold ($\beta_1 \vert \tau = \tau^*$)	.151***	.094**	.128**	.142***
	(.042)	(.030)	(.048)	(.030)
Constant (β_0)	.606***	.598***	.613***	.609***
	(.006)	(.005)	(.006)	(.006)
Optimal Threshold Level (τ^*)	.35	.35	.35	.33
F-statistic	13.0	9.7	7.3	22.4
1% critical value for F-statistic	6.6	6.6	6.6	6.6
Cohort Fixed Effects	No	No	No	No
Slopes in Nationality Share	No	No	No	No
Observations	1746	2617	1307	1746

Notes. **** denotes significance at 0.1%, ** at 1%, and * at 5% level, when treated as a standalone regression. The outcome measures the married proportion of the cohort that is 'in-married' i.e. married to someone of the same nationality. Cohort fixed effects are composed of nationality, arrival year (grouped), and tenure (grouped) fixed effects.*
All columns are regressions of share in-married on a sequence of nationality share thresholds. We vary the threshold between .20 and .40, at intervals of .01. We provide the results for the threshold among these which produced a regression with the highest F-statistic when tested against the null of no threshold. The value for this threshold, the estimated F-statistic, and the 1% critical value for this statistic (Andrews, 1993) are provided at the bottom of the table.
Column (1) requires a minimum cohort size of 30, as in our main specifications. Columns (2) and (3) use minimum sizes of 20 and 40 respectively. Column (4) retains the minimum cohort size of 30, but first divides cohorts in the unrestricted sample into deciles based on nationality share, and then weights observations in each of these deciles in the analysis sample inversely proportionally to the size of that decile in the analysis sample.

Chapter 5

The Dynamic Effects of Tax Audits[1]

5.1 Introduction

Central to the efficient functioning of a modern tax authority is the ability to assess and collect tax revenue owed by taxpayers in a timely and cost-effective manner. One tool used by many tax authorities to help achieve this aim is taxpayer audits. Audits have a mechanical benefit in terms of the unpaid revenue they identify and recover. Historically, this is primarily what tax authorities such as the Internal Revenue Service (IRS) have focused on when selecting tax returns for further scrutiny (see Bloomquist, 2013). However, since most taxpayers both pay taxes over many years and interact with other taxpayers, it is important to understand the extent of any indirect (behavioural) effects of audits that arise due to updated information about the likelihood, effectiveness or cost of audits. These indirect effects take two forms: dynamic effects and spillover effects. Dynamic effects are changes in the future behaviour of audited taxpayers. Spillover effects are changes in the behaviour of other taxpayers who know an audited taxpayer.[2] Measuring these wider impacts of audits is crucial for determining audit strategy including the optimal extent of enforcement.

[1]This chapter is co-authored with William Elming and Jonathan Shaw. The authors thank Richard Blundell, Tracey Bowler, Monica Costa Dias, Costas Meghir, Imran Rasul and seminar participants at the Tax Systems Conference in October 2014, Royal Economic Society Annual Conference in April 2015, the International Institute of Public Finance Annual Congress in August 2015, the Louis-André Gérard-Varet Conference in June 2016 and the European Economic Association Annual Conference in August 2016 for helpful comments. This work contains statistical data from HMRC which is Crown Copyright. The research datasets used may not exactly reproduce HMRC aggregates. The use of HMRC statistical data in this work does not imply the endorsement of HMRC in relation to the interpretation or analysis of the information.

[2]There is also the possibility of a general deterrence effect of audits that extends beyond the group of taxpayers who know someone who has been audited.

In this chpater we study the dynamic effects of audits on individual income tax payers in the UK. Our contribution is twofold: we demonstrate how the response to audits evolves over time and show how it varies across different income sources subject to varying degrees of third-party reporting. Typically, this is challenging to do robustly because most audits are targeted towards taxpayers believed to be noncompliant. We address this by exploiting a random audit programme run by the UK tax authority (HM Revenue and Customs, HMRC) that focuses on income tax self-assessment taxpayers. As a control group, we use individuals who could have been selected for a random audit but weren't. The parameter we estimate is an intention to treat (ITT) parameter because some of those assigned to audit do not end up being audited (e.g. because their case is not worked before the required deadline). Thus, our results should be interpreted as the effect of being selected for random audit relative to a baseline of not being selected for random audit but facing the normal policy environment, which includes a small chance of a targeted audit.

Our main results are as follows. First, we show that audits lead to a modest increase in survival in self assessment for two years after the audit that peaks at 1.1 percentage points and is consistent with individuals being more diligent in filing their tax returns while audit investigations are ongoing. Second, we uncover little in the way of an impact on reported total taxable income and total tax, except for a statistically significant increase in total taxable income of £795 (3.0 per cent of taxable income) in the second year after audit. Nevertheless, while not significant, some of the point estimates remain substantial. Third, we demonstrate that there are marked differences in the response across different reported income components. Components not subject to third-party reporting tend to experience a significant increase after audit. For example, self-employment earnings and property income peak 5.6 per cent and 7.9 per cent higher respectively, but these increases largely fall away after four years, suggesting individuals are reverting to their old behaviour. In contrast, audits do not seem to have an impact on income components that are subject to third-party reporting, such as employment earnings, suggesting compliance levels here are already high. Pension income does record a significant and sustained increase after audit despite extensive third-party reporting, a pattern that might be due to the correction of previously undetected errors.

This chapter contributes to a small but growing literature on the dynamic effect of audits. Kleven et al. (2011) analyse the effect of a tax enforcement field experiment in Denmark. Half of a sample of income tax filers were randomly audited based on their 2006 return, while the other half were not. The following year, before 2007 returns had been filed, letters threatening an audit were randomly assigned to a subset of tax filers in both groups. Declarations in 2007 returns were then followed up. Their results suggest that the overall deterrence effect of audits is positive but modest, corresponding to an increase in reported total net income of about 2,500 Danish kroner, or around 1% of income. This effect is driven entirely by self-reported income; there is no impact on third-party reported income. They do not follow taxpayers beyond the first year.

Gemmell and Ratto (2012) investigate behavioural responses to taxpayer audits in the UK in the year 2000 using a much more limited version of the random audit data we use. They find no impact on overall tax declared. They also distinguish between taxpayers found to be non-compliant and those found to be compliant, arguing that the former are likely to increase their subsequent compliance while the latter could reduce their compliance. However, this distinction between compliant and non-compliant taxpayers is endogenous, making it hard to interpret the comparison with an unconditionally randomly selected control group as causal.

In work conducted concurrently with this study, DeBacker et al. (2015) investigate the impact of random audits conducted by the US Internal Revenue Service (IRS) between 2006 and 2009 on subsequent taxpaying behaviour. They find that audits increase reported taxable income by over \$1,000 per year, equivalent to 2.7% of average income. The effect is only 0.4% for wage income compared to 7.5% for self-employment income but the former lasts over time while the latter doesn't. Over a five-year horizon, the static revenue gain from the audit understates the total revenue gain by more than half. Earlier studies for the US include Long and Schwartz (1987), Erard (1992) and Tauchen et al. (1993).

We extend this work by considering a longer horizon (up to eight years after audit) and by providing a detailed breakdown for different income sources (see DeBacker et al., 2015). For example, we find somewhat larger initial results than Kleven et al. (2011) but show that these effects are largely transitory.

Using laboratory experiments, Maciejovsky et al. (2007) and Kastlunger et al. (2009) find that compliance decreases immediately after an audit, suggestive of a 'bomb-crater effect'.[3] These studies are based on experiments using students. Choo et al. (2013), whose results are based on an experiment with taxpayers rather than students, do not find any evidence of a bomb crater effect. Bruttel and Friehe (2014) explore in an experimental context whether past enforcement regimes have an impact on current income declarations. We build on this work by demonstrating that a bomb crater effect does not exist for taxpayers as a whole, based on a set of real (rather than experimental) compliance decisions.

[3]The 'bomb-crater effect' refers to the idea that individuals might perceive the risk of being audited to fall immediately after an audit. The name originates from preference of World War I soldiers to hide out in bomb craters, believing that it was unlikely that a bomb would strike exactly the same place again (see Mittone, 2006). A competing explanation for the decline in reported tax following an audit is the mechanism of loss repair: experiencing an audit may make taxpayers "want to evade more in the future in an attempt to 'get back' at the tax agency" (see Andreoni et al., 1998, p. 844).

The remainder of the chapter is organised as follows. Section 5.2 outlines the policy context and data sources, while Section 5.3 sets out the method we use. The results are presented in Section 5.4. Section 5.5 concludes.

5.2 Background

5.2.1 Policy context

In this chapter, we focus on individuals who file an income tax self assessment return in the UK. Self assessment returns are used to collect three main taxes: income tax, National Insurance contributions and capital gains tax. Income tax is the largest of all UK taxes, contributing 25.0 per cent of total government receipts in 2014-15. Most sources of income are subject to income tax, including earnings, retirement pensions, income from property, interest on deposits in bank accounts, dividends and some welfare benefits. Income tax is levied on an individual basis and operates through a system of allowances and bands. Each individual has a personal allowance, which is deducted from total income. The remainder—taxable income—is then subject to a progressive schedule of tax rates. National Insurance contributions comprised 16.8 per cent of government receipts in 2014-15. It operates through a similar system of allowances and bands as income tax but is levied only on employment earnings and self employment profits. Capital gains tax made up 0.9 per cent of government receipts in 2014-15. It is levied on the increase in the value of assets when ownership changes. The structure of the tax has been reformed a number of times; for much of our period of interest the tax rate individuals faced depended on the type of asset, the holding period and their marginal income tax rate.

Unlike the US, not all taxpayers have to file a tax return in the UK: for the tax year 2014-15, around 10 million individuals filed a return (around 30 million individuals were liable for income tax). Individuals required to submit a tax return tend to be individuals with forms of income not subject to withholding or for whom the tax system struggles to calculate and withhold the right amount of tax. It includes self-employed individuals, those with incomes over £100,000, company directors, landlords and many pensioners.

Since incomes covered by self assessment tend to be harder to verify, there is a significant risk of non-compliance. As a result, HM Revenue and Customs (HMRC, the UK tax authority) carries out audits each year to deter non-compliance and recover lost revenue. HMRC runs two types of audit. Targeted audits are based on perceived risks of non-compliance. Random audits are used to ensure that all self-assessment taxpayers face a positive probability of being audited, as well as to collect information about the scale of non-compliance and predictors of non-compliance.

The timeline for the audit process is as follows. The tax year ends on 5 April. Shortly after the end of the tax year, HMRC issues a 'notice to file' to taxpayers who they believe need to submit a tax return. This is based on information that HMRC held shortly before the end of the tax year. Random audit cases are provisionally selected from the population of individuals issued with a notice to file. The deadline by which taxpayers must submit their tax return is 31 January the following calendar year (e.g. 31 January 2010 for the 2008-09 tax year). Once returns have been submitted, HMRC deselects some random audit cases (e.g. due to an audit already being open or having recently concluded). At the same time, targeted audits are selected on the basis of the information provided in self-assessment returns and other intelligence. Random audits are selected before targeted audits, and individuals cannot be selected for a targeted audit in the same tax year as a random audit. The list of taxpayers to be audited is passed on to local compliance teams who carry out the audits. Audits must be opened within a year of the date when the return was filed.[4] Taxpayers subject to an audit are informed when it is opened but they are not told whether it is a random or targeted audit. A small number of taxpayers on the list passed on to local compliance teams end up not being audited, typically due to resource constraints. Because deselections and non-worked cases are not random, we cannot simply compare individuals who actually get audited under the random audit program with non-audited individuals. We discuss the implications for what parameter we are able to recover in Subsection 5.3.1.

5.2.2 Data

We exploit data on income tax self assessment random audits together with information on income tax returns. This combines a number of different HMRC datasets, linked together on the basis of encrypted taxpayer reference number and tax year.

Audit records for tax years 1998-99 to 2008-09 come from CQI (Compliance Quality Initiative), an HMRC operational dataset that records audits made into income tax self assessment and corporation tax self assessment returns. It does not include audits by the HMRC's Large Business Service, Special Investigations or Employer Compliance Reviews. It includes operational information about the audits, such as start and end dates and audit outcomes (whether noncompliance was found, and the size of any correction, penalties and interest).

We track individuals before and after the audit using information from tax returns for the years 1998-99 to 2011-12. This comes from two datasets: SA302 and Valid View. The SA302 dataset contains information that is sent out to taxpayers summarising their income and tax liability (the SA302 tax calculation form). It is derived from self assessment returns, which

[4]For tax returns relating to 2006-07 or earlier, audits had to be opened within a year from the 31 January filing deadline for returns filed on time.

have been put through a tax calculation process. It contains information about total income and tax liability as well as a breakdown into different income sources (employment earnings, self employment profits, pensions, property income etc). For all of these variables, we uprate to 2012 using the Consumer Prices Index (CPI) and Winsorise at the 99.95th percentile to avoid outliers having an undue impact on the results. We supplement these variables with information from Valid View, a dataset that draws directly from individuals' tax returns. For our purposes, Valid View provides demographics and filing information (e.g. filing date). Note that we cannot identify actual compliance behaviour subsequent to the audit: the number of random audit taxpayers that are re-audited is far too small for it to be possible to focus just on them.

5.2.3 Audit descriptives

Table 5.1 shows the average number of individuals per year who face an income tax random audit over the period 1998-99 to 2008-09.[5] HMRC selected an average of 3,299 cases for random audits and 178,829 cases for targeted audits per year. The corresponding probabilities of being selected for an audit are 0.04 per cent (four in 10,000) for random audits and 2.1 per cent for targeted audits.

Table 5.1: *Average number of cases selected for income tax audits*

	Number per year	Audit probability
Random	3,299	0.0004
Targeted	178,829	0.0210

Notes: *Annual averages for tax years 1998-99 to 2008-09. Includes all individuals who face an audit.*
Source: *Authors' calculations based on HMRC administrative datasets.*

Table 5.2 provides some summary statistics for lags in, and durations of, the audit process among random audit cases. As described above, up to and including the 2006-07 return, HMRC had to begin an audit within 12 months of the 31 January filing deadline; since then, HMRC has had to begin an audit within 12 months of the filing date. The average lag between when the tax return was filed and when the random audit was started is 8.9 months, but 10 per cent have a lag of 14 months or more. The average duration of audits is 5.3 months, but 10 per cent experience a duration of 13 months or more. Taken together, this means that the average time between when a return is filed and when the audit is concluded is 14.3 months but there is a long tail for whom the experience is much more drawn out: for 10 per cent it is almost two years or more. This means that individuals will generally have filed at least one subsequent tax return before the outcome of the audit is clear and some will have filed two tax returns. This will be relevant for interpreting the results in Section 5.4.

[5]There are also a small number of partnerships and trusts that are audited, but we exclude these from our analysis.

Table 5.2: *Random audit lags and durations*

	Mean	Std. dev.	Median	75th	90th
Lag to audit start (months)	8.9	4.0	9	11	14
Audit duration (months)	5.3	6.6	3	7	13
Total time to audit end (months)	14.3	7.3	13	17	23

Notes: *Annual averages for tax years 1998-99 to 2008-09. Includes all individuals with a completed random audit.*
Source: *Authors' calculations based on HMRC administrative datasets.*

Table 5.3 summarises the outcomes of random audits on individuals. More than half of all returns are found to be correct, 11 per cent are found to be incorrect but with no underpayment of tax and 36 per cent are 'non-compliant' and have a tax underpayment. Among the non-compliant, the average additional tax owed is £1,912 (across all taxpayers, the figure is £683). However, the distribution is heavily skewed: 65 per cent of non-compliant individuals owe additional tax of £1,000 or less, whilst a small fraction (three per cent) owe more than £10,000.

Table 5.3: *Random audit outcomes*

	Mean	Std. dev.
Proportion of audited returns deemed		
Correct	0.533	0.499
Incorrect but no underpayment	0.111	0.314
Incorrect with underpayment (non-compliant)	0.357	0.479
Mean additional tax if non-compliant (£)	1,912	8,177
Distribution of additional tax if non-compliant		
Share £1-100	0.149	0.356
Share £101-1,000	0.498	0.500
Share £1,001-10,000	0.324	0.468
Share £10,001+	0.029	0.169

Notes: *Annual averages for tax years 1998-99 to 2008-09. Includes all individuals with a completed random audit.*
Source: *Authors' calculations based on HMRC administrative datasets.*

5.3 Method

5.3.1 Parameters of interest

For each taxpayer in self assessment, we observe the amount of tax paid in year $t + h$, Y_{t+h}. If $D(t)$ is the event that a taxpayer is audited in year t, then we can define the tax paid h years after audit as $Y_{t+h}|D(t)$. The question we wish to answer is: what effect do audits have on individuals subsequent tax reporting decisions? A second question is: how does this vary across income components subject to differing degrees of third-party reporting? More formally, we want to know the average effect of an audit on the expected tax liability h years after audit:

$$\beta_{th}^{ATE} = \mathbb{E}\left[Y_{t+h}|D(t) = 1\right] - \mathbb{E}\left[Y_{t+h}|D(t) = 0\right] \tag{5.1}$$

This parameter is an average across tells us the average amount of additional tax HMRC should expect to receive h years after an audit in year t if they select taxpayers to audit uniformly randomly from the population of self-assessment taxpayers. We are interested in knowing this parameter averaged across all audit years, $\beta_h^{ATE} = \mathbb{E}\left[[]\, \beta_{th}^{ATE}\right]$. We are also interested in the corresponding parameters for reported total income and income components (rather than for tax liability).

To answer these questions, we compare a treatment group of all individuals who are selected for a random tax audit in year t, denoted $Z(t) = 1$, to a control group of individuals who *could have* been selected for a random audit, but who weren't, $Z(t) = 0$. Since we have data on the full population of taxpayers, our potential control group is very large. For computational reasons, we limit the control group to being six times the size of the treatment group in each year t. A simple comparison of the average tax paid between the two groups, gives us the average effect of being *assigned* to the audit group (sometimes described as the 'intention to treat' parameter):

$$\beta_{th}^{ITT} = \mathbb{E}\left[Y_{t+h}|Z(t) = 1\right] - \mathbb{E}\left[Y_{t+h}|Z(t) = 0\right] \tag{5.2}$$

Again, we average this across audit years, $\beta_h^{ITT} = \mathbb{E}\left[[]\, \beta_{th}^{ITT}\right]$. If compliance were perfect, so all individuals in the treatment group and none in the control group were audited, i.e. $Z(t) = D(t)$, then this would be equal to the average treatment effect parameter, β_h^{ATE}. However, there are three reasons why these parameters differ in the present context: (i) targeted audits; (ii) deselections; and (iii) incomplete auditing. We discuss these in turn.

First, HMRC's targeted audit programme selects for audit individuals who they believe are likely to be cheating. This happens after random audit cases have been selected (see Subsection 5.2.1). As a result, our control group will contain a small number of high-risk taxpayers who have been audited (Table 5.1 suggests that around 2 per cent of cases will be affected). To the extent that these individuals respond to audit, the intention to treat parameter will differ from the desired average treatment effect such that β_h^{ITT} will be attenuated relative to β_h^{ATE}.

A second issue is that some taxpayers are deselected from audit after being assigned to the treatment group. In the majority of cases, this happens because the notice to file has been sent to a taxpayer that does not need to file a return (e.g. they no longer meet the self-assessment criteria). This does not cause a problem for us, since the parameter of interest is defined for individuals who are in self assessment in the audit year. We therefore exclude these individuals from both the treatment and control groups. However, a small number of individuals are deselected despite filing a tax return. In 2008-09, this affected 4 per cent of those initially assigned to the treatment group. The main reason this happens is if an audit is already open or has recently concluded. In these cases, no audit will be carried out. This will again mean that the estimated β_h^{ITT} parameter will be attenuated relative to β_h^{ATE}.

A third reason why the intention to treat parameter will differ from the average treatment effect is that not all taxpayers who are selected for a random audit actually end up being audited. Having been selected for a random audit, taxpayers' details are passed to local compliance teams, where random (and targeted) audits are worked by compliance officers. While these officers are informed that random audit cases are considered high priority, they often receive more cases than they can complete and so some (around 5 per cent in 2008-09) end up not being worked. This again will tend to attenuate the estimated β_h^{ITT} parameter relative to β_h^{ATE}, but the effect of this may not be large if (as seems reasonable) the cases left until last are the ones with the lowest expected return.

These three reasons mean that the intention to treat parameter we recover will differ from the average treatment effect. Given the arguments above, the estimated β_h^{ITT} parameter will attenuated relative to β_h^{ATE}.

5.3.2 Assessing balancing

The approach just outlined exploits the randomisation of the random audit programme to select a suitable control group. For this to be a valid approach, we need random audits to have been selected randomly. We check this by carrying out a set of balancing tests to confirm whether outcomes in the treatment and control groups are balanced with each other in the run-up to the audit. We do this between five years and one year before the audit for income and tax totals, income components and individual characteristics.

Overall balancing statistics suggest that the samples are fairly well balanced. The p-value of the likelihood-ratio test of the joint insignificance of all the regressors is 0.181, while the mean and median absolute standardised percentage bias across all outcomes of interest are low at 2.4 per cent and 1.7 per cent respectively.[6] Similarly, Rubin's B and R statistics are 10.8 and 0.983, well within reasonable thresholds to consider the samples to be balanced.[7]

[6] The standardised percentage bias is the difference in the sample means between treated and control groups as a percentage of the square root of the average of the sample variances in the treated and control groups (see Rosenbaum and Rubin, 1985).

[7] Rubin's B is the absolute standardized difference of the means of the linear index of the propensity score in the treated and control group. Rubin's R is the ratio of treated to control variances of the propensity score index. Rubin (2001) recommends that B be less than 25 and that R be between 0.5 and 2 for the samples to be considered sufficiently balanced.

Table 5.4: *Balancing tests of characteristics, income and tax totals, and income components*

Years after audit		-5	-4	-3	-2	-1
			Characteristics			
Female	Mean	0.274	0.276	0.278	0.282	0.287
	Difference	-0.006	-0.004	-0.002	-0.001	-0.002
	p-value	0.221	0.359	0.606	0.627	0.863
Age	Mean	49.2	49.3	49.3	49.4	49.5
	Difference	0.2	0.3	0.3	0.2	0.2
	p-value	0.756	0.586	0.390	0.610	0.057
In London or SE	Mean	0.333	0.334	0.335	0.333	0.331
	Difference	-0.006	0.001*	0.004*	0.002	0.002
	p-value	0.177	0.025	0.011	0.281	0.152
Has tax agent	Mean	0.628	0.614	0.603	0.589	0.573
	Difference	0.000	0.002	0.001	0.003	0.005
	p-value	0.547	0.508	0.405	0.396	0.412
Survives	Mean	0.624	0.669	0.728	0.803	0.892
	Difference	0.032***	0.039***	0.047***	0.050***	0.050***
	p-value	0.000	0.000	0.000	0.000	0.000
			Income and tax totals			
Total taxable income	Mean	35,075	34,670	34,030	32,912	31,755
	Difference	881	492	403*	1,051*	1,095*
	p-value	0.374	0.157	0.028	0.012	0.012
Total tax	Mean	9,646	9,539	9,321	8,979	8,635
	Difference	260	63	82	310	337*
	p-value	0.539	0.303	0.055	0.064	0.027
			Income components			
Employment	Mean	22,508	22,534	22,266	21,708	21,145
	Difference	-31	-136	180*	909**	721*
	p-value	0.162	0.371	0.028	0.006	0.027
Self employment	Mean	6,546	6,379	6,200	5,950	5,581
	Difference	356	328	173	99	200*
	p-value	0.151	0.174	0.311	0.106	0.025
Interest and dividends	Mean	4,007	3,905	3,895	3,759	3,645
	Difference	-36	208	18	63	112
	p-value	0.767	0.432	0.700	0.578	0.580
Pensions	Mean	3,493	3,542	3,561	3,562	3,531
	Difference	176	168	128	148	159
	p-value	0.425	0.478	0.642	0.307	0.327
Property	Mean	869	844	811	769	726
	Difference	18	-2	37	47	31
	p-value	0.813	0.952	0.576	0.498	0.134
Foreign	Mean	194	193	194	181	169
	Difference	23	-1	-5	-6	1
	p-value	0.759	0.240	0.956	0.317	0.766
Trusts and estates	Mean	150	145	145	131	123
	Difference	46	17	8	19	8
	p-value	0.245	0.367	0.290	0.125	0.367
Share schemes	Mean	91	104	68	62	55
	Difference	17*	24	22**	11	-6**
	p-value	0.043	0.062	0.004	0.132	0.008
Other	Mean	80	75	76	73	71
	Difference	-1	-4	0	5	6
	p-value	0.747	0.675	0.645	0.796	0.167

Notes: *'Mean' is the mean outcome in the control (not selected for audit) group across all years. 'Difference' is the coefficient on the treatment dummy in a regression of the outcome on a treatment dummy. Treatment dummy equals 1 if taxpayer was selected by HMRC for a random audit. p-values are derived from an F-test that coefficients on all interactions between treatment and tax year dummies are all zero in a regression of the outcome of interest on tax year dummies and interactions between treatment and tax year dummies. Tests for all outcomes other than 'survives' are conditional on survives = 1. Monetary values are in 2012 prices. Standard errors are clustered by taxpayer. * $p < .05$, ** $p < .01$, *** $p < .001$.*
Source: *Authors' calculations based on HMRC administrative datasets.*

There are, however, some problems with individual outcome variables, as shown by the t-tests in Table 5.4. Most important of these is survival (whether the individual remains in self assessment): there is a strongly significant difference of between three and five percentage points in all five years before the audit. There are also statistically significant imbalances in total income and employment income in the three years immediately before audit, total tax and self-employment income in the year before audit, and income from share schemes in three of the five years before audit. Other income components and most demographics are well balanced. Overall, 18 out of 80 tests (22.5 per cent) are significant at the five per cent level.

As a consequence of these differences, we are reluctant to judge the audited and non-audited samples as being adequately balanced. Imbalances are the result of the audit timeline described in Subsection 5.2.1. To understand how this happens, consider random audits for the 2006-07 tax year. These were selected on the basis of information held by HMRC in around February 2007. Suppose that by then HMRC had the full list of filers for the 2005-06 tax year, and that they used only this to select the 2006-07 random audit sample. Then, by definition, every random audit case they draw will have a 2005-06 tax return. In contrast, we take the actual list of 2006-07 filers. Some of these will be new to self assessment in 2006-07, so will not be in the 2005-06 self assessment data. Consequently, the proportion of taxpayers that survive one year before the audit will be higher in the audited group than in the non-audited group—just as we find.

In reality, things were not quite that simple, but the same underlying arguments apply. In February 2007, HMRC did already know about some taxpayers who would be first-time filers for the 2006-07 tax year (e.g. because they had reported it to HMRC). This explains why not every randomly audited case has a tax return the year before the audit. In addition, HMRC did not have a complete list of all filers for the 2005-06 tax year: the filing deadline for these returns (31 January 2007) had only just passed so many of the returns were yet to be processed. This meant HMRC was partly working off information for 2004-05, two years prior to the tax year in question.

5.3.3 Propensity score matching approach

One seemingly simple solution to the lack of balancing would be for us to go back and draw our control sample on the basis of information HMRC held at the time random audits were selected. Unfortunately, this is not possible because HMRC has not held onto the relevant datasets. We therefore use propensity score matching techniques to reweight the non-audited group to look like the audited group.

Table 5.5: *Matched balancing tests of characteristics, income and tax totals, and income components*

Years after audit		-5	-4	-3	-2	-1
			Characteristics			
Survives	Mean	0.653	0.705	0.770	0.851	0.941
	Difference	0.002	0.002	0.004	0.001	0.000
	p-value	0.532	0.467	0.117	0.654	0.918
			Income and tax totals			
Total taxable income	Mean	22,718	24,120	25,910	27,729	29,480
	Difference	235	167	147	278	256
	p-value	0.562	0.639	0.663	0.367	0.447
Total tax	Mean	6,230	6,591	7,049	7,499	8,083
	Difference	55	-1	27	98	91
	p-value	0.704	0.991	0.818	0.378	0.452
			Income components			
Employment	Mean	14,534	15,681	17,087	18,667	20,174
	Difference	-146	-156	-55	158	26
	p-value	0.618	0.556	0.831	0.531	0.922
Self employment	Mean	4,340	4,569	4,844	5,116	5,415
	Difference	161	155	62	1	43
	p-value	0.107	0.086	0.466	0.991	0.604
Interest and dividends	Mean	2,588	2,693	2,893	3,116	3,438
	Difference	-79	111	60	23	2
	p-value	0.357	0.198	0.447	0.756	0.982
Pensions	Mean	2,364	2,592	2,823	3,130	3,507
	Difference	27	27	26	17	21
	p-value	0.660	0.651	0.635	0.756	0.709
Property	Mean	573	594	629	667	703
	Difference	3	-3	17	17	14
	p-value	0.915	0.882	0.432	0.405	0.508
Foreign	Mean	117	123	133	135	141
	Difference	14	-2	-3	-4	3
	p-value	0.304	0.887	0.791	0.712	0.775
Trusts and estates	Mean	87	93	107	100	97
	Difference	15	-5	-11	2	5
	p-value	0.284	0.663	0.292	0.825	0.621
Share schemes	Mean	59	76	50	40	24
	Difference	6	5	15	11	5
	p-value	0.701	0.758	0.232	0.252	0.412
Other	Mean	48	51	56	59	57
	Difference	0	-4	0	1	3
	p-value	0.966	0.419	0.952	0.876	0.463

Notes: *The variables matched upon are survives, total taxable income, total tax and all income components. 'Mean' is the mean outcome in the matched control (not selected for audit) group. 'Difference' is the mean difference between matched treated and control groups. p-values are derived from a test that the outcome of interest is not significantly different between matched treated and control groups. Monetary values are in 2012 prices.* * $p < .05$, ** $p < .01$, *** $p < .001$.
Source: *Authors' calculations based on HMRC administrative datasets.*

We use radius matching with a small caliper of 0.0005, though we have experimented with other matching techniques and it makes little difference to the results. We control for survival, income and tax totals and all the income components in the two years prior to audit.[8] We choose two years because of the two-year lag in information described above. We match separately for each audit year and then weight together the results across audit years based on the number of observations in the audited group. Table 5.5 shows that, once we perform the matching, all outcomes of interest are balanced across treated and control samples in all five years before audit.

5.4 Results

5.4.1 Aggregate results

In this subsection we present results for survival in self assessment, total taxable income and total tax; in the following subsection we break these results down by type of income. We consider outcomes between the tax year selected for audit and eight years after audit. We include the year of audit as a year potentially affected by the audit because, in some cases, HMRC adjusts tax returns to reflect underpayments they uncover. Table 5.6 sets out overall results for survival, total income and total tax liability. Taking survival first, it is not clear a priori what direction to expect the effect of audit to work: it could increase survival (e.g. because individuals become more diligent about filing their tax returns) or it could decrease survival (e.g. because borderline businesses are forced out of existence). In fact we find the former: survival increases by a modest but statistically significant 0.6 and 1.1 percentage points one and two years after audit. Given the lags involved in starting and completing audits (see Table 5.2), this corresponds to the period while many audits are being conducted. The increase in survival is short-lived, however, with estimates insignificant from year three onwards. This pattern is shown in Figure 5.1, which also plots pre-audit outcomes to demonstrate balancing. Note that there is no difference in the year of the audit because everyone in the treatment and control groups is necessarily in self assessment in that year.

Results for reported total taxable income and total tax in Table 5.6 are averages across all individuals, including those who do not survive in self assessment. For reported total taxable income, we find a statistically significant increase of £795 two years after audit, equivalent to 3.0 per cent of average total taxable income. This somewhat larger than the 1 per cent effect found by Kleven et al. (2011) for Denmark and of a comparable magnitude to the 2.7 per cent found by

[8] To get balanced samples, it turns out to be sufficient to control just for survival (results available from the authors on request). We additionally control for income and tax totals and the income components to reduce the amount of noise in outcomes in the run-up to the audit.

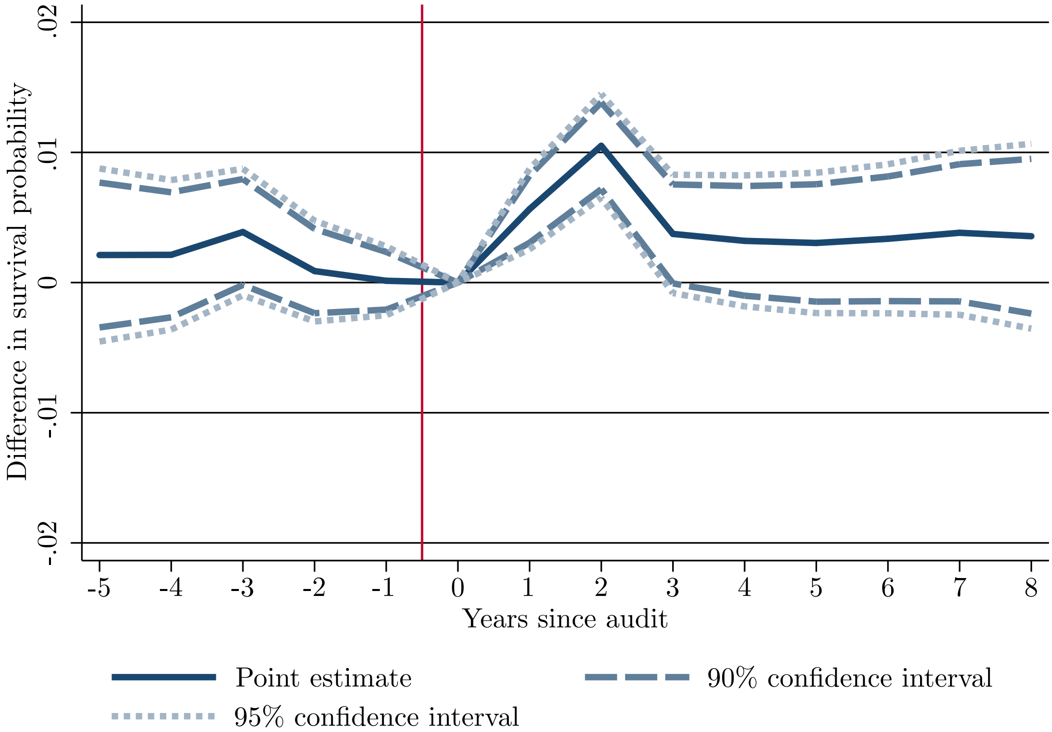
Table 5.6: *Impact of audit: aggregate results*

Years since audit		0	1	2	3	4	5	6	7	8
Survives	Mean	1	0.902	0.823	0.758	0.703	0.666	0.637	0.609	0.584
	Difference	0	0.006***	0.011***	0.004	0.003	0.003	0.003	0.004	0.004
	Std. err.	(0.000)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Reported total taxable income	Mean	30,783	28,693	26,886	25,190	24,456	24,161	23,712	22,837	22,296
	Difference	315	180	795*	510	633	457	447	280	615
	Std. err.	(367)	(376)	(388)	(393)	(460)	(486)	(539)	(559)	(630)
Total tax	Mean	8,452	7,876	7,394	6,923	6,772	6,696	6,649	6,385	6,156
	Difference	73	23	212	141	164	177	155	118	326
	Std. err.	(131)	(132)	(138)	(139)	(162)	(173)	(192)	(199)	(221)
N		278,449	278,449	278,449	278,449	259,957	241,744	223,321	190,295	152,899

Notes: *'Mean' is the mean outcome in the matched control group. 'Difference' is the mean difference between matched treated and control groups. 'Std. err.' is the standard error for the mean difference reported above. Estimates are derived from a propensity score matching procedure described in Subsection 5.3.3. Monetary values are in 2012 prices.* * $p < .05$, ** $p < .01$, *** $p < .001$.

Source: *Authors' calculations based on HMRC administrative datasets.*

Figure 5.1: *Impact of audit on survival in self assessment*



Notes: *'Point estimate' is the difference between treated and control groups and corresponds to the relevant 'Difference' row in Tables 5.5 and 5.6. Estimates are derived from a propensity score matching procedure described in Subsection 5.3.3.*
Source: *Authors' calculations based on HMRC administrative datasets.*

DeBacker et al. (2015) for the US. Increases in subsequent years are not significant but are still substantial at around £500 until at least the sixth year after audit. This is shown in Figure 5.2. For total tax, the point estimates are all positive—and indeed are all above £100 from year two onwards—but none are significant. This is plotted in Figure 5.3. Taking these point estimates at face value, they suggest a dynamic effect of audit (the amount of additional tax raised aside from any audit adjustment) of around 0.5 per cent of taxable income each year. The cumulative additional revenue raised between years one and eight is over £1,000, roughly the same order of magnitude as the average initial audit adjustment of £700.[9] Thus, for total taxable income and total tax, there is no evidence of a 'bomb-crater effect' or loss repair whereby individuals report less income after the audit. Since we cannot link self-assessment tax records to other tax records, we cannot see whether non-survivors leave the tax system altogether, or merely switch to paying tax only through withholding e.g. on employment earnings. For simplicity, our estimates ignore any spillover effects onto other means of tax collection. Also, we only observe what taxpayers report, so can't be sure whether any change in reported income reflects a change in true underlying income or just in what is reported (though the latter seems more likely).

[9] £700 is the product of the noncompliant share and the mean additional tax if noncompliant in Table 5.3.

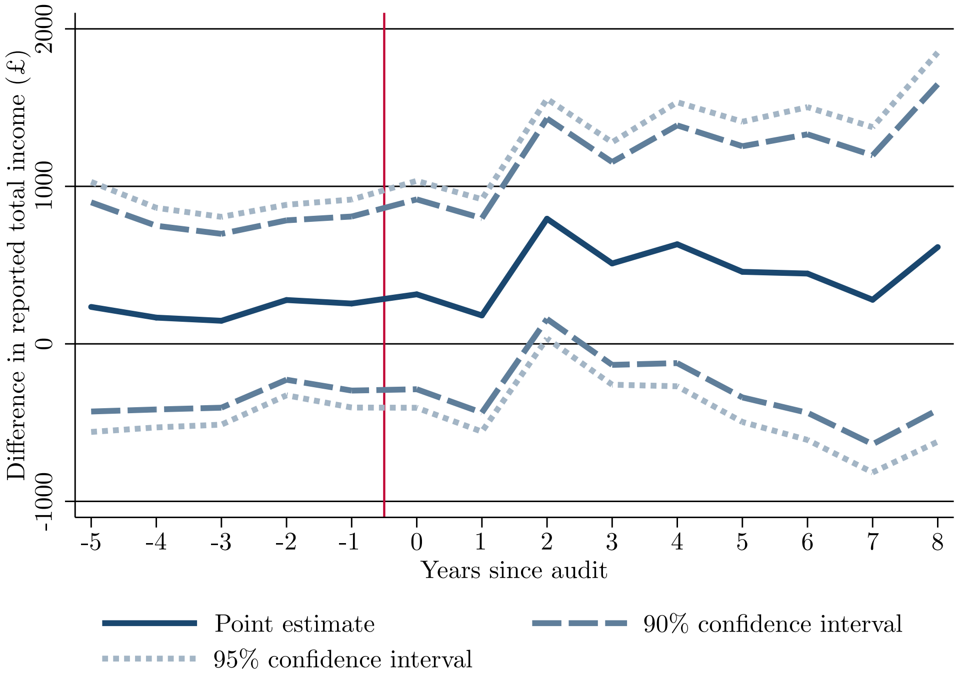


Figure 5.2: *Impact of audit on reported total taxable income*

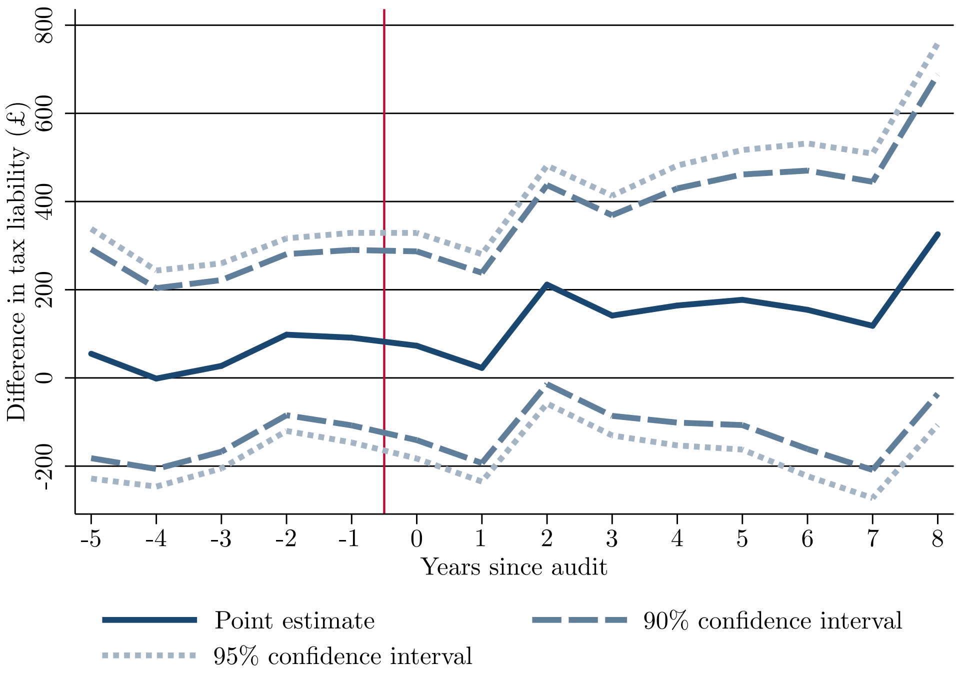


Figure 5.3: *Impact of audit on total tax*

Notes: *'Point estimate' is the difference between audited and non-audited groups and corresponds to the relevant 'Difference' row in Tables 5.5 and 5.6. Estimates are derived from a propensity score matching procedure described in Subsection 5.3.3.*
Source: *Authors' calculations based on HMRC administrative datasets.*

Table 5.7: *Proportion of individuals reporting positive income, by income component*

Income component	Proportion
Employment	0.309
Self employment	0.240
Interest and dividends	0.420
Pensions	0.193
Property	0.087
Foreign	0.031
Trusts and estates	0.007
Share schemes	0.001
Other	0.020

Notes: *Denominator includes non-surviving individuals.*
Source: *Authors' calculations based on HMRC administrative datasets.*

5.4.2 Results for different income components

We now consider how the results vary across different income components. To help interpret the results, Table 5.7 shows what proportion of individuals in the control group (including non-survivors) report positive amounts of each type of income. 31 per cent report employment earnings, 24 per cent report self employment profits, 42 per cent report interest and dividends and 19 per cent report pensions. All other categories are reported by less than 10 per cent of individuals.

Results split by income component are set out in Table 5.8. This table shows that there are significant impacts on reporting for three of the major types of income: self employment, pensions and property. The largest of these is reported self-employment profits, where we find positive dynamic impacts that peak at £272 (5.6 per cent of self-employment profits) in year two and last until year three but fall away sharply after that. Given that only 24 per cent of individuals report positive amounts of self employment income, this corresponds to over £1,000 for each self-employed individual, or a total of over £3,000 across the three years after audit. These patterns are plotted in Figure 5.4.

Reported pension income records a sustained increase that is significant at the five per cent level in two years (£135 in year four and £167 in year seven) and borderline significant in many others. This is shown in Figure 5.5. The values in the two significant years correspond to 3.6 and 4.5 per cent of pension income respectively. Given that 19 per cent of individuals report pension income, this equates to around £700 per pensioner.

There is a significant increase in reported property income of £48, £57 and £51 in years one, three and four respectively (see Figure 5.6). These figures correspond to between 6.6 and 7.9 per cent of property income. By year six, the difference has fallen back close to zero.

Table 5.8: *Impact of audit: results for reported income components*

Years since audit		0	1	2	3	4	5	6	7	8
Employment	Mean	20,907	18,418	16,375	14,580	13,556	13,003	12,679	12,130	11,964
	Difference	-162	-452	316	151	202	337	283	155	169
	Std. err.	(286)	(275)	(289)	(288)	(322)	(341)	(366)	(375)	(401)
Self employment	Mean	5,435	5,180	4,823	4,472	4,210	3,958	3,689	3,401	3,099
	Difference	243**	251**	272**	225**	117	41	13	34	117
	Std. err.	(84)	(86)	(85)	(85)	(83)	(84)	(87)	(90)	(102)
Interest and dividends	Mean	3,673	3,725	3,792	3,870	3,880	3,870	3,829	3,627	3,441
	Difference	-29	70	29	143	137	107	87	151	118
	Std. err.	(83)	(90)	(91)	(99)	(101)	(107)	(115)	(131)	(143)
Pensions	Mean	3,776	3,775	3,762	3,731	3,724	3,755	3,742	3,728	3,723
	Difference	102	87	92	100	135*	90	122	167*	174
	Std. err.	(58)	(60)	(61)	(62)	(66)	(69)	(73)	(80)	(90)
Property	Mean	724	730	723	721	727	734	742	734	745
	Difference	26	48*	34	57*	51*	42	14	22	14
	Std. err.	(22)	(23)	(22)	(23)	(24)	(25)	(26)	(28)	(32)
Foreign	Mean	153	154	155	153	157	153	158	152	137
	Difference	-2	1	-5	0	5	10	13	11	3
	Std. err.	(12)	(12)	(12)	(12)	(13)	(14)	(16)	(16)	(15)
Trusts and estates	Mean	100	100	98	89	88	85	83	78	77
	Difference	12	2	-2	6	6	4	7	7	4
	Std. err.	(11)	(11)	(10)	(11)	(11)	(11)	(12)	(12)	(13)
Share schemes	Mean	23	9	8	6	8	6	7	8	3
	Difference	8	0	-3*	-2	2	0	1	-4**	1
	Std. err.	(7)	(2)	(1)	(1)	(2)	(2)	(2)	(1)	(1)
Other	Mean	65	60	58	59	57	54	54	54	54
	Difference	3	7	5	-3	5	-1	-3	-4	-13*
	Std. err.	(5)	(5)	(5)	(5)	(5)	(5)	(5)	(6)	(6)
N		278,449	278,449	278,449	278,449	259,957	241,744	223,321	190,295	152,899

Notes: *'Mean' is the mean outcome in the matched control group. 'Difference' is the mean difference between matched treatment and control groups. 'Std. err.' is the standard error for the mean difference reported above. Estimates are derived from a propensity score matching procedure described in Subsection 5.3.3. Monetary values are in 2012 prices.* $* \ p < .05$, $** \ p < .01$, $*** \ p < .001$.
Source: *Authors' calculations based on HMRC administrative datasets.*

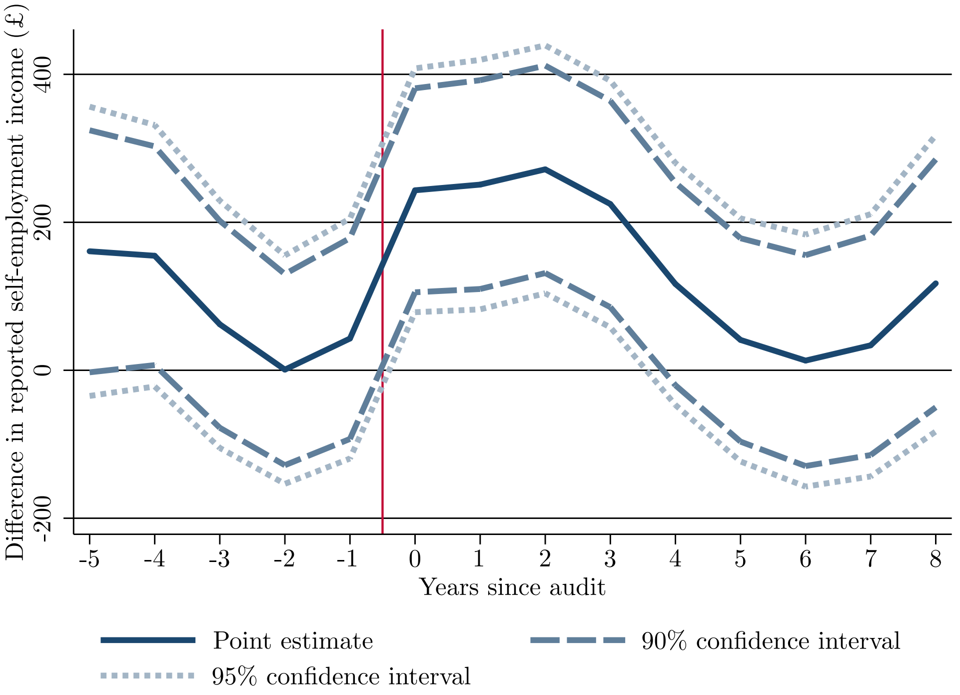


Figure 5.4: *Impact of audit on reported self employment profits*

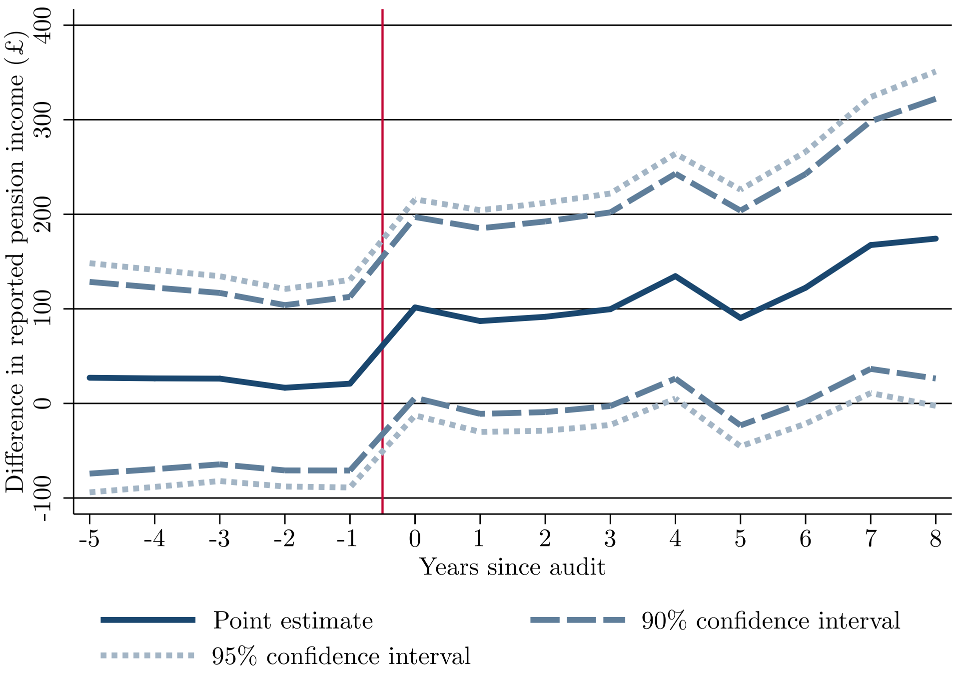
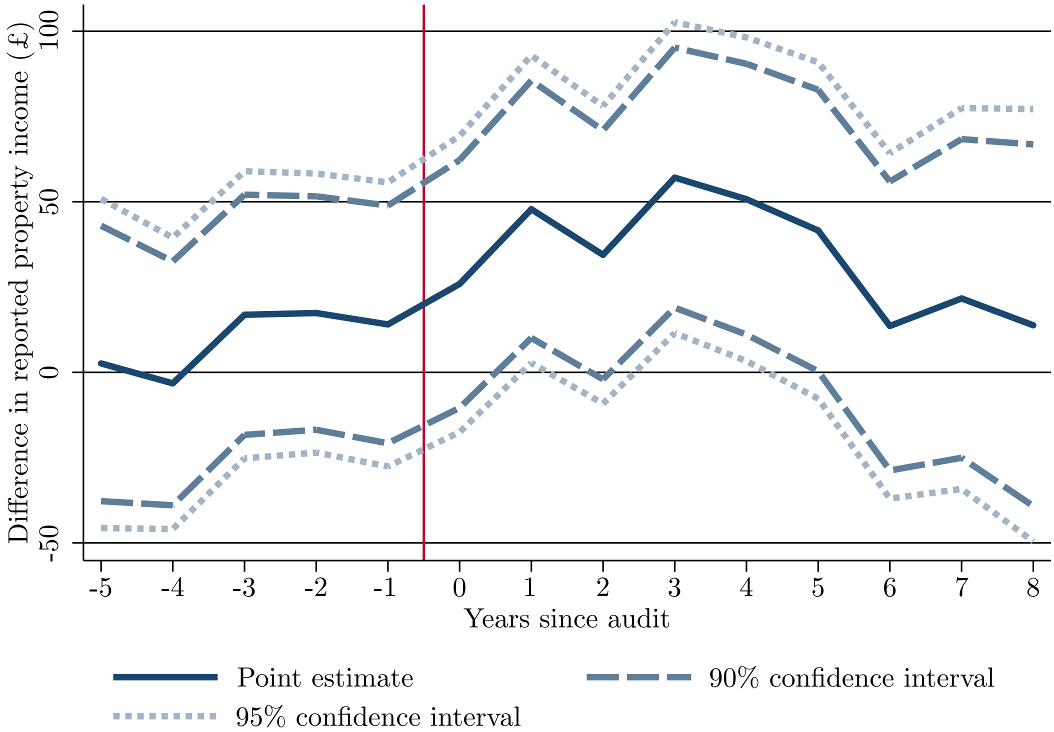


Figure 5.5: *Impact of audit on reported pension income*

Notes: *'Point estimate' is the difference between audited and non-audited groups and corresponds to the relevant 'Difference' row in Tables 5.5 and 5.6. Estimates are derived from a propensity score matching procedure described in Subsection 5.3.3.*
Source: *Authors' calculations based on HMRC administrative datasets.*

Figure 5.6: *Impact of audit on reported property income*



Notes: *'Point estimate' is the difference between treatment and control groups and corresponds to the relevant 'Difference' row in Tables 5.5 and 5.8. Estimates are derived from a propensity score matching procedure described in Subsection 5.3.3.*
Source: *Authors' calculations based on HMRC administrative datasets.*

Reported share schemes record a significant decrease in two years and other income a significant decrease in one year, but the magnitudes involved are small because such a small proportion of individuals report income of these types. As for the other income types—including employment income and interests and dividends—none of the differences in reported income are significant.

What might explain the differing patterns across income components? To get a handle on the mechanisms involved, Table 5.9 sets out the extent of third-party reporting arrangements across different income types. Third-party reporting describes the situation in which there is an obligation on an agent other than the taxpayer (e.g. the employer) to pass information to the tax authority about income received by the taxpayer. The table shows that there is third-party reporting for employment earnings, interest, pensions and share schemes; partial third-party reporting for foreign income, trusts and estates and other income; and no third-party reporting for dividends, self-employment profits and property income.

There are two main points to note. First, the biggest responses are for income components where there is little or no third-party reporting. Self-employment profits and property income—for which there is little third-party reporting—both exhibit substantial reported increases, peak-

Table 5.9: *Third-party reporting arrangements in the UK for different income components*

Income component	Degree of third-party reporting
Employment	Yes through the employer
Self employment	No unless an entertainer, sportsman or contractor in the construction industry
Interest and dividends	Yes through bank for interest, no for dividends
Pensions	Yes through pension provider
Property	No
Foreign	Depends on the jurisdiction
Trusts and estates	Partial but complicated
Share schemes	Yes
Other	Partial

Source: *Personal communication with Tracey Bowler, Tax Law Review Committee.*

ing at 5.6 per cent and 7.9 per cent respectively. In contrast, in most cases where there is extensive third-party reporting (e.g. employment income) no significant response is recorded. This is consistent with the findings in Kleven et al. (2011) and DeBacker et al. (2015), both of which highlight the importance of third-party information reporting. It also makes intuitive sense: in the absence of third-party reporting, opportunities for noncompliance are greater, and therefore it is more likely that there is misreporting that can be corrected. Second, although the impact on components with little or no third-party reporting lasts for a number of years, effects are ultimately transitory: four years after the audit the difference between treatment and control groups has largely returned to zero. This may suggest that taxpayers think they are being monitored for some time after audit but that this effect dies away so taxpayers revert to their previous behaviour.

For pension income, the story is different. Here there is widespread third-party reporting, but nevertheless audits have a dynamic impact and this impact is persistent. The most likely explanation is accidental errors in withholding. Pensioners often end up in self assessment because they have a number of different pension (and other) income sources that PAYE struggles to deal with accurately in combination. An audit may bring to light errors that previously lay hidden, allowing them to be corrected for all future years through revised withholding arrangements.

5.5 Conclusion

This chapter investigated the dynamic effects of audits on income reported in subsequent tax returns. Understanding these effects is important both from the perspective of quantifying the returns to the tax authority an audit, and for assessing the mechanisms by which audits might influence taxpayer behaviour. To answer this question we exploited a random audit program run by the UK tax authority (HMRC) under which an average of around 3,000 individuals are selected for audit each year. We used data on audits over the period 1998-99 to 2008-09 and tracked responses in tax returns between 1998-99 and 2011-12.

Our results suggest that audits lead to a modest increase in survival in self assessment for two years after the audit, peaking at 1.1 percentage points in year two before falling back. This pattern is consistent with individuals being more diligent in filing their tax returns while audit investigations are ongoing. For reported total taxable income, we find a statistically significant increase of £795 two years after audit. Increases in subsequent years are not significant but are still substantial at around £500 until at least the sixth year after audit. For total tax, the point estimates are all positive but are not significant.

Breaking these overall results down by income component reveals some marked differences that point towards the importance of third-party reporting for maintaining high levels of compliance. Reported self-employment profits and property income—neither of which are subject to third-party reporting—both record significant increases after audit, peaking at the equivalent of 5.6 per cent of self-employment profits and 7.9 per cent of property income. But these increases largely fall away after year four suggesting individuals are reverting to their old behaviour. In contrast, income components subject to third-party reporting, such as reported employment earnings, do not exhibit any significant impact suggesting there is already a high level of compliance for these components. Reported pension income does record a significant and persistent increase after audit despite extensive third-party reporting, a pattern that might be due to the correction of previously undetected errors.

Our results have three main policy implications. First, taking dynamic effects into account substantially increases the estimated revenue impact of audits. This suggests that the optimal audit rate should be increased relative to the situation where there are no dynamic effects. Second, the variation in dynamic effects observed across different income components alters the way in which targeted audits should be targeted: audits should focus more on individuals reporting types of income with the largest dynamic effects. For example, the maximum annual impact on reported self-employment income for each self-employed individual is over £1,000, higher than other components. This suggests focusing more on individuals reporting self-employment income. Likewise, although the maximum annual impact on pension income is lower, it is persistent, so there may be more incentive to target individuals reporting pension of income. Third, there are implications for setting optimal re-auditing strategies. Impacts for reported self-employment income and property income die away after about four years, so it might make sense to revisit these individuals after this sort of horizon. In contrast, the impact on reported pension income seems to persist for at least eight years, implying that there is less need to re-audit these individuals so soon.

We have only been able to investigate the dynamic impacts of audits on the subsequent behaviour of audited taxpayers. In future work it would be interesting to explore the extent of any spillovers from the audit onto the behaviour of other taxpayers who know an audited taxpayer.

Chapter 6

Conclusion

This thesis comprises four self-contained chapters on the economics of networks, each with its own conclusions. Rather than recapitulate these here, in this chapter I propose some directions for future work and lessons for policy, building on the work set out in the previous chapters. Following the structure of Chapter 2, these can be grouped into three broad areas: (i) social effects, (ii) network formation, and (iii) measurement.

6.1 Social Effects

Each of Chapters 3 to 5 offers a natural direction for the further study of social effects. In Chapter 3, I examined the potential for informal risk-sharing networks to allow borrowing for the purpose of (lumpy) investment. To model this, I studied a dynamic contracting model, in the spirit of Kocherlakota (1996) and Ligon et al. (2002), where any risk-sharing or borrowing agreement had to be self-enforcing. A useful simplification made in this chapter was that transfers could take place between any pair of agents, so network *structure* played no role. Work by e.g. Ambrus et al. (2014), Ambrus et al. (2015), and Milán (2016) shows how different results in pure risk-sharing models can look when network structure constrains the ability to risk share. Combining this network-constrained risk-sharing with the possibility of investment would be an interesting and useful path for further work if a more detailed quantitative understanding of the mechanism is to be achieved.

In Chapter 4, we explored the trade-off between economic and cultural incentives in determining choices by migrants. Recent work has shown the importance of migrant networks for labour market outcomes (Beaman, 2012; Patacchini and Zenou, 2012a; and Patel and Vella, 2013). A fruitful path for further study would be to combine these ideas, and use the discontinuity in network structure to examine directly how social incentives translate into economic outcomes, in terms of employment as well as return migration (which was important at this time; see Bandiera et al., 2013). This allows a quantitative assessment of the value of culture, based on the financial cost to migrants of retaining their culture, as well as the 'price of anarchy', the inefficiency cost from being potentially unable to coordinate on the efficient equilibrium.

In Chapter 5 we studied the indirect revenue raised due to *dynamic effects*: the impact that being audited has on a taxpayer's future tax declarations. A natural avenue for further work, discussed above, is to study *spillover effects*: the indirect revenue raised from audits by changing the declarations of taxpayers connected to the audited taxpayer. These might arise due to salience effects of the likelihood of audit, additional information about how taxes should be correctly filed, or shifts in beliefs about the ability of the tax authority to detect errors (or some combination of these). As with dynamic effects, capturing spillovers is important for determining the optimal audit rate.

6.2 Network Formation

Another broad area of research in networks that is ripe for further work is the interplay between economic incentives and network formation decisions. Chapter 4 studied this explicitly, considering the choice by migrants of what links to form, when the costs and benefits were endogenous. These represented the costs of interacting with someone of similar or different culture, and the gains from this interaction. Extending this idea into, say, the model of Chapter 3 would allow the study of additional related questions.

The model presented assumed a complete network, among an appropriately selected subgroup of the village. This was justified by the lack of fixed costs of creating or maintaining links. With a complete network, risk-sharing transfers would in principle be possible between all pairs of households, though in practice some pairs may never need to make any bilateral transfer. If instead there were fixed costs of some intermediate size, an insurance network would exist but it would be incomplete, as some links may not provide enough benefit to be worth maintaining.

As noted in the previous section, this has important implications for the amount of risk-sharing that takes place, the distribution of surplus from risk-sharing, and the ability to borrow for investment. Households will make choices about what links to form based on the cost of these links, and the gains that arise endogenously from being able to make transfers. Capturing this is important both for welfare analysis and policymaking. From a welfare standpoint, if network structure is not a constraint but a choice, then differences in network position (partly) reflect differences in households' valuations of those links, although there may also be externalities that affect choices. Malleability of network structure is also important for policy, since it drives a wedge between the direct, partial equilibrium effect of the policy, when network structure has not responded, and the general equilibrium effect once link choices adjust.[1, 2] This is closely related to the well-understood result that formal institutions affect the functioning of informal ones e.g. Attanasio and Rios-Rull (2000), and Ligon et al. (2000).

6.3 Measurement

A final important direction for future research is in thinking about issues of measurement. In Chapter 2, we discussed two important measurement issues in networks: the choice of sampling scheme, and the problems of measurement error. However, a third crucial issue, which has not yet received much attention, is the deceptively difficult decision of how to define a link between two nodes. Links may be directed or undirected: for example, a loan will typically be *from* one individual *to* another, whilst friendship is typically two-way. Links may be binary or weighted: they may measure *whether* two individuals are friends, or how much time they spend together. Links may also have many dimensions: a student might study with one group of people, play sports with another group, and ask advice from yet another.

These choices of definition can be critical. For example, if only the 'extensive margin' of link presence were measured, one might conclude that a policy led to link severing, when instead only the intensity of the link changed. It is also important in the absence of a program: when we observe informal transfers between some set of individuals, it does not necessarily tell us whether these are the same paths through which information flows occur. Recent work (e.g. Banerjee et al., 2015a) has begun to recognise the importance of considering intensive as well as extensive margins, and different dimensions of relationships, but there is little firm guidance in how to handle these issues generally. Given the particular problems caused by measurement error in the context of networks, as discussed in Chapter 2, this is a particularly important area for future research.

[1] For example, work by Banerjee et al. (2015a) finds that offering microfinance in India can crowd out or crowd in informal credit, depending on the borrower type

[2] Some inital discussion of how to estimate such a model was provided in Chapter 2. See also Badev (2013) for an example in the context of friendship formation and smoking decisions.

Bibliography

Ábrahám, A. and E. Cárceles-Poveda (2009): "Competitive Equilibria with Production and Limited Commitment," *mimeo.*

Ábrahám, A. and S. Laczó (2014): "Efficient Risk Sharing with Limited Commitment and Storage," *mimeo.*

Abramitzky, R., L. P. Boustan, and K. Eriksson (2014): "Cultural Assimilation during the Age of Mass Migration," *mimeo.*

Abreu, D., D. Pearce, and E. Stacchetti (1990): "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, 58, 1041–1063.

Advani, A. N. and I. Rasul (2015): "Social Networks and Success in the US," *mimeo.*

Aghion, P. and P. Bolton (1997): "A Theory of Trickle-Down Growth and Development," *Review of Economic Studies*, 64, 151–172.

Akerlof, G. A. and R. E. Kranton (2000): "Economics And Identity," *Quarterly Journal of Economics*, 115, 715–753.

——— (2002): "Identity and Schooling: Some Lessons for the Economics of Education," *Journal of Economic Literature*, 40, 1167–1201.

Alatas, V., A. Banerjee, A. G. Chandrasekhar, R. Hanna, and B. A. Olken (2014): "Network Structure and the Aggregation of Information: Theory and Evidence from Indonesia," *mimeo, MIT.*

Alba, Richard, D. (1985): *Italian Americans. Into the Twilight of Ethinicity*, Prentice-Hall, Inc. New Jersey, second ed.

Alesina, A., R. Baqir, and W. Easterly (1999): "Public Goods And Ethnic Divisions," *Quarterly Journal of Economics*, 114, 1243–1284.

Alesina, A., J. Harnoss, and H. Rapoport (2013): "Birthplace Diversity and Economic Prosperity," NBER Working Papers 18699, National Bureau of Economic Research.

Alesina, A. and E. La Ferrara (2005): "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, 43, 762–800.

Alesina, A. and B. Reich (2015): "Nation Building," *NBER Working Paper No. 18839*.

Alesina, A. and E. Zhuravskaya (2011): "Segregation and the Quality of Government in a Cross Section of Countries," *American Economic Review*, 101, 1872–1911.

Algan, Y., T. Mayer, and M. Thoenig (2013): "The Economic Incentives of Cultural Transmission: Spatial Evidence from Naming Patterns Across France," *CEPR discussion paper DP9416*.

Alvarez, F. and U. J. Jermann (2000): "Efficiency, Equilibrium, and Asset Pricing with Risk of Default," *Econometrica*, 68, 775–797.

Ambrus, A., A. G. Chandrasekhar, and M. Elliott (2015): "Social Investments, Informal Risk Sharing, and Inequality," NBER Working Papers 20669, National Bureau of Economic Research.

Ambrus, A., M. Mobius, and A. Szeidl (2014): "Consumption Risk-Sharing in Social Networks," *American Economic Review*, 104, 149–182.

Andreoni, J., B. Erard, and J. Feinstein (1998): "Tax Compliance," *Journal of Economic Literature*, 36, 818–860.

Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–56.

Angelucci, M., G. De Giorgi, M. A. Rangel, and I. Rasul (2010): "Family Networks and School Enrolment: Evidence from a Randomized Social Experiment," *Journal of Public Economics*, 94, 197–221.

Angelucci, M., G. De Giorgi, and I. Rasul (2015): "Resource Sharing Within Family Networks: Insurance and Investment," *mimeo*.

——— (forthcoming): "Consumption and Investment in Resource Pooling Family Networks," *Economic Journal*.

Angrist, J. (2013): "The Perils of Peer Effects," NBER Working Papers 19774, National Bureau of Economic Research.

Ashraf, Q. and O. Galor (2013): "The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *American Economic Review*, 103, 1–46.

Attanasio, O. and N. Pavoni (2011): "Risk Sharing in Private Information Models with Asset Accumulation: Explaining the Excess Smoothness of Consumption," *Econometrica*, 79, 1027–1068.

Attanasio, O. and V. Rios-Rull (2000): "Consumption smoothing in island economies: Can public insurance reduce welfare?" *European Economic Review*, 44, 1225–1258.

Austen-Smith, D. and R. G. Fryer (2005): "An Economic Analysis of Acting White," *Quarterly Journal of Economics*, 120, 551–583.

Azevedo, F. A., L. R. Carvalho, L. T. Grinberg, J. M. Farfel, R. E. Ferretti, R. E. Leite, W. Jacob Filho, R. Lent, and S. Herculano-Houzel (2009): "Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain," *Journal of Computational Neurology*, 513, 532–541.

Badev, A. I. (2013): "Discrete Games in Endogenous Networks: Theory and Policy," *mimeo*.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006): "Who's Who in Networks. Wanted: the key player," *Econometrica*, 74, 1403–1417.

Bandiera, O., R. Burgess, N. Das, S. Gulesci, I. Rasul, and M. Sulaiman (forthcoming): "Labor Markets and Poverty in Village Economies," *Quarterly Journal of Economics*.

Bandiera, O., M. Mohnen, I. Rasul, and M. Viarengo (2017): "Nation-Building Through Compulsory Schooling During the Age of Mass Migration," *mimeo*.

Bandiera, O. and I. Rasul (2006): "Social Networks and Technology Adoption in Northern Mozambique," *Economic Journal*, 116, 869–902.

Bandiera, O., I. Rasul, and M. Viarengo (2013): "The Making of Modern America: Migratory Flows in the Age of Mass Migration," *Journal of Development Economics*, 102, 23–47.

Banerjee, A., E. Breza, E. Duflo, and C. Kinnan (2015a): "Do Credit Contstraints Limit Entrepreneurship? Heterogeneity in the Returns to Microfinance," *mimeo*.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. Jackson (2013): "The Diffusion of Microfinance," *Science*, 341.

Banerjee, A. V., E. Duflo, N. Goldberg, R. Osei, W. Parienté, J. Shapiro, B. Thuysbaert, and C. Udry (2015b): "A multifaceted program causes lasting progress for the very poor: Evidence from six countries," *Science*, 348, 1260799.

Banerjee, A. V. and S. Mullainathan (2010): "The Shape of Temptation: Implications for the Economic Lives of the Poor," .

Banerjee, A. V. and A. F. Newman (1993): "Occupational Choice and the Process of Development," *Journal of Political Economy*, 101, 274–298.

Beaman, L. (2012): "Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the US," *Review of Economic Studies*, 79, 128–161.

Berman, E. (2000): "Sect, Subsidy, And Sacrifice: An Economist's View Of Ultra-Orthodox Jews," *Quarterly Journal of Economics*, 115, 905–953.

Bernheim, D., D. Ray, and Ş. Yeltekin (2015): "Self-Control and Poverty," *Econometrica*, 83, 1877–1911.

Berry, J. W. (1997): "Immigration, Acculturation, and Adaptation," *Applied Psychology: An International Review*, 46, 5–68.

Besag, J. (1974): "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society, Series B*, 36, 192–236.

——— (1975): "Statistical Analysis of Non-Lattice Data," *The Statistician*, 24, 179–195.

Bhamidi, S., G. Bresler, and A. Sly (2008): "Mixing Time of Exponential Random Graphs," Arxiv preprint arXiv:0812.2265.

Bisin, A., A. Moro, and G. Topa (2011a): "The Empirical Content of Models with Multiple Equilibria in Economies with Social Interactions," NBER Working Papers 17196, National Bureau of Economic Research.

Bisin, A., E. Patacchini, T. Verdier, and Y. Zenou (2011b): "Formation and Persistence of Oppositional Identities," *European Economic Review*, 55, 1046–1071.

——— (2013): "Bend it Like Beckham: Ethnic Identity and Integration," *mimeo.*

Bisin, A., G. Topa, and T. Verdier (2004): "Religious Intermarriage and Socialization in the United States," *Journal of Political Economy*, 112, 615–664.

Bisin, A. and T. Verdier (2000): "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits," *Quarterly Journal of Economics*, 115, 955–988.

Bloch, F., G. Genicot, and D. Ray (2008): "Informal Insurance in Social Networks," *Journal of Economic Theory*, 143(1), 36–58.

Bloomquist, K. (2013): "Incorporating Indirect Effects in Audit Case Selection: An Agent-Based Approach," *IRS Research Bulletin*, forthcoming.

Blume, L. E., W. A. Brock, S. N. Durlauf, and Y. M. Ioannides (2010): "Identification of Social Interactions," in *Handbook of Social Economics*, ed. by J. Benhabib, A. Bisin, and M. Jackson, North Holland, vol. 1B.

Blume, L. E., W. A. Brock, S. N. Durlauf, and R. Jayaraman (2013): "Linear Social Interaction Models," NBER Working Papers 19212, National Bureau of Economic Research.

Booij, A. S., E. Leuven, and H. Oosterbeek (2015): "Ability Peer Effects in University: Evidence from a Randomized Experiment," IZA Discussion Papers 8769, Institute for the Study of Labor (IZA).

Boucher, V. and I. Mourifié (2013): "My Friend Far Far Away: Asymptotic Properties of Pairwise Stable Networks," Working papers, University of Toronto Dept. of Economics.

Bramoullé, Y., S. Currarini, M. O. Jackson, P. Pin, and B. W. Rogers (2012): "Homophily and Long-run Integration in Social Networks," *Journal of Economic Theory*, 147, 1754–1786.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009): "Identification of Peer Effects through Social Networks," *Journal of Econometrics*, 150, 41–55.

Bramoullé, Y. and R. Kranton (2007): "Public Goods in Networks," *Journal of Economic Theory*, 135(1), 478–494.

Bramoullé, Y., R. Kranton, and M. D'Amours (2014): "Strategic Interaction and Networks," *American Economic Review*, 104(3), 898–930.

Brock, W. A. and S. N. Durlauf (2001): "Discrete Choice with Social Interactions," *Review of Economic Studies*, 68, 235–260.

——— (2007): "Identification of Binary Choice Models with Social Interactions," *Journal of Econometrics*, 140, 52–75.

Bruttel, L. and T. Friehe (2014): "On the Path Dependence of Tax Compliance," *European Economic Review*, 65, 90–107.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009): "Peer Effects and Social Networks in Education," *Review of Economic Studies*, 76, 1239–1267.

Card, D., A. Mas, and J. Rothstein (2008): "Tipping and the Dynamics of Segregation," *Quarterly Journal of Economics*, 123, 177–218.

Carrell, S., R. Fullerton, and J. West (2009): "Does Your Cohort Matter? Estimating Peer Effects in College Achievement," *Journal of Labor Economics*, 27(3), 439–464.

Carrell, S., B. Sacerdote, and J. West (2013): "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 81(3), 855–882.

Carvalho, J.-P. (2013a): "Coordination and Culture," *mimeo.*

——— (2013b): "Veiling," *Quarterly Journal of Economics*, 128, 337–370.

Chandrasekhar, A. G. and M. O. Jackson (2014): "Tractable and Consistent Exponential Random Graph Models'," NBER Working Papers 20276, National Bureau of Economic Research.

Chandrasekhar, A. G. and R. Lewis (2011): "Econometrics of Sampled Networks," *mimeo, Massachusetts Institute of Technology.*

Chatterjee, S., P. Diaconis, and A. Sly (2010): "Random Graphs with a Given Degree Sequence," Arxiv preprint arXiv:1005.1136.

Chay, K. and K. Munshi (2013): "Black Networks After Emancipation: Evidence from Reconstruction and the Great Migration," *mimeo.*

Chen, X., H. Hong, and D. Nekipelov (2011): "Nonlinear Models of Measurement Errors," *Journal of Economic Literature*, 49(4), 901–937.

Chinchalkar, S. (1996): "An Upper Bound for the Number of Reachable Positions," *ICCA Journal*, 19, 181–183.

Choo, L., M. Fonseca, and G. Myles (2013): "Lab Experiment to Investigate Tax Compliance: Audit Strategies and Messaging," Research Report 308, HM Revenue and Customs.

Christakis, N. A., J. H. Fowler, G. W. Imbens, and K. Kalyanaraman (2010): "An Empirical Model for Network Formation," NBER Working Papers 16039, National Bureau of Economic Research.

Chuang, Y. and L. Schechter (2014): "Social Networks In Developing Countries," .

Cohen-Cole, E., X. Liu, and Y. Zenou (forthcoming): "Multivariate Choice and Identification of Social Interactions," *Journal of Econometrics.*

COMOLA, M. AND M. FAFCHAMPS (2014): "Estimating Mis-reporting in Dyadic Data: Are Transfers Mutually Beneficial?" *mimeo, Paris School of Economics*, 124, 954–976.

COMOLA, M. AND S. PRINA (2014): "Do Interventions Change the Network? A Dynamic Peer Effect Model Accounting for Network Changes," *SSRN Working Paper No. 2250748*.

CONLEY, T. G. AND C. R. UDRY (2010): "Learning About a New Technology: Pineapple in Ghana," *American Economic Review*, 100, 35–69.

CONTI, G., A. GALEOTTI, G. MUELLER, AND S. PUDNEY (2013): "Popularity," *Journal of Human Resources*, 48(4), 1072–1094.

COSTENBADER, E. AND T. W. VALENTE (2003): "The stability of centrality measures when networks are sampled," *Social Networks*, 25, 283–307.

CURRARINI, S., M. O. JACKSON, AND P. PIN (2009): "An Economic Model of Friendship: Homophily, Minorities, and Segregation," *Econometrica*, 77, 1003–1045.

——— (2010): "Identifying the Roles of Race-based Choice and Chance in High School Friendship Network Formation," *Proceedings of the National Academy of Sciences of the USA*, 107, 4857–4861.

CURRARINI, S. AND F. VEGA-REDONDO (2011): "A Simple Model of Homophily in Social Networks," *mimeo*.

CUTLER, D. M. AND E. L. GLAESER (1997): "Are Ghettos Good or Bad?" *Quarterly Journal of Economics*, 112, 827–72.

DASGUPTA, P. AND S. GOYAL (2009): "Narrow Identities," *mimeo*.

DE GIORGI, G., M. PELLIZZARI, AND S. REDAELLI (2010): "Identification of Social Interactions through Partially Overlapping Peer Groups," *American Economic Journal: Applied Economics*, 2, 241–275.

DE MEL, S., D. MCKENZIE, AND C. WOODRUFF (2008): "Returns to Capital in Microenterprises: Evidence from a Field Experiment," *The Quarterly Journal of Economics*, 123, 1329–1372.

DEBACKER, J., B. HEIM, A. TRAN, AND A. YUSKAVAGE (2015): "Once Bitten, Twice Shy? The Lasting Impact of IRS Audits on Individual Tax Reporting," Mimeo.

DEGROOT, M. (1974): "Reaching a Consensus," *Journal of the American Statistical Association*, 69, 118–121.

Ductor, L., M. Fafchamps, S. Goyal, and M. van der Leij (2014): "Social Networks and Research Output," *Review of Economics and Statistics*, Forthcoming.

Duflo, E., P. Dupas, and M. Kremer (2011): "Peer Effects and the Impacts of Tracking: Evidence from a Randomized Evaluation in Kenya," *American Economic Review*, 101, 1739–1774.

Dupas, P. and J. Robinson (2013): "Why Don't The Poor Save More? Evidence from health savings experiments," *American Economic Review*, 103, 1138–1171.

Dyen, I., J. B. Kruskal, and P. Black (1992): "An Indoeuropean Classification: A Lexicostatistical Experiment," *Transactions of the American Philosophical Society*, 82, iii–iv+1–132.

Dzemski, A. (2014): "An empirical model of dyadic link formation in a network with unobserved heterogeneity," *mimeo.*

Easterly, W. and R. Levine (1997): "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, 112, 1203–50.

Eguia, J. X. (2013): "Discrimination and Assimilation," *mimeo.*

Erard, B. (1992): "The Influence of Tax Audits on Reporting Behaviour," in *Why People Pay Taxes: Tax Compliance and Enforcement*, ed. by J. Slemrod, University of Michigan Press, chap. 5, 95–114.

Erdős, P. and A. Rényi (1959): "On Random Graphs," *Publicationes Mathematicae*, 6, 290–297.

Fafchamps, M. and F. Gubert (2007): "The Formation of Risk Sharing Networks," *Journal of Development Economics*, 83, 326–350.

Fafchamps, M. and J. Pender (1997): "Precautionary Saving Credit Constraints And Irreversible Investment: Theory And Evidence From Semiarid India," *Journal of Business and Economic Statistics*, 15, 180–194.

Fafchamps, M. and P. Vicente (2013): "Political Violence and Social Networks: Experimental Evidence," *Journal of Development Economics*, 101(C), 27–48.

Fearon, J. D. (2003): "Ethnic and Cultural Diversity by Country," *Journal of Economic Growth*, 8, 195–222.

Foster, A. and M. R. Rosenzweig (1994): "A Test for Moral Hazard in the Labor Market: Contractual Arrangements, Effort, and Health," *Review of Economics and Statistics*, 76, 213–27.

Fouka, V. (2014): "Backlash: The Unintended Effects of Language Prohibition in US Schools after World War I," *mimeo.*

Frank, O. (1978): "Sampling and Estimation in Large Social Networks," *Social Networks*, 1, 91–101.

——— (1980a): "Estimation of the Number of Vertices of Different Degrees in a Graph," *Journal of Statistical Planning and Inference*, 4, 45–50.

——— (1980b): "Sampling and Inference in a Population Graph," *International Statistical Review/Revue Internationale de Statistique*, 48(1), 33–41.

——— (1981): "A Survey of Statistical Methods for Graph Analysis," *Sociological Methodology*, 23, 110–155.

Frank, O. and D. Strauss (1986): "Markov Graphs," *Journal of the American Statistical Association*, 81, 832–842.

Frantz, T. L., M. Cataldo, and K. Carley (2009): "Robustness of Centrality Measures Under Uncertainty: Examining the Role of Network Topology," *Computational and Mathematical Organization Theory*, 15, 303–328.

Fulford, S. L., I. Petkov, and F. Schiantarelli (2015): "Does It Matter Where You Came From: Ancestry Composition and Economic Performance of U.S. Counties, 1850-2010," *IZA Discussion Paper no. 9060.*

Galaskiewicz, J. (1991): "Estimating Point Centrality Using Different Network Sampling Techniques," *Social Networks*, 13, 347–386.

Galor, O. and J. Zeira (1993): "Income Inequality and Income Growth," *Review of Economic Studies*, 60, 35–52.

Gemmell, N. and M. Ratto (2012): "Behavioral Responses to Taxpayer Audits: Evidence from Random Taxpayer Inquiries," *National Tax Journal*, 65, 33–58.

Geyer, C. and E. Thompson (1992): "Constrained Monte Carlo Maximum Likelihood for Dependent Data," *Journal of the Royal Statistical Society, Series B*, 54 (3), 657–699.

Ghatak, M. (2015): "Theories of Poverty Traps and Anti-Poverty Policies," *World Bank Economic Review, Papers and Proceedings of the ABCDE*, 29, S77–S105.

Gilbert, E. N. (1959): "Random Graphs," *Annals of Mathematical Statistics*, 30, 1141–1144.

Glaeser, E., B. Sacerdote, and J. Scheinkman (1996): "Crime and Social Interactions," *Quarterly Journal of Economics*, 115, 811–846.

Goldberg, D. and F. Roth (2003): "Assessing experimentally derived interactions in a small world," *Proceedings of the National Academy of Sciences of the USA*, 100 (8), 4372–4376.

Goldsmith-Pinkham, P. and G. W. Imbens (2013): "Social Networks and the Identification of Peer Effects," *Journal of Business and Economic Statistics*, 31, 253–264.

Gordon, Milton, M. (1964): *Assimilation in American Life*, Oxford University Press, New York, second ed.

Goyal, S. and F. Vega-Redondo (2005): "Network Formation and Social Coordination," *Games and Economic Behavior*, 50, 178–207.

Graham, B. S. (2008): "Identifying Social Interactions through Conditional Variance Restrictions," *Econometrica*, 76, 643–660.

——— (2015): "Methods of Identification in Social Networks," *Annual Review of Economics*, 7.

Granovetter, M. S. (1973): "The Strength of Weak Ties," *American Journal of Sociology*, 78, 1360–1380.

Grimmett, G. R. (1973): "A theorem about random fields," *Bulletin of the London Mathematical Society*, 5, 81–84.

Gualdani, C. (2017): "An Econometric Model of Network Formation with an Application to Board Interlocks between Firms," *mimeo*.

Hammersley, J. and P. Clifford (1971): "Markov fields on finite graphs and lattices," *mimeo*.

Handcock, M. S. (2003): "Assessing Degeneracy in Statistical Models of Social Networks," CSSS Working Paper 39, Center for Statistics and the Social Sciences.

Hansen, B. E. (1999): "Threshold Effects in Non-Dynamic Panels: Estimation, Testing and Inference," *Journal of Econometrics*, 93, 345–368.

Henderson, D. J., C. F. Parmeter, and L. Su (2015): "Nonparametric Threshold Regression: Estimation and Inference," *mimeo*.

Hoff, P. (2009): "Multiplicative latent factor models for description and prediction of social networks," *Computational and Mathematical Organization Theory*, 15 (4), 261–272.

Holland, P. W. and S. Leinhardt (1981): "An Exponential Family of Probability Distributions for Directed Graphs," *Journal of the American Statistical Association*, 76, 33–50.

Hsieh, C.-S. and L.-F. Lee (2014): "A Social Interactions Model with Endogenous Friendship Formation and Selectivity," *Journal of Applied Econometrics*.

Hubert, L. J. (1987): *Assignment Methods in Combinatorial Data Analysis*, Marcel Dekker.

Hubert, L. J. and J. Schultz (1976): "Quadratic Assignment as a General Data Analysis Strategy," *British Journal of Mathematical and Statistical Psychology*, 29, 190–241.

Iannaccone, L. R. (1992): "Sacrifice and Stigma: Reducing Free-Riding in Cults, Communes, and Other Collectives," *Journal of Political Economy*, 100, 271–91.

Ikegami, M., M. R. Carter, C. B. Barrett, and S. A. Janzen (2016): "Poverty Traps and the Social Protection Paradox," NBER Working Papers 22714, National Bureau of Economic Research.

Jackson, M. (2008): *Social and Economic Networks*, Princeton University Press.

Jackson, M., T. Rodriguez-Barraquer, and X. Tan (2012): "Social Capital and Social Quilts: Network Patterns of Favor Exchange," *American Economic Review*, 102(5), 1857–1897.

Jackson, M. O. (2009): "Networks and Economic Behavior," *Annual Review of Economics*, 1, 489–513.

Jackson, M. O. and A. Watts (2002): "On the Formation of Interaction Networks in Social Coordination Games," *Games and Economic Behavior*, 41, 265–291.

Jackson, M. O. and A. Wolinsky (1996): "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71, 44–74.

Jakiela, P. and O. Ozier (2016): "Does Africa Need a Rotten Kin Theorem? Experimental Evidence from Village Economies," *Review of Economic Studies*, 83, 231–268.

Karlan, D., R. Osei, I. Osei-Akoto, and C. Udry (2014): "Agricultural Decisions after Relaxing Credit and Risk Constraints," *Quarterly Journal of Economics*, 129, 597–652.

Kastlunger, B., E. Kirchler, L. Mittone, and J. Pitters (2009): "Sequences of Audits, Tax Compliance, and Taxpaying Strategies," *Journal of Economic Psychology*, 30, 405–418.

Kelejian, H. H. and G. Piras (2014): "Estimation of Spatial Models with Endogenous Weighting Matrices, and an Application to a Demand Model for Cigarettes," *Regional Science and Urban Economics*, 46, 140–149.

Kim, P. and H. Jeong (2007): "Reliability of Rank Order in Sampled Networks," *The European Physical Journal B*, 55, 109–114.

Kleven, H. J., M. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011): "Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark," *Econometrica*, 79, 651–692.

Kocherlakota, N. (1996): "Implications of Efficient Risk Sharing without Commitment," *Review of Economic Studies*, 63, 595–609.

Kolaczyk, E. (2009): *Statistical Analysis of Network Data*, Springer.

König, M., X. Liu, and Y. Zenou (2014): "R&D Networks: Theory, Empirics, and Policy Implications," CEPR Discussion Paper 9872, Centre for Economic Policy Research.

Kossinets, G. (2006): "Effects of Missing Data in Social Networks," *Social Networks*, 28, 247–268.

Krackhardt, D. (1988): "Predicting with Networks: A Multiple Regression Approach to Analyzing Dyadic Data," *Social Networks*, 10, 359–381.

Kuran, T. and W. H. Sandholm (2008): "Cultural Integration and its Discontents," *Review of Economic Studies*, 75, 201–228.

Kwok, H. H. (2013): "Identification Problems in Linear Social Interaction Models: A General Analysis Based on Matrix Spectral Decompositions," *mimeo*.

Lazear, E. P. (1999): "Culture and Language," *Journal of Political Economy*, 107, S95–S126.

Lee, L.-F. (2007): "Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects," *Journal of Econometrics*, 140, 333–374.

Lee, L.-F. and X. Liu (2010): "Identification and GMM Estimation of Social Interactions Models with Centrality," *Journal of Econometrics*, 159, 99–115.

Lee, S. H., P. Kim, and H. Jeong (2006): "Statistical Properties of Sampled Networks," *Physical Review E*, 73(1).

Leung, M. (2015): "Two-Step Estimation of Network-Formation Models with Incomplete Information," *Journal of Econometrics*, 188, 182–195.

Ligon, E., J. P. Thomas, and T. Worrall (2000): "Mutual Insurance, Individual Savings, and Limited Commitment," *Review of Economic Dynamics*, 3, 216–246.

——— (2002): "Informal Insurance Arrangements with Limited Commitment: Theory and Evidence from Village Economies," *Review of Economics Studies*, 69, 209–244.

Liu, X. (2013): "Estimation of a Local-aggregate Network Model with Sampled Networks," *Economics Letters*, 118, 243–246.

Liu, X., E. Patacchini, and Y. Zenou (2014a): "Endogenous Peer Effects: Local Aggregate or Local Average?" *Journal of Economic Behavior and Organization*, 103, 39–59.

Liu, X., E. Patacchini, Y. Zenou, and L.-F. Lee (2014b): "Criminal Networks: Who is the Key Player?" *mimeo*.

Long, S. and R. Schwartz (1987): "The Impact of IRS Audits on Taxpayer Compliance: A Field Experiment in Specific Deterrence," Paper presented at the Annual Meeting of the Law and Society Association, Washington DC.

Maciejovsky, B., E. Kirchler, and H. Schwarzenberger (2007): "Misperception of Chance and Loss Repair: On the Dynamics of Tax Compliance," *Journal of Economic Psychology*, 28, 678–691.

Manski, C. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, 531–542.

——— (2013): "Identification of Treatment Response with Social Interactions," *Econometrics Journal*, 16, S1–S23.

Mantel, N. (1967): "The Detection of Disease Clustering and a Generalised Regression Approach," *Cancer Research*, 27, 209–220.

Marmaros, D. and B. Sacerdote (2006): "How do Friendships Form," *Quarterly Journal of Economics*, 121(1), 79–119.

Matsuyama, K. (2002): "Explaining Diversity: Symmetry-Breaking in Complementarity Games," *American Economic Review*, 92, 241–246.

——— (2004): "Financial Market Globalization, Symmetry-Breaking, and Endogenous Inequality of Nations," *Econometrica*, 72, 853–884.

——— (2011): "Imperfect Credit Markets, Household Wealth Distribution, and Development," *Annual Review of Economics*, 3, 339–362.

Mayer, A. and S. L. Puller (2008): "The Old Boy (and Girl) Network: Social network formation on university Campuses," *Journal of Public Economics*, 92, 329–347.

Mazzocco, M. and S. Saini (2012): "Testing Efficient Risk Sharing with Heterogeneous Risk Preferences," *American Economic Review*, 102, 428–468.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001): "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 27, 415–444.

Méango, R. (2014): "International Student Migration: A Partial Identification Analysis," *mimeo.*

Mele, A. (2013): "A Structural Model of Segregation in Social Networks," *mimeo.*

Mihaly, K. (2009): "Do More Friends Mean Better Grades? Student Popularity and Academic Achievement," *RAND Working Papers*, WR-678.

Milán, P. (2016): "Network-Constrained Risk Sharing in Village Economies," *mimeo.*

Milgrom, P. and C. Shannon (1994): "Monotone Comparative Statics," *Econometrica*, 62, 157–180.

Mittone, L. (2006): "Dynamic Behaviour in Tax Evasion: An Experimental Approach," *Journal of Socio-Economics*, 35, 813–835.

Moffitt, R. (2001): "Policy Interventions, Low-level Equilibria, and Social Interactions," in *Social Dynamics*, ed. by S. Durlauf and H. P. Young, MIT Press, Cambridge, 45–82.

Montalvo, J. G. and M. Reynal-Querol (2005): "Ethnic Polarization, Potential Conflict, and Civil Wars," *American Economic Review*, 95, 796–816.

Morduch, J., S. Ravi, and J. Bauchet (2015): "Failure vs Displacement: Why an innovative anti-poverty program showed no net impact," *Journal of Development Economics*, 116, 1–16.

Morten, M. (2015): "Temporary Migration and Endogenous Risk Sharing in Village India," *mimeo.*

Munshi, K. and J. Myaux (2006): "Social Norms and the Fertility Transition," *Journal of Development Economics*, 80 (1), 1–38.

Munshi, K. and M. Rosenzweig (2016): "Networks and Misallocation: Insurance, Migration and the Rural-Urban Wage Gap," *American Economic Review*, 106, 46–98.

Murphy, K. M., A. Shleifer, and R. W. Vishny (1989): "Industrialization and the Big Push," *Journal of Political Economy*, 97, 1003–1026.

Nurkse, R. (1953): *Problems of Capital Formation in Underdeveloped Countries*, Oxford University Press.

Patacchini, E. and Y. Zenou (2012a): "Ethnic networks and employment outcomes," *Regional Science and Urban Economics*, 42, 938–949.

——— (2012b): "Juvenile Delinquency and Conformism," *Journal of Law, Economics and Organization*, 1, 1–31.

Patel, K. and F. Vella (2013): "Immigrant networks and their implications for occupational choice and wages," *Review of Economics and Statistics*, 95, 1249–1277.

Patnam, M. (2013): "Corporate Networks and Peer Effects in Firm Policies," *mimeo, ENSAE-CREST.*

de Paula, A. (2013): "Econometric Analysis of Games with Multiple Equilibria," *Annual Review of Economics*, 5, 107–131.

de Paula, A., S. Richards-Shubik, and E. Tamer (2014): "Identification of Preferences in Network Formation Games," .

Pavoni, N. and G. L. Violante (2007): "Optimal Welfare-To-Work Programs," *Review of Economic Studies*, 74, 283–318.

Phelan, C. and R. M. Townsend (1991): "Multi-Period, Information-Constrained Optima," *Review of Economic Studies*, 58, 853–881.

Popescul, A. and L. Ungar (2003): "Statistical relational learning for link prediction," *Proceedings of the Workshop on Learning Statistical Models from Relational Data at IJCAI-2003.*

Prescott, E. and R. M. Townsend (1984a): "General Competitive Analysis in an Economy with Private Information," *International Economic Review*, 25, 1–20.

——— (1984b): "Pareto Optima and Competitive Equilibria with Adverse Selection Moral Hazard," *Econometrica*, 52, 21–46.

Preston, C. J. (1973): "Generalised Gibbs states and Markov random fields," *Advances in Applied Probability*, 5, 242–261.

Quandt, R. E. (1960): "Tests of the Hypothesis that a Linear Regression Obeys Two Separate Regimes," *Journal of the American Statistical Association*, 55, 324–330.

Rasul, I. and D. Rogger (2015): "The Impact of Ethnic Diversity in Bureaucracies: Evidence from the Nigerian Civil Service," *American Economic Review*, 105, 457–461.

Robinson, P. M. (1988): "Root-n Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

Rosenbaum, P. and D. Rubin (1985): "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score," *American Statistician*, 39, 33–38.

Rosenstein-Rodan, P. N. (1943): "Problems of Industrialization of Eastern and Southeastern Europe," *Economic Journal*, 53, 202–211.

Rosenzweig, M. R. and K. I. Wolpin (1993): "Credit Market Constraints, Consumption Smoothing, and the Accumulation of Durable Production Assets in Low-Income Countries: Investment in Bullocks in India," *Journal of Political Economy*, 101, 223–244.

Rubin, D. (2001): "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation," *Health Services and Outcomes Research Methodology*, 2, 169–188.

Ruggles, S., J. T. Alexander, K. Genadek, R. Goeken, M. B. Schroeder, and M. Sobek (2010): *Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]*, Minnesota Population Center.

Sacerdote, B. (2011): "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" in *Handbook of the Economics of Education*, ed. by E. Hanushek, S. Machin, and L. Woessman, Elsevier, vol. 3.

Schutz, B. (2003): *Gravity from the Ground Up*, Cambridge Univ Press.

Shaban, R. A. (1987): "Testing Between Competing Models of Sharecropping," *Journal of Political Economy*, 95, 893–920.

Sheng, S. (2012): "Identification and Estimation of Network Formation Games," *mimeo*.

Sherman, S. (1973): "Markov Random Fields and Gibbs Random Fields," *Israel Journal of Mathematics*, 14, 92–103.

Slemrod, J. and C. Gillitzer (2013): *Tax Systems*, MIT Press.

Snijders, T. A. B. (2002): "Markov Chain Monte Carlo Estimation of Exponential Random Graph Models," *Journal of Social Structure*, 3, 1–40.

Spear, S. E. and S. Srivastava (1987): "On Repeated Moral Hazard with Discounting," *Review of Economic Studies*, 54, 599–617.

SPOLAORE, E. AND R. WACZIARG (2009): "The Diffusion of Development," *Quarterly Journal of Economics*, 124, 469–529.

STRAUSS, D. AND M. IKEDA (1990): "Pseudolikelihood Estimation for Social Networks," *Journal of the American Statistical Association*, 85, 204–212.

SUBRAMANIAN, S. AND A. DEATON (1996): "The Demand for Food and Calories," *Journal of Political Economy*, 104, 133–162.

TAUCHEN, H., A. WITTE, AND K. BERON (1993): "Tax Compliance: An Investigation Using Individual Taxpayer Compliance Measurement Program (TCMP) data," *Journal of Quantitative Criminology*, 9, 177–202.

THOMAS, J. AND T. WORRALL (2016): "Dynamic Relational Contracts under Complete Information," *mimeo.*

THOMPSON, S. K. (2006): "Adaptive Web Sampling," *Biometrics*, 62(4), 1224–1234.

TOPA, G. AND Y. ZENOU (2015): "Neighborhood and Network Effects," in *Handbook of Regional and Urban Economics*, ed. by G. Duranton, V. Henderson, and W. Strange, Elsevier, vol. 5A, chap. 9.

TOPKIS, D. M. (1978): "Minimizing a Submodular Function on a Lattice," *Operations Research*, 26, 305–321.

TOWNSEND, R. (1994): "Risk and Insurance in Village India," *Econometrica*, 62, 539–591.

UDRY, C. (1994): "Risk and Insurance in a Rural Credit Market: An Empirical Investigation in Northern Nigeria," *Review of Economic Studies*, 61, 495–526.

WALDINGER, F. (2010): "Quality Matters: The Expulsion of Professors and the Consequences for PhD Students Outcomes in Nazi Germany," *Journal of Political Economy*, 118 (4), 787–831.

——— (2012): "Peer Effects in Science - Evidence from the Dismissal of Scientists in Nazi Germany," *Review of Economic Studies*, 79 (2), 838–861.

WANG, D. J., X. SHI, D. MCFARLAND, AND J. LESKOVEC (2012): "Measurement error in network data: A re-classification," *Social Networks*, 34(4), 396–409.

WASSERMAN, S. AND P. PATTISON (1996): "Logit models and logistic regressions for Social Networks: I. An Introduction to Markov Graphs and p*," *Psychometrika*, 61, 401–425.

Williams, K. Y. and C. A. O'Reilly (1998): "Demography and Diversity in Organizations. A Review of 40 Years of Research," *Organizational Behavior*, 20, 77–140.

Young, H. P. (1993): "The Evolution of Conventions," *Econometrica*, 61, 57–84.