Essays in Credit Markets and Development Economics

by

Anil Jain

M.Phil. Economics, Trinity College, University of Cambridge (2009)
M.A. Economics, Trinity College, University of Cambridge (2008)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2014

© 2014 Anil Jain. All rights reserved.

The author hereby grants to MIT permission to reproduce and distribute publicly paper and electronic copies of this thesis document in whole or in part.

Signature redacted

Author .............................................

Department of Economics
May 15, 2014

Signature redacted

Certified by.

Abhijit Banerjee
Ford Foundation International Professor of Economics
Thesis Supervisor

Signature redacted

Certified by...

Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics
Thesis Supervisor

Signature redacted

Accepted by .......... ...................................

Michael Greenstone
3M Professor of Economics
Chairman, Departmental Committee on Graduate Studies
Abstract

Chapter 1 (co-authored with Ali Choudhary) exploits exogenous variation in the amount of public information available to banks about a firm to empirically evaluate the importance of adverse selection in the credit market. A 2006 reform introduced by the State Bank of Pakistan (SBP) reduced the amount of public information available to Pakistani banks about a firm's creditworthiness. Prior to 2006, the SBP published credit information not only about the firm in question but also (aggregate) credit information about the firm's group (where the group was defined as the set of all firms that shared one or more director with the firm in question). After the reform, the SBP stopped providing the aggregate group-level information. We propose a model with differentially informed banks and adverse selection, which generates predictions on how this reform is expected to affect a bank's willingness to lend. The model predicts that adverse selection leads less informed banks to reduce lending compared to more informed banks. We construct a measure for the amount of information each lender has about a firm's group using the set of firm-bank lending pairs prior to the reform. We empirically show those banks with private information about a firm lent relatively more to that firm than other, less-informed banks following the reform. Remarkably, this reduction in lending by less informed banks is true even for banks that had a pre-existing relationship with the firm, suggesting that the strength of prior relationships does not eliminate the problem of imperfect information.

Chapter 2 examines the provision of public goods in developing countries is a central challenge. This paper studies a model where each agent's effort provides heterogeneous benefits to the others, inducing a network of opportunities for favor-trading. We focus on a classical efficient benchmark – the Lindahl solution – that can be derived from a bargaining game. Does the optimistic assumption that agents use an efficient mechanism (rather than succumbing to the tragedy of the commons) imply incentives for efficient investment in the technology that is used to produce the public goods? To show that the answer is no in general, we give comparative statics of the Lindahl solution which have natural network interpretations. We then suggest some welfare-improving interventions.

In chapter 3 (co-authored with Robert Townsend) we present a tractable model of platform competition in a Walrasian equilibrium. Rochet and Tirole (2003) sparked a decade of extensive study on two-sided markets. However, the analysis of two-sided markets with multiple platforms has been largely ignored. We endogenize the size of each platform for different utility functions, different types of agents, and different levels of capital. Contrary to the prior literature, our economy is efficient – platforms internalize the network effects of adding more users by offering bundles which state both the number of users and the price to join the platform. Further, we show that the first and second welfare theorems are still able to be applied. Our model suggests how the equilibrium characterization of two-sided markets changes when we alter the cost structure or wealth of agents and subsequently we analyse the welfare implications of various placebo interventions.

Thesis Supervisor: Abhijit Banerjee
Title: Ford Foundation International Professor of Economics

Thesis Supervisor: Esther Duflo
Title: Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics
Contents

1 How Public Information Affects Asymmetrically Informed Lenders: Evidence from a Credit Registry Reform .................................................. 1
   1.1 Introduction ........................................................................... 11
   1.2 Literature Review .................................................................. 13
   1.3 Data and Institutional Background ....................................... 15
       1.3.1 Why did the State Bank of Pakistan alter the information available to lending institutions? .................................................. 16
       1.3.2 Other sources of firm information for a bank. ...................... 17
       1.3.3 Building firm links. ......................................................... 18
       1.3.4 What does the network of firms look like? ......................... 18
       1.3.5 The data. ...................................................................... 18
   1.4 Model .................................................................................. 21
       1.4.1 Setup............................................................................... 21
       1.4.2 Timing .......................................................................... 22
       1.4.3 The game. ..................................................................... 22
           1.4.3.1 The firm’s problem. ................................................. 22
           1.4.3.2 The informed (I) lender’s problem. ......................... 23
           1.4.3.3 The uninformed (N) lender’s problem. .................... 24
       1.4.4 Equilibrium ..................................................................... 24
       1.4.5 Mapping the model to the data. ....................................... 26
   1.5 The Effect of the Reform on a Firm’s Source of Credit .............. 26
       1.5.1 Econometric specification .............................................. 26
           1.5.1.1 Outcomes of interest .............................................. 28
       1.5.2 Results ........................................................................ 29
           1.5.2.1 Following the reform, loan sizes were relatively larger within informed banking relationships. 29
1.5.2.2 There were substantial differences in the effect of the policy on the credit
market depending on the strength of the interfirm relationship. 

1.5.2.3 The effect of the change in information was predominantly felt by small to
medium sized firms. 

1.5.2.4 The effect of the policy was largest on those firms who were overdue on a
loan in December 2004. 

1.5.2.5 Overdue rates are similar across informed and uninformed banking relation-
ships. 

1.6 The Effect of the Reform on a Firm's Access to Total Credit 

1.6.1 Econometric specification 

1.6.2 Results 

1.6.2.1 Firms with an informed lending relationship borrowed more after the reform. 

1.6.2.2 The borrowing capacity of the smallest firms were most affected
by the reform. The effect of the reform was greatest on the smallest firms' total borrowings. 

1.7 Conclusion 

1.8 Figures and Tables 

1.9 Appendix 

1.9.1 Proof of Proposition 1 

1.9.2 Proof of Proposition 2 

1.9.3 Proof of Proposition 3 

1.9.4 Proof of Proposition 4 

2 Efficient Public Good Provision in Networks: Revisiting the Lindahl Solution 

2.1 Introduction 

2.2 Model 

2.2.1 Assumptions on stage game utilities 

2.2.2 Lindahl equilibrium 

2.3 Individual Characteristics 

2.4 Network-wide Characteristics 

2.5 How the network affects the individual 

2.6 Conclusion 

2.7 Proofs 

2.7.1 Proof of Proposition 5 

2.7.2 Proof of Proposition 6 

2.7.3 Proof of Proposition 7 

2.7.4 Proof of Proposition 9 

2.7.5 Proof of Lemma 1
Acknowledgments

I owe a great debt of gratitude to many people for their advice, feedback, support, and kindness during the writing of this thesis.

First and foremost, I am grateful to three individuals who shaped my perspective and my economic reasoning. Colin McFall stimulated my interest in economics, challenged my reasoning and provided invaluable advice. Rupert Gatti focused my intuition and provided the tools with which to do a PhD. Abhijit Banerjee inspired me and finessed my ideas — his insights were always precise and able to isolate the main issue in a single sentence. Abhijit has always been concerned for my research and my personal development.

I would also like to thank my advisors Esther Duflo, Tavneet Suri, Benjamin Golub, Robert Gibbons and Robert Townsend. Tavneet in particular has shown great empathy during the last five years and has been incredibly generous with her time and always available to mentor and guide me. Ben willingly listens and works through my most inarticulate of ideas. He sets an example of excellence in scholarship and compassion I can only hope to aspire too.

I am privileged to have a great set of friends and classmates who have all helped my ability to write this PhD and my personal development. Geoff Wright has been a constant companion through happy days and frustrating ones. As my sports buddy, Geoff ensured a happier PhD even if this meant propping up my self-confidence by losing at squash — and also a healthier one, given my fondness for American cheesecake. All my friends and colleagues are too numerous too name, but a most heartfelt thank you to Emily Breza, Ali Choudhary, Rebecca Dizon-Ross, Ben Feigenberg, Gary King, Conrad Miller, Jennifer Peck, Adam Sacarny, Indrani Saran, Annalisa Scogamiglio, Ashish Shenoy and Pawel Zaczkowski.

Finally, I thank my family for their patience, generosity and love during the PhD. Specifically, I must thank my grandfather, Dada. Dada has patiently taught me and instilled a desire to learn — although I do wonder why it had to start on early Sunday mornings! Nidhi, for always listening to my latest obstacle and her willingness to learn about the "dismal science." I thank Aaron, Dillon and Isha for their cheeky antics which gave me a timely boost to complete the PhD. I thank my amazing brothers, Sunil and Rajan, for always believing in me and their invaluable advice. As always, I owe a great deal to my wonderful parents.
List of Tables

1.1 Summary statistics for the entire dataset in December 2004 .................. 50
1.2 Summary statistics for group firms in December 2004 ......................... 50
1.3 Summary statistics for group firms who have both an informed and uninformed loan in December 2004 ............................................ 51
1.4 The effect of the policy on the likelihood of a loan being renewed between informed and uninformed lending relationships .................................... 52
1.5 The effect of the policy on the loan sizes between informed and uninformed lending relationships .................................................. 53
1.6 The effect of the policy on the loan sizes between informed and uninformed lending relationships restricting to only those loans that lasted until June 2008 or December 2008. 54
1.7 The effect of the policy on the loan sizes between informed and uninformed lending relationships - censoring at baseline ............................ 55
1.8 The effect of greater shareholding on the effect of the policy on loan sizes and likelihood of renewing the loan ........................................ 56
1.9 The effect of the greater number of firms borrowing from the same institution on loan sizes and likelihood of renewing the loan .................. 57
1.10 The effect of the policy by decile of firm size ..................................... 58
1.11 The effect of the policy by decile of group size .................................. 59
1.12 Firms who were overdue at baseline were more likely to renew their informed loan ................................................................. 60
1.13 The difference in the likelihood of a loan going overdue ......................... 61
1.14 Difference in total loan sizes between firms who have an informed lender and those who do not .......................... 62
1.15 The difference in total loan borrowings by firm decile and whether they had an informed lending relationship .......................... 63
3.1 Equilibrium platforms and user utility for Example 1 .......................... 124
3.2 Equilibrium platforms and user utility for Example 2 .......................... 125
List of Figures

1-1 The network of firm connections ......................................................... 19
1-2 Extensive form of the game. ............................................................. 23
1-3 Identification Strategy: comparing loan details for the same firm borrowing from two different banks. ............................................................. 27
1-4 Multiple overlapping groups ............................................................. 36
1-5 The network of firm connections - using a cutoff of 25% equity .................. 37
1-6 The number of connections for each firm ........................................... 38
1-7 The size of the components ............................................................... 38
1-8 The histogram of log total firm borrowings ......................................... 39
1-9 The histogram of log loan size ........................................................... 39
1-10 The histogram of log loan size for informed and uninformed banking relationships ............................................................. 40
1-11 The histogram of total firm borrowings by lending relationship. ............... 41
1-12 The monthly coefficients for the difference in likelihood of a loan being renewed ............................................................. 42
1-13 The monthly coefficients for the difference in log loan size ....................... 43
1-14 The effect of the policy on the likelihood of the loan being renewed by different share thresholds. 44
1-15 The effect of the policy on the likelihood of a loan being renewed by different number of firm’s who borrow from the same bank. ............................................................. 45
1-16 The effect of the policy on the likelihood of a loan being renewed by decile of firm size. ............................................................. 46
1-17 The effect of the policy on the likelihood of a loan being renewed by decile of group size. ............................................................. 47
1-18 The monthly coefficients for the difference in total borrowings between borrowers with informed lending relationships and those who do not. ............................................................. 48
1-19 The effect of the reform on a firm’s total borrowings by decile of firm borrowings in December 2004. ............................................................. 49
1-20 Credit Worthiness Report from 2004 .................................................... 72
1-21 Credit Worthiness Report from 2012 .................................................... 73

2-1 The benefit matrix depicted as a network of benefit flows. The thickness of the arrows represents the magnitude of the marginal benefits. ............................................................. 81
2-5 Agents A and C are unconnected clones in this matrix, since they receive the same marginal benefits from the rest of the network and confer no benefits to each other .......... 105
2-6 Agents A and C are connected clones in this matrix, since they receive the same marginal benefits from the rest of the network and confer equal benefits to each other .......... 106
2-7 Agents A and C are identical clones in this matrix, since they receive and confer the same marginal benefits from and to the rest of the network. .................. 107
2-2 How costs, efforts and utilities are related for unconnected clones i and j if \( c_i > c_j \) and elasticity is decreasing in \( a \) .......................... 108
2-3 A stylised benefit flow through the network. The arrows demonstrate who confers benefits on whom. .......................... 109
2-4 The flow of a cost increase on person G through the network .......................... 109
3-1 Redistributing wealth within and across type .................................. 127
3-2 How does the utility for each subtype change as we alter the Pareto weight for subtype B,2 127
3-3 How does utility change as we alter the platform’s cost function .......................... 129
3-4 How does the utility for each subtype change as we alter the fixed cost of building a platform 130
3-5 Redistributing wealth within and across type – the effects on all agents’ utilities 134
Chapter 1

How Public Information Affects
Asymmetrically Informed Lenders:
Evidence from a Credit Registry Reform
1.1 Introduction

Credit markets are crucial to economic growth but asymmetric information may hinder their effectiveness. Jaffee and Russell [1976] and Stiglitz and Weiss [1981] theoretically show that adverse selection is a factor that can substantially constrain the effectiveness of credit markets, yet the extent to which this factor is a problem in credit markets remains largely unknown. It is possible, for example, that lenders face other binding constraints (such as those due to moral hazard) that make adverse selection less relevant.

The key empirical challenge to isolating the importance of adverse selection in the real world is that the variation in the individual observations is unlikely to be exogenous. For example, lending could drive the asymmetry of information rather than the reverse: one bank may know more about a firm than another bank simply because it has lent to that firm in the past.

An innovative recent paper by Karlan and Zinman [2009] on adverse selection (and moral hazard) in credit markets tackles one aspect of this issue. Karlan et. al examine if exogenously lowering the interest rate changes selection of the borrower pool in ways that the lender cannot observe. While they are able to highlight borrower behavior, they are still unable to observe changes in lender behavior in the presence of information asymmetries. In particular, they are not able to demonstrate that lenders react to these changing unobservables by altering the terms they offer – since the set of contracts offered by the lenders is fixed – which is the key mechanism behind credit market failure in the theoretical literature.

The purpose of this paper is to show that banks do change their lending behavior in reaction to a change in the distribution of information about a particular firm. In particular, among a group of differentially informed lenders, we ask: What is the effect of reducing public information about a borrower's creditworthiness?

A regulatory reform by the State Bank of Pakistan (SBP) in April 2006 offers a unique opportunity to answer this question. The reform exogenously reduced the amount of public information available to lenders about a firm's creditworthiness and did so in a way that varied across firm-lender pairs. Specifically, the change limited a lender's capacity to procure information about a firm's relationships to other firms.

Until April 2006, the SBP had supplemented credit information about prospective borrowers with information about that firm's "group." This group was defined as all other firms which shared at least one director with the borrowing firm.¹ But in April 2006, the SBP stopped defining a firm's group and in doing so stopped providing group-specific information (see section 1.3.1 for more detail).

Lenders value such information about the credit of other firms in a borrower’s network because assets and profits may be transferred within a group of firms – especially if one firm is in financial difficulty. For instance, when Lehman Brothers Holdings Inc. filed for bankruptcy, an unscheduled transfer of $8 billion occurred from the European operations (Lehman Brothers International) to the US operations.²

¹Firms within a group have complex interfirm relationships which subsequently have important economic implications. There is both theoretical and empirical evidence that interfirm relationships may be a mechanism for tax reduction (Desai and Dharmapala [2009]), tunneling (Bertrand et al. [2002]), risk sharing (Khwaja and Mian [2005a]), the efficient working of organizations (Williamson [1975, 1981]) and internal capital markets (Stein [1997], Almeida and Wolfenzon [2006], Gopalan et al. [2007]).

firm’s bankruptcy led to $38 billion in claims among the various arms of Lehman Brothers and took over three years of litigation to settle.\(^3\)

We take advantage of the natural experiment the SBP reform generated: We use a difference-in-difference methodology to estimate the causal effect of the reduction in public information on a bank’s willingness to lend.\(^4\) In particular, we exploit variation in the impact of the policy across firm-bank pairs generated by each bank’s other lending relationships. Suppose a firm borrowed from two different banks in Dec 2004. Following the reform, we predict that one bank— the informed bank— could have more information about the firm’s actual liabilities, if other members of the firm’s group also borrow from that bank. We then compare how the loans that the firm receives from the informed and the uninformed banks changes after the reform. In addition, we examine whether the reform affected a firm’s ability to access total credit by comparing loan amounts received by firms who had informed lenders, and those who did not, before and after the policy reform.

Our main result shows that banks with private information about other firms in a firm’s group lent more to that firm than other, less-informed lenders did, after the reform, both on an extensive margin (5.4 percent more likely to renew the firm’s loan) and on an intensive margin (larger loans). This is the primary evidence that the reduction in public information amplified the problem of adverse selection in the credit market.

Second, those firms that borrowed from informed lenders were likely to borrow 11-14% more than those firms who did not have access to informed lenders. In other words, both the level of credit and its source were affected.

Third, using the distribution of firm-bank pairs and the strength of interfirm relations we construct a measure for the quality of private information each bank possessed about a firm’s group prior to the reform. Following the reform, those lenders with greater private information about a firm’s group were more likely to renew a firm’s loan.\(^5\) This further supports the claim that a greater information differential between lenders leads to a greater disparity in the likelihood of a loan being renewed.

Fourth, there is substantial heterogeneity in the measured effect across firms: The smaller firms in a group were the most disadvantaged from the information change. There was no effect on the largest firms in a group, which suggests either that information about these borrowers was already pervasive, or that the relative cost of procuring the information was lower.

Fifth, those firms which had negative information in their credit reports in December 2004 were the most affected by the change in information reporting. They were 18% more likely to renew their loan at an informed lender. Those firms with a poor credit history are likely to be the most risky for a bank, making public information all the more important for these firms. A firm may have overdue loans, but if the rest of


\(^4\) For brevity, in this paper we use “bank” to denote any financial institution.

\(^5\) Since the strength of interfirm ties vary, we can construct various measures for the quality of private information a bank may possess about a firm dependent on who the bank lends to within that firm’s group, and the relative strength of those interfirm relationships.
the group is prompt in their repayment, it may signal a sufficiently creditworthy borrower. However, if both the firm and the group are overdue, this could signal wider systemic issues in the firm's creditworthiness.

The rest of the paper proceeds as follows: Section 1.2 reviews the relevant literature on credit markets, and section 1.3 describes the institutional background, and explains the SBP's reasoning in instituting the change in lending policy. Section 1.4 presents a stylized model to explain the results and shows how information asymmetries between lenders can be important. Section 1.5 outlines our econometric framework for analysing the importance of information asymmetries across lenders and details our results. Section 1.6 outlines the effects of the reform on a firm's ability to procure credit and section 1.7 presents a summary and concluding remarks for future research.

1.2 Literature Review

This paper's main contribution is to empirically assess the impact of adverse selection in credit markets with differentially informed lenders. There is a long theoretical literature describing how asymmetric information in credit markets causes lenders to alter what contracts they offer. Compared to an environment with full information, this can lead to the misallocation of capital (Jaffee and Russell [1976], Stiglitz and Weiss [1981, 1983], De Meza and Webb [1987]), even to the complete unraveling of the credit market (Akerlof [1970]). Yet, the empirical evidence for adverse selection is relatively limited.

Karlan and Zinman's novel experiment separately identified adverse selection and moral hazard in microcredit but only found weak evidence for adverse selection effects. It is unclear if the small effects they find are due to the small loans in the microcredit sector, or because the experiment only used individuals who borrowed from the lender previously, or if adverse selection is not a key problem in lending. Ausubel [1999] uses a randomized trial which varied the contractual terms for a pre-approved credit card for 600,000 individuals. Contrary to Karlan and Zinman, Ausubel finds large effects of adverse selection in the credit card industry. To the best of our knowledge our paper is the first paper that examines adverse selection in corporate lending. Filling the gap in the literature, we examine the effects of adverse selection in a market with much larger loans and in a setting where we expect the effect of adverse selection to be very different. Additionally, we examine how lenders react to the change in a set of observable characteristics – which is the key mechanism leading to credit market failure in the theoretical literature.

The classic literature on information problems in credit markets highlights the effect of asymmetric information on symmetrically informed lenders. Recent papers examine the theoretical and empirical effect of information asymmetries across lenders. Stroebel [2013] theoretically models the interest rates a borrower receives when mortgage lenders are differentially informed about the borrower's collateral. He shows empirically that the return is higher for more informed lenders. Moreover, due to the winner's curse, less informed lenders charge higher interest rates when competing against more informed lenders.

Consistent with Stroebel's results we find the more informed lenders to be better off. Yet our environment
has richer intertemporal variation that allows us to address different issues. First, our model endogenizes a firm's total borrowings and we empirically test whether the more informed lenders offer larger loan sizes than other lenders. Second, while Stroebel relies on pre-existing differences in whether the lender also built the property, we identify the effect of differential information exploiting an exogenous change in public information reporting. Third, because firms in our setting have multiple loans, we can estimate the effect of differential information across lenders for the same firm.

Public information performs three key roles in credit markets: (i) it reduces information asymmetries between the borrower and the lender, (ii) reduces information asymmetries between lenders and (iii) reduces a borrower's incentive to default. However, public information does have drawbacks. First, it may reduce a lender's incentive to procure information if the information will be later revealed by a credit registry. Second, public information may be imprecise or noisy causing excessive volatility in the observed public information (Morris and Shin [2002]).

Public information is often considered a substitute for private information, or a leveler of the playing field, but this is not always true. For example, consider a scenario where two banks can lend to the same firm but only one bank knows the identity of all the directors of this firm. Providing a directory of director creditworthiness to both firms, would only benefit the firm who is privately informed about the firm's composition of directors. Therefore, it is possible for public information to be a complement to private information.

Hertzberg et al. [2011] empirically demonstrate how merely the prospect of publicly announcing a bank's private credit rating about a firm can lead to strategic effects by lenders, and the subsequent reduction of credit. Lenders strategically reduced lending to borrowers who they had previously labeled as a poor credit. They argue that lenders reduced lending since the public information revelation would lead to other lenders reappraising their lending terms and potentially reducing their credit lines to the borrower, causing the borrower financial stress.

Our paper is similar to Hertzberg et al. in that we analyze the effects of altering public information in an environment with multiple lenders. Since the change in public information is common to all lenders, lenders must take into account how the change in public information affects their willingness to lend and how the change affects their competitors' willingness to lend. Although both Hertzberg et al. and our paper examine changes in public information, the form of the public information is quite different. Hertzberg et al. analyze a reform which publicly released a lender's appraisal of a firm's creditworthiness whereas in the reform that we analyze, the credit registry altered the information available about the firm's ongoing credit history. Hertzberg analyzes a reform where lenders respond to the expectation of a bad credit rating becoming public knowledge limiting the firm's capacity to procure credit; by contrast, we study a reform where the SBP removed the ongoing capacity to monitor a firm.

This paper contributes to the literature on relationship banking (Petersen and Rajan [1994], Berger and
Udell [1995], Degryse and Van Cayseele [2000]). The prior literature has highlighted the potential for banking relationships to overcome the problem of information problems. Our paper emphasizes that the strength of these prior relationships does not completely eliminate the problem of imperfect information.

There is a nascent literature examining how variation in private information affects loan officers’ decisions. Hertzberg et al. [2010], Cole et al. [2012], Paravisini and Schoar [2012] all examine how loan officers evaluate whether they should offer loans, and at what terms, under different information structures. This is similar to our paper’s environment except the loan officers must consider not only what information they can observe, but what other information may be available to other lenders.

Our paper examines the importance of interfirm relationships in a firm’s ability to procure credit. Khwaja and Mian [2005a, 2008], Khwaja et al. [2011] show two key and related points: first, interfirm relationships allow insurance against idiosyncratic shocks; and second, politically connected firms are able to garner political favors for the entire set of related group firms. There is similar evidence in developed economies, Haselmann et al. [2013] show how social connections between banks and firms can facilitate more favorable lending terms for the firm. This paper emphasizes a mechanism for improved access to credit that is distinct from the intragroup lending channels studied in previous work. In particular, we suggest that the interfirm relationships may help firms procure loans since a lender considers the creditworthiness of both the firm and the firm’s group. Therefore, being part of a group with a perfect credit history can facilitate a firm’s access to credit. To our knowledge, this is the first paper that shows that information about a firm’s group has implications on a firm’s ability to borrow.

1.3 Data and Institutional Background

Bank lending is the primary source of formal funding in Pakistan. For instance in 2002, Pakistan’s main stock exchange, the Karachi Stock Exchange, only had a market capitalization of 16% of GDP, which is much smaller than the more mature NYSE which had a market capitalization of 92% of GDP (Khwaja and Mian [2005b]). The small size of public equity and debt issuance is a common feature amongst emerging stock markets. In part this is due to institutional failings, for example Khwaja and Mian [2005b] show that brokers in Pakistan manipulated public stock prices through “pump and dump” schemes to earn rates of return 50-90 percentage points higher than outside investors.

The data comes from the State Bank of Pakistan’s electronic Credit Information Bureau (e-CIB), which legally requires all banks and lending institutions to submit data on all borrowing firms with outstanding loan amounts greater than 500,000 Pakistani Rupees (equivalent to about $8500 in 2004). Some of the information collected by the SBP was passed back to the banks to facilitate lending. The information was provided through “credit worthiness reports.” A sample report is shown in figure 1-20 in the Appendix.

The creditworthiness report provided information about the firm’s total borrowing, overdue loans, ongoing

---

6This limit was removed in April 2006, and all loans (regardless of size) were required to be reported to the e-CIB.
litigation against the firm, and amounts written off in the last five years. In addition, the central bank provided information on a firm’s group borrowing, that is, the total borrowing by all firms which shared a mutual director, and whether the group had any amount overdue. However, the central bank did not provide information on which firms were in the firm’s group – it was solely the aggregate group’s credit history. Therefore, the individual banks would have different information about who – and who was not – in the firm’s group.

Financial institutions use the credit reporting system as an initial appraisal and to monitor the ongoing creditworthiness of a firm. In an interview a former loan officer remarked: “The eCIB is used to verify credit history and monitor exposure, both during and after approval [of the loan].” Additionally, the banks’ written notes on a firm’s credit worthiness mention that credit reports were checked.

In April 2006, the central bank instituted new policies about the amount of information they would provide about a firm through the credit reports. Specifically, the central bank reduced the amount of information they would provide about a firm’s “group.” An example of the new report is shown in Figure 1-21 in the Appendix. The main difference between the reports is that the key terms detailing the group’s outstanding loans were removed in 2006, hampering the lending institution’s ability to conduct due diligence on the firm and group. A bank has limited capacity to recoup its funds on non-performing loans – emphasizing the need to conduct sufficient due diligence prior to offering a loan. In 2005, Pakistani banks recouped a mere 14.1% of the value of the loans which were classified as Non-Performing Assets (Ministry of Finance [2006]).

Further there is important information in the creditworthiness of a firm’s group. The probability of any corporate loan in our dataset being overdue in December 2004 was 2.6%. However, conditional on any firm in the firm’s group being overdue on a loan, that firm was 6.2 percentage points more likely to be overdue on his loan. After controlling for observables such as loan size, total borrowings, bank fixed effects, those firms who had an overdue firm in their group were 6.8 percentage points more likely to be overdue on their loans than firms with perfectly creditworthy partners. This highlights the importance of understanding the creditworthiness the firm’s group in assessing the creditworthiness of the firm.

This paper argues that this reduction in information had a major impact on bank lending decisions. It led banks to lend more to those firms about which they had private information.

1.3.1 Why did the State Bank of Pakistan alter the information available to lending institutions?

Prior to April 2006, the State Bank of Pakistan defined a firm’s group as those firms which shared a director. However, due to firms lobbying the SBP, this definition of groups was altered, because the SBP believed that the definition of a group was too broad and not an adequate measure of control.

The following quote comes from the minutes of an interview with Mr. Inayat Hussein, head of the Banking and Regulation at the SBP — the group responsible for designing and implementing the prudential
regulation titled “Criteria of Grouping Companies for the CIB Report,” and outlines the motivations behind the change:

“The 2004 criteria of grouping for the purpose of CIB reports resulted in tying together some companies/firms/individuals, otherwise historically financially sound, with defaulters. This happened due to common directorship definition which also includes nominee directors who have little influence on the management of the firm. Hence, the SBP management decided to recognize and differentiate between those controlling shares in a group from common directors having no influence over the management of the group in question. Notwithstanding, SBP also appreciated financial institution’s need to take informative decisions if complete information of allied companies is not provided in the credit worthiness reports.”

These views were represented in the confidential internal minutes of a meeting of the SBP in May 2004. The State Bank’s intention was to provide a new definition for groups – one which would offer a better measure of control than the previous definition. However, the SBP delegated the responsibility for constructing the group relations to the banks, stating that “the onus for correct formation of the group as per definition given in the Prudential Regulations will be on Banks” (State Bank of Pakistan, 2004 Prudential Regulation). So the banks – the main beneficiaries of receiving information about a firm’s group – were the ones expected to provide the information to the SBP.

The State Bank specifically warned against reporting too large a group, stating that: “Banks are advised to be very careful while reporting the names of group entities in the CIB data. In case any party disputes the group relationship, the reporting Bank should be able to defend its position with documentary evidence” (State Bank of Pakistan, 2004 Prudential Regulation), further reducing the bank’s incentive to report a firm’s group to the SBP.

Ultimately, the regulation led banks to report almost no group information to the central bank. Since the SBP was no longer constructing group liabilities, and banks were not reporting group entities, banks were left to construct their own groups. Therefore, they were forced to make their own definition of groups and conduct their own due diligence. A current credit officer remarked: “With the new e-CIB system, you have do your own intel and also consider past information for firms” and in the new system “the full group information was not declared – overall I preferred the older system.”

1.3.2 Other sources of firm information for a bank.

Lenders access multiple different sources of information about a borrower prior to offering a loan. They collect information from the borrower using a standard form – the Basic Borrower Fact Sheet (BBFS), the firm’s accounts, and some information held by the Securities and Exchange Commission of Pakistan (SECP). Lenders also consult other people in the banking industry. The information reported in the BBFS and SECP

7It was surprising to us that loan officers mentioned talking to other banks to learn more about a client’s credit worthiness. A further way to procure knowledge of a firm’s group is via the SBP, lenders can query which firms a director is part of, however,
detail who are the firm’s directors and the amount of shareholding each director holds.

Thus if a lender has a loan application from two firms, who share a mutual director, using the information from the BBFS and the SECP, the lender can determine the existence of an interfirm relationship and the strength of the relationship (as proxied via the shareholdings). Further, we suspect if a bank lent to two members of the same group, the bank would have better information about that firm’s group.

1.3.3 Building firm links.

The State Bank of Pakistan collects information on all directors of a firm that borrows from a bank. This includes the director’s name, father’s name, a common identifier, shareholding and home address. At the baseline of December 2004 we have details of 174,244 director relationships and 97,449 firms.

1.3.4 What does the network of firms look like?

Figure 1-1 shows the network of firm connections across the entire set of borrowing firms in Pakistan in December 2004. A connection between two firms is shown if both firms have at least one mutual director.

The most visually striking aspect of the network is the huge dense network of firm connections in the center of the figure. The largest component is a total of 2395 firms in December 2004.

Figure 1-5 in section 1.8 shows what the network of firms looks like if we restrict attention to interfirm relationships where a mutual director owns at least 25% of the firm’s equity. Figures 1-6 and 1-7 show the distribution of the number of connections each firm has and the distribution of component size.

1.3.5 The data.

The loan level data comes from the SBP’s credit registry. The data includes information on all directors for each firm and the amount of equity each director holds. Furthermore, the data set in December 2004 included data on a total of 97,449 firms. 11,395 are corporate firms and 86,053 are sole proprietorship or small firm loans. In April 2006, as part of an overhaul of the credit registry database, two separate registries were created: (1) corporate firms and (2) all consumer. The data on sole proprietor loans was moved into the consumer database and the data on corporate firms was kept within the corporate database. Simultaneously, some of the original corporate firms were placed in the consumer dataset – therefore, to maintain a consistent dataset of firms throughout the period, we use the set of firms which were ex-post defined to be corporate firms and maintained within the credit registry. Further, we exclude publicly listed

---

"...our interviews with loan officers suggest that the knowledge of this facility is not apparent."

"The e-CIB is missing the common identifier for a total of 19,473 of these director relationships and these are omitted from our sample. Since the common identifier was also used to compile which firms were in groups, our definition of which firms were in groups should not be affected.

"A component is defined as, the set of nodes such that every node within the component has either a direct or indirect connection to every other node within the component.

"It should be noted, we are only building the component of firms using the set of firms currently borrowing. Therefore, the giant component should really be seen as a underestimate of the true size of the largest component since there are possibly non-borrowing firms which would link other firms into the giant component."
Only firms with at least one connection are shown in the graph above. The circles and edges in the picture correspond to firms and connections between firms respectively, where a connection is defined if both firms have at least one mutual director. Those firms with more connections are slightly darker and larger in the diagram.
firms due to our expectation that there should not be effected by the reform and a lack of sufficient sample size.

The SBP defined a firm's group as the set of firms with a common director in the entire borrowing data set of 97,449 firms. This paper follows the same definition of a group, using the entire database of 97,449 firms in December 2004.

The data set stretches from December 2004 to December 2008, however there are data validation concerns immediately following the regulation change. Therefore the months from April to August 2006 are dropped from the data set. The new credit registry required more information on borrowers to be uploaded by banks (for SBP's role of supervising the banking system, not for the purposes of the credit registry\(^{11}\) and subsequently required banks to upload the data in a new format which led to some initial teething problems. In all our main within-firm regressions we include a bank interacted with time fixed effect, therefore, if there was any time-varying bank-specific measurement error (which was common across all the bank's loans), then this fixed effect should alleviate this concern.

In addition, one of the banks, Union Bank, was taken over in late 2006. Since data reporting was poor during the takeover, that bank is omitted from all specifications.

The paper examines “funded” loan balances. A funded loan is a credit which is backed by the bank, such that in the case of defaults, the bank must attempt to recover the loan directly from that firm or person. The bank is the residual claimant on the loan. On the other hand, non-funded loans are backed with a letter of credit or a personal guarantee — if the borrower defaults, the bank can repossess funds from the guarantor.

The majority of lending are working capital balances, which are normally renewed every 12 months. They are similar to an overdraft facility, where firms are able to borrow more or less at any stage subject to their total borrowing limit. Unfortunately, data on the borrowing limit was not collected by the SBP prior to April 2006, therefore, the paper restricts attention to the total amount borrowed in any month.

The data set details a loan to be overdue in any particular month if the loan amount was overdue for more than 90 days.

There are a total of 94 banks which offer loans to corporate firms in 2004, but the sample is restricted to banks offering at least 50 corporate loans in December 2004. This leaves a final data set of 55 banks and 96.4% of the original data.

The sample is restricted to both harmonize the data set and for tractability. Many of the specifications use three high dimensional fixed effects: ‘firm interacted with date’, ‘firm interacted with bank’ and ‘bank interacted with date’. Therefore, by removing the very smallest banks, the computation is greatly sped up.

All variables which include nominal amounts have been discounted to December 2004 prices using the official Pakistani CPI index published by the State Bank of Pakistan.

Tables 1.1 and 1.2 present summary statistics on the entire dataset as well as showing more detailed statis-

\(^{11}\)It is important that the extra information being collected in 2006, is solely for purpose of SBP’s role of supervising the credit market since if the credit registry was displaying more information than it was in 2004, this could potentially conflate some of the observed effects of the regulation change.
tics on group firms. Figures 1-8 and 1-9 show the distribution of total firm borrowings and the distribution of loan sizes for group firms.

1.4 Model

The model demonstrates how lenders who are differentially informed about a firm’s creditworthiness decide whom to lend to, and how much to lend, after assessing the creditworthiness of borrowing firms and what information other lenders may have.

The model aims to show how lending patterns change between informed and uninformed lenders as we alter the composition of the borrower pool, and as we alter the informational differential between lenders.

1.4.1 Setup

There are three players in the model: two lenders and a single firm. There are two different types of firms, a high type, \( H \), and a low type, \( L \), which vary in the probability of repayment. The probability of the low type is \( \gamma \) and the probability of the high type is \( (1 - \gamma) \).

Both lenders have the same cost of capital \( (\rho) \), however one of the lenders (\( I \)) - the informed lender - is more informed of the firm's type than the other (\( N \)) - the uninformed lender.

The firm has no outside source of funds, no collateral and limited liability. The firm is able to undertake a project such that if a firm of type \( i \) invests \( k \) in the project, the firm’s output \( Y_i(k) \) is:

\[
Y_i(k) = \begin{cases} 
Ak & \text{with probability } \frac{X_L}{1+k} \\
0 & \text{else}
\end{cases}
\]

Where \( X_L < X_H < 1 \), therefore the high type is more likely to have a successful project conditional on the amount borrowed, \( k \). The project’s expected output, \( E[Y_i(k)] \) is increasing and concave in capital, so the expected return \( E[Y_i(k)/k] \) is decreasing in capital.

For simplicity, we shall assume that the interest rate \( (R) \) is exogenously fixed at a rate greater than the cost of capital \( (\rho) \) and lower than the return of the project \( (A) \) if successful:

\[
\rho < R < A
\]

This is not an innocuous assumption since it stops competition over interest rates. Yet it seems empirically plausible as Khwaja and Mian [2008] demonstrate for Pakistani corporate firms and Petersen and Rajan [1994] for small business firms in the US, interest rates are not responsive to changes in lending costs or information respectively. Both papers find large effects on the amount of credit each bank is willing to offer.

To ensure there is the prospect of there being “lemons” in the model we assume:
Inequality (1.1) states the maximum expected return from the low (high) type borrower is lower (greater) than the cost of capital for the lender. Therefore inequality (1.1), states that if the lenders knew the firm was a low-type, it would not be profitable to offer that firm any loans.

Further, for ease of exposition in the proofs\(^{12}\) we assume that \(\frac{RX_L}{\rho} > \frac{1}{2}\) and we assume that \(\frac{RX_H}{\rho} < 4\).

We assume a lender can only make non-negative loan offers, \(k_j \geq 0\).

The firm has limited liability, no outside wealth and no collateral. Further, the firm cannot strategically default on a loan – this could be due to legal requirements or the lender can repossess the firm’s output. We define a firm default at lender \(i\), \(D^i\), if a firm received a loan and the project was unsuccessful.

For simplicity we have assumed that both the informed lender and the uninformed lender know that there exists one lender of each type. In the context of the empirical setting this is not so clear. Banks do not know what information other banks do and do not possess.

1.4.2 Timing

1. Nature chooses the firm’s type \(X_i\).
2. The informed lender observes \(X_i\).
3. The informed and uninformed lenders make simultaneous bids \((k_j)\) over how much to lend at an interest rate \(R\).
4. The firm accepts none, one, or both of the loan offers.
5. The project is successful, or not, and payoffs are assigned.

Figure 1-2 outlines the extensive form of the game.

1.4.3 The game.

We will consider the Perfect Bayesian Equilibrium of the game.

To solve for the equilibrium of the game, we will use backward induction. Firstly we solve for the contract offers the firm will accept. Second, conditional on the firm’s strategy, we solve for the optimal loan offer by each of the lenders.

1.4.3.1 The firm’s problem.

The firm wants to maximise its expected utility, which takes the form:

\(^{12}\)These bounds on the set of parameter values are sufficient conditions which greater simplify the proofs for certain boundary cases (when the uninformed lender does not enter).
Figure 1-2: Extensive form of the game.

\[ U_i(k, R) = \Pr(\text{success}) \times (Ak - Rk) \]
\[ = \frac{X_i}{1 + k} (A - R)k \]

Where \( k \) is the firm's total borrowings. This specific utility function has some key advantages, which will greatly simplify the model. The firm’s utility \( U_i(k, R) \) exhibits strictly increasing returns in capital for all interest rates below the productivity parameter, \( A \).

Therefore, the firm’s weakly dominant strategy will be to accept all loan offers.

1.4.3.2 The informed (I) lender’s problem.

Having solved the firm’s problem, we know the firm will accept all loan offers. The informed lender’s problem becomes:
\[ \pi_i^j(k_i^j, k_N, X_i) = \max_{k_i^j} \Pr(\text{success}) \times (k_i^j) - \rho k_i^j \]
\[ = \frac{X_i}{1 + k_i^j + k_N} - R - \rho \] \[ k_i^j \geq 0 \]

Where \( k_i^j \) is the informed lender's loan offer to the firm of type \( i \) and \( k_N \) is the uninformed lender's loan offer. The informed lender's problem is to maximise the expected return from lending to a borrower (conditional on the uninformed lender's loan amount) minus the cost of lending.

Recalling the assumption that the cost of capital is greater than the maximum repayment from the low type (\( \rho > X_L R \)). In this case, the informed lender will not lend to the low-type firm, since the expected profit from any non-zero capital offer is negative:

\[ \pi_i^j(k_i^j, k_N, X) < 0 \ \forall k_i^j > 0 \]

Therefore, the informed lender will make the offer \( k_i^L = 0 \) in equilibrium.

1.4.3.3 The uninformed (N) lender's problem.

\[ E(\pi_N(k_I, k_N, X)) = \max_{k_N} E_X \{ \Pr(\text{success}) \times (Rk_N) - \rho k_N \} \]
\[ = E_X \left\{ \frac{X_i}{1 + k_i^j(X_i) + k_N} - R - \rho \right\} \]

\[ E(\pi(k_I, k_N, X)) \geq 0 \]
\[ k_N \geq 0 \]

The uninformed lender's problem is similar to the informed lender's problem. However the uninformed lender does not observe the firm's type, so the lender must maximise over the expectation of the firm's type. It should be noted that the uninformed lender must also consider that the informed lender's loan offer will be a function of the firm's type. In particular, the uninformed lender will face more competition on the high-type firms than the low-type.

1.4.4 Equilibrium

Proposition 1. A Perfect Bayesian equilibrium of the game is:
(informed lender) \( k^f = 0, k^H = k^H* > 0 \)

(uninformed lender) \( k_N^* \geq 0 \)

(firm) \( s^f = (\text{Acc, Acc}) \)

\[ \text{Proof. In the appendix.} \]

Therefore, the informed lender will always offer the null offer \( k^f_L = 0 \) to the low-type firm. Further the informed lender, will always make a positive loan offer \( k^H* \) to the high-type firm.

The uninformed lender will make an offer an offer \( k^*_N \), which may be the null offer. The uninformed lender makes the null offer \( k^*_N = 0 \) when the problem of adverse selection is sufficiently severe that the uninformed lender is unable to make positive profits when entering the market.

The firm's weakly dominant strategy is to accept all loan offers\(^{13} \).

**Proposition 2.** The informed lender will lend more in expectation than the uninformed lender:

\[ \Delta k \equiv \gamma k^f_L + (1 - \gamma)k^H - k_N > 0 \]

\[ \text{Proof. In the appendix.} \]

Proposition 2 shows that the informed lender will make larger loans on average than the uniformed lender. If the firm is a high-type, the uninformed lender competes with the informed lender on offering a loan, therefore, they split the profits from servicing the high-type firm. However, on the low-type firms, the uninformed lender is the sole provider of loans and makes a loss. Overall, the losses from the low-type firm leads the uninformed lender to make smaller loans on average.

**Proposition 3.** If the uninformed continues to lend, the uninformed lender will have greater rates of default than the informed lender:

\[ \Delta D \equiv E [D^N - D^f | k_N > 0] > 0 \]

\[ \text{Proof. In the appendix.} \]

The information asymmetry leads the uninformed lender to offer loans to the low-type borrowers leading to greater default rates. Combining proposition's 2 and 3 the uninformed lender lends less than the informed lender and makes worse quality loans.

The lack of public information leads to two sources of welfare loss: (1) the good borrowers can receive too much capital and (2) the low-type borrowers receive too much capital.

\(^{13}\text{There is another PBE where both the informed and uninformed lender's make the null offers } k^f_L = k^H = k_N = 0, \text{ and the firm rejects all loan offers. Given these strategies, no lender or firm could be better off. However, we ignore this equilibrium in our analysis, since it involves the firm playing a weakly dominated strategy.} \)
Proposition 4. As we reduce $X_L$, the expected difference between how much the informed and uniformed lender offer is increasing.

$$\Delta \equiv (1 - \gamma)k_I - k_N \text{ is decreasing in } X_L$$

Proof. In the appendix.

As we decrease the quality of the "lemons" in the model, the uninformed lender makes greater losses by servicing the entire market. This leads to the uninformed lender reducing her overall lending.

Since the informed lender never makes an offer to the low-type borrower, the informed lender is affected solely through the reduction in the uninformed lender's willingness to offer loans. If the uninformed lender reduces the size of her loan offer, the informed lender's optimal reaction is actually to increase her offer.

1.4.5 Mapping the model to the data.

Empirical predictions of proposition 2: The informed lenders will offer larger loans on average to those firms for whom they have better information following the reform.

Empirical predictions of proposition 3: The informed lenders will have default rates similar to the uninformed lender following the reform on those firms they both continue to serve, but the uninformed lender will have greater overdue rates overall.

Empirical predictions of proposition 4: The informed lenders will lend more to those firms where the quantity of lemons is greatest, since determining which firms are good credit risks, and which are not, is more important as we increase the proportion of bad types. Generally, those firms which have had an amount overdue in the past would be the riskiest firms. We would expect the effect of the reform to be the largest on those firms.

1.5 The Effect of the Reform on a Firm’s Source of Credit

1.5.1 Econometric specification

The paper's main question is: What is the effect of asymmetric information between lenders on a firm's source of credit? We answer this question using a reform that exogenously reduced banks information about firms in a way that varied across bank-firm lending pairs. To test the model's predictions, we examine whether those banks with private information about a firm's group were more likely than other banks to renew a firm's loans following the reform.

In the paper's main specifications, a "group" is defined using a relatively narrow definition: An overlapping director must own a substantial amount of equity in two firms before those two firms are grouped together. We will demonstrate that altering the definition of a group in economically meaningful ways will lead to different results.
**Definition 1.** Firm $f$'s group is all other firms with whom firm $f$ has a common director and at least 25\% shareholding in both firms as of December 2004.\(^{14}\)

The main source of identification in the paper will be to compare borrowing amounts for the same firm from two different lenders, before and after the regulation, where each lender lends to different members in the group at baseline (so we are restricting the sample to lenders who were already lending before the reform, and therefore, by definition have some information). For example, assume there are two firms “F1” and “F2” who are in the same group. Both “F1” and “F2” borrow from two banks each in December 2004. “F1” borrows from banks “A” and “B” and “F2” borrows from banks “B” and “C,” as shown in Figure 1-3.

Notice that bank “B” is lending to both firms in the group in December 2004, whereas bank “A” and bank “C” only lend to one firm in the group. After the regulation change, only bank “B” is able to compile the group’s lending.

**Definition 2.** A loan from bank $b$ to firm $f$ is labeled a “informed loan” if at least one other member of firm $f$’s group borrows from the same bank, $b$, in December 2004 (baseline). Similarly, a loan from bank $b$ to firm $f$ is labeled an “uninformed loan” if no other member of firm $f$’s group borrows from the same bank, $b$, in December 2004.

\(^{14}\)This definition of groups allows firms to be in multiple groups and that groups are not mutually exclusive. Therefore, a group is defined with respect to a firm. For a quick pictorial representation how firms can be in multiple groups see figure 1-4.
In the stylized example in figure 1-3, firm F1 has one informed loan (the loan with bank "B") and one uninformed loan (the loan with bank "A"). The paper's identification will be examining how lending changes between the informed and uninformed loans, before and after the regulation change.

We estimate equations of the form:

\[ Y_{bft} = \alpha_{bf} + \alpha_{bt} + \alpha_{ft} + \beta_1 \times \text{Post}_t \times \text{Informed}_{bf} + \epsilon_{bft} \]

The unit of observation is at the bank-firm-date level, so \( Y_{bft} \) is the variable of interest at bank \( b \), firm \( f \) in month \( t \). For example, it could be the size of the loan outstanding by firm \( f \) at bank \( b \), in month \( t \). Informed\( _{bf} \) is a dummy variable equal to one if the loan between bank \( b \) and firm \( f \) is an informed loan. Post is a dummy variable equal to one for a loan after April 2006.

The standard errors \( \epsilon_{bft} \) are clustered at the level of the component – at the level such that every firm within the component has at least one indirect link to every other firm in the component.\(^{15}\)

All the main regressions contain a firm x date fixed effect, \( \alpha_{ft} \). This fixed effect implies that we are estimating the difference in firm lending using differences for the same firm and in the same month. Therefore, we are estimating \( \beta_1 \) from only those firms that have both an informed and an uninformed loan. We have a total of 449 firms and 1,784 loans in December 2004 who had both an informed and an uninformed lending relationship. Table 1.3 presents more details about the firms which identify \( \beta_1 \).

Figure 1-11 demonstrates that the firms who identify \( \beta_1 \) in general have larger loans. This is in part mechanical, since we are only identifying the effect of the policy from those firms who have at least two loans.

Including the fixed effect \( \alpha_{bt} \) ensures that we are allowing for any aggregate change in bank lending for each month, and the fixed effect \( \alpha_{bf} \) ensures we are controlling for any firm-bank specificity.

This section restricts attention to those banks that were lending in December 2004, therefore, the paper does not include any new relationship in the analysis because any new borrowing relationship may be endogenous. Furthermore, in the estimation procedure, firms who discontinue relationships with all of their original lenders are dropped\(^{16}\).

1.5.1.1 Outcomes of interest

There are three main outcomes of interest in the paper:

- Log loan size\( _{bft} \)
- Renewed loan\( _{bft} \)

\(^{15}\)A component is defined as the set of firms for whom their exists a direct or indirect link to every other firm within the component. A direct link between two firms is defined if there is at least one mutual director who owns at least 25% of each firm and an indirect link exists between two firms if there is at least one path of direct links between the two firms. Conceptually, if firm \( f \) is in component \( c \) then firm \( f \)'s group must be a subset of a component \( c \). The definition of a link, is similar to the one used in all benchmark specifications.

\(^{16}\)The paper shows robustness results showing similar effects with the set of firm-bank connections which always have an active borrowing relationship.
• Overdue$_{bft}$

"Log loan size$_{bft}$" is defined as the log of real funded loan size outstanding in date $t$ at bank $b$ by firm $f$. If there is no loan observed this is coded in the data as equal to the minimum of what is observable in the data set (the log of 500,000 Rs.).

"Renewed loan$_{bft}$" is defined as whether firm $f$ at bank $b$ at date $t$ has an outstanding funded loan amount above 500,000 Rs. (in 2004 Pakistani Rupees).

"Overdue$_{bft}$" is defined as whether firm $f$ at bank $b$ at date $t$ is overdue at date $t$. There are certain endogeneity issues when looking at overdue rates because a firm can only go overdue if a firm has a loan which will be shown to be a function of the amount of information a bank has about a borrower. This is discussed in greater detail in the results.

1.5.2 Results

We explore the implications of differentially informed lenders on a firm’s source of credit in three ways: (i) to explore the overall effects of the reform, we estimate the difference in loan sizes between informed and uninformed lenders (ii) to explore the role of private information, we estimate how the difference in loan sizes changes as we use various measures of informed lenders and (iii) to explore whether there were heterogeneous effects for different firms, we estimate the effect of the reform on different firm sizes and differing credit risks.

1.5.2.1 Following the reform, loan sizes were relatively larger within informed banking relationships.

Our main results indicate that when public information available to banks was reduced, the change caused those banks to lend more to firms for whom they had greater private information. Banks with a better knowledge of a firm’s group were more likely to renew the credit facility, and grant greater credit for renewed loans.

This is most strikingly represented in Figure 1-12. Figure 1-12 plots the coefficients from a regression of loan renewal on an informed dummy, interacted with date dummies, and all second-order fixed effects, firm×date, date×bank, and firm×bank. The figure clearly shows that the difference in renewal rates between informed and uninformed loans is relatively constant before the reform, suggesting that the common trends identifying assumption holds. Following the reform, there is a sharp and persistent increase in the renewals of informed loans relative to uninformed loans. This indicates that the reform causes banks to increase their lending to firms for which they have more information.

Table 1.4 shows the estimates for the policy’s effect under different specifications. The first column contains date, firm and bank fixed effects, and therefore is being estimated from between- and within-firm differences. Columns 2-6 all include a firm interacted with date fixed effects. This ensures that we only use the set of borrowers who have both, an informed and uninformed banking relationship, to identify whether an informed banking relationship was more likely to be renewed.
The estimates are all relatively similar and the various specifications suggest that an informed banking relationship was between 5-8% more likely to be renewed than an uninformed banking relationship for the same firm. The preferred specification is column 6, which includes all three second order-fixed effects. In doing so, we are controlling for any aggregate changes in a bank's willingness to offer credit over time and firm-bank match specificity, and identifying the effect from firms which had both, an informed and uninformed lending relationship. Figure 1-12 is the graphical counterpart of the regression in column 6, except that we interact date dummies with the informed dummy variable.

While the previous results examine the effects of the policy on the extensive margin, here we can examine if the total size of the funded loan was larger in an informed banking relationship after the regulation change.

Figure 1-13 plots the coefficients from the log loan size on an informed dummy variable interacted with date dummies and all second-order fixed effects. Though the estimates are less precise than the extensive margin, we clearly see a similar trend as the extensive margin. Table 1.5 shows the same specifications as Table 1.4 but the dependent variable is the log of real funded loan size. The various estimates suggest that an informed banking relationship was between 8-12% larger following the policy change.

As a robustness check, we restrict the sample to those lending relationships which last until June 2008 or December 2008, and compare loan balances across informed and uninformed banking relationships for the same firm. Table 1.6 shows the estimates from this regression. Although power is an issue, the size of the effects look similar across the different specifications to the results in table 1.5.

The SBP only required banks to report details of a loan if the firm's total loan outstanding was greater than 500,000 Rupees ($8,500) could lead to the following bias in our results. Loans which were initially just above the cutoff could be partially repaid, subsequently falling below the 500,000 Rupees threshold and as such incorrectly categorized as a non-renewed loan. This is a greater concern since the uninformed loans are in general smaller than the informed loans. Table 1.7 excludes firms which had a loan close to the cutoff in December 2004. The observed effect of the policy seems consistent when we exclude firms with a loans below $12,750, $17,000 or $21,250 in December 2004. This suggests that the censoring of the data at 500,000 Rupees does not affect the results.

1.5.2.2 There were substantial differences in the effect of the policy on the credit market depending on the strength of the interfirm relationship.

The results in section 1.5.2.1 demonstrate that the reform led banks with private information about a group to lend more than other banks. Here we consider whether the effects vary by the amount of private information an informed lender has. In particular we consider variation in: the strength of the observed interfirm relationship and the number of firms who borrow from the same lender.

We observe substantial heterogeneity in the measured effect of the reform depending on the level of control implied within the interfirm relationship. If we separate the interfirm relationships according to the

17If a loan is not observed, we code the loan size to be the log of 500,000 Rs., which is the minimum threshold at which we are able to observe a loan.
amount of equity a director owns, it is clear there was little or no effect on those firms which had only overlapping directors. Further, the size of the effect was increasing in the interfirm relationship according to equity levels, which is a proxy for control. Therefore, this suggests that merely overlapping directors had no informational content on a firm’s creditworthiness.

Table 1.8 and figure 1-14 shows the effect of the reform for different levels of equity held by a mutual director. If a bank observed two firms who shared a director who owned at least 40 percent of both firms, the bank was 7.6% more likely to renew a loan from either firm, compared to a bank who only observed one firm. However if a director had no ownership stake in either firm, the bank was only 2.5% (not statistically significant) more likely to renew their loans.

An additional measure of the strength of the firm-bank relationship is observing the number of other group firms who borrow from the same bank. If more firms in the same group borrow from a bank, that lender is expected to have more private information about interfirm links, and therefore be more willing to lend. Tables 1.9 and figure 1-15 demonstrate that as we increase the number of interfirm relationships borrowing from the same bank, the effect of the policy was much larger. Those bank-firm pairs where the bank had greater information over the group are observed to increase their lending more.

1.5.2.3 The effect of the change in information was predominantly felt by small to medium sized firms.

From a policy perspective it is important to examine what type of firms were most affected by the reform. We examine if the effect varies by firm size.

Table 1.10 and figure 1-16 show the effects of the policy on the likelihood of a loan being renewed by different deciles, where the deciles are created according to the total amount the firm borrows at baseline. We see almost no effect (and certainly no statistically significant effect) on the largest decile of borrowers.

These results suggest that public information was most important for small to medium sized firms. These results can be interpreted in two ways: (i) the relative cost of procuring information is largest for the smallest firms or (ii) information about the largest firm’s groups is already well known by the banks. We consider each possibility in turn.

As discussed in section 1.3.2, banks procure more information about a firm in addition to what is provided by the SBP’s credit registry service. This information — such as calling other bank managers, or accessing the SECP database — may be costly to acquire. It is plausible that loan officers will conduct greater scrutiny over larger loans (assuming the cost of default is linear in loan size). Consequently the effect of the regulation change would be largest on those firms that borrowed the least (as this is where the proportional cost of acquiring more information to dollar lent is the largest).

Mr. Mansoor Siddiqi, the ex-Director Banking Policy and Regulation Department at the SBP remarked in an interview: “The number of corporate [firms] is not large in Pakistan therefore people tend to know about reputations,” further strengthening the assertion that we would expect little to no effect on those firms.
that had large groups. Table 1.11 and figure 1.11 show the effects of the policy on the likelihood of a loan being renewed by different deciles, where the deciles are created according to the total amount the firm’s group borrows at baseline. Similar to the results in table 1.10, we see no effect on those firms who are part of the largest groups.

1.5.2.4 The effect of the policy was largest on those firms who were overdue on a loan in December 2004.

Another way to analyze how the reform may have had heterogeneous effects across firms is to consider those firms who were observably worse credit risks prior to the reform. Firms that had loans overdue in December 2004 are expected to be at the greatest risk of being overdue in the future. Further, the problem of incomplete information on the firm’s group may be more pronounced, since one firm with an amount overdue might indicate financial distress within the wider group. So, a lending institution may be more willing to lend to a firm if it is able to inspect the wider set of group firms too. Also, when one firm defaults, the institution could react by reducing lending to the group at large.

Therefore, a firm that has an amount overdue may be more likely to borrow more from an informed banking relationship after the regulation change. Confirming this economic intuition, these results are shown in table 1.12.

These results are consistent with proposition 4 in the model: In environments where there is largest number of lemons, the uninformed lender is relatively more likely to stop lending.

1.5.2.5 Overdue rates are similar across informed and uninformed banking relationships.

The previous results highlighted how the reform affected who was able to borrow. In this section we examine if the reform affected overdue rates.

The reduction in public information may alter a firm’s incentive to default on a loan for two reasons. First, a firm may be more willing to be overdue because the impact on the rest of the firm’s group will be limited. Second, the lender would be less willing to offer a loan. In section 1.5.2.1 we demonstrated that the uninformed lender was 5.4% less likely to renew a firm’s loan than the informed lender.

Only if all loans were renewed would we be able to identify the effect of the change information on a firm’s incentive to default.

However, if strategic default was a key problem in these corporate markets you would expect the relative likelihood of a loan being overdue to be greater at an uninformed lender than an informed lender for the same firm. The firm is less likely to have his loan renewed by the uninformed lender, reducing the dynamic incentive to repay the loan. Also, the repercussions on a firm’s group would be smaller at an uninformed lender. Therefore, the absence of any differential on the likelihood of a loan being overdue between an informed and uninformed lender suggests there was no greater strategic default by the firm upon the uninformed lender.

The regressions in column 2 of Table 1.13, show little difference in overdue rates between informed and
uninformed banking relationships. Column 2, which estimates the difference in the likelihood of a loan being overdue at an informed and uninformed lender for the same firm, shows a relatively precise zero estimate. Column 2 suggests there was no strategic default by the firm following the regulation reform – even though they were less likely to have their loan renewed.

Column 1 is estimated from both, between-firm and within-firm differences, since it does not include a firm x date fixed effect. Column 1 suggests uninformed lenders select a worse set of loans to renew than informed lenders which is consistent with the proposition 3. In particular, defaults rates for uninformed lenders were 3% higher than informed lenders following the reform.

1.6 The Effect of the Reform on a Firm’s Access to Total Credit

In the previous section we established that a firm was more likely to receive a loan from an informed lender following the reform. In this section we wish to examine whether those firms with such informed lenders were more likely to have larger credit lines following the reform. Consequently, we ask: Did firms merely substitute their lending partners and receive the same total loan amounts? Or, did the reform lead to real effects in how much a firm was able to borrow?

1.6.1 Econometric specification

Definition 3. Firm $f$ has an informed lending relationship if firm $f$ borrows from a bank who also lends to at least one other member of firm $f$’s group in December 2004.

To examine whether the reform affected a firm’s access to credit we create a dummy variable “informed lender” which takes a value of “1” if the firm had an informed lending relationship in December 2004.

We identify the effect of the reform by comparing total loan amounts for those firms with and without an informed lending relationship before and after the reform. Formally, we estimate an equation of the form:

$$ Y_{ft} = a_f + a_t + \beta_1 \times \text{post}_t \times \text{Informed Lender}_f + \gamma X_{ft} + \epsilon_{ft} $$

Where $Y_{ft}$ is the log of real total funded borrowings by firm $f$ in month $t$. Informed lender$_f$ is a dummy variable equal to one, if the firm has at least one informed lending relationship in December 2004. Post is a dummy variable equal to one for a loan after April 2006. $a_f$ and $a_t$ are firm and date fixed effects respectively. In section (1.5) our identification relied on comparing the loan outcomes for the same firm, whereas in this section we rely on comparing the total credit borrowed between firms. Consequently, to allow for firm differences across regions and sectors, we include a set of firm by time controls $X_{ft}$, such as the firm’s province interacted with time fixed effects.

As before, the standard errors $\epsilon_{ft}$ are clustered at the level of the component – at the level such that
every firm within the component has at least one indirect link to every other firm in the component. As in the previous section, we do not observe loans smaller than 500,000 Rupees. Therefore, if we do not observe any loan for a firm, we code that firm's total borrowings to be 500,000 Rupees.

1.6.2 Results

1.6.2.1 Firms with an informed lending relationship borrowed more after the reform.

Table 1.14 estimates the effect of the reform on firms who had an informed lending relationship in December 2004. The difference-in-difference estimates suggest that those firms with an informed lender were able to borrow 11%-14% more following the reform, compared to firms with no informed lender. The results are robust to various controls such as a province interacted with date fixed effect and business sector interacted with date fixed effect.\(^{18}\)

Combining the results from section 1.5.2.1 with the results in table 1.14 suggests that the reform had two key effects: (i) lenders with greater information lent more to the same firm following the reform and (ii) those firms with informed lenders were able to access more credit following the reform.

This suggests that the reform not only led to a reallocation of lending across lenders but also led to real effects in the credit market by altering a firm's capacity to procure credit. Interpreting these results, the reduction in public information led to a reallocation of credit to borrowers for whom the lender had greater information. Ultimately, this disadvantaged firms who borrowed from banks which had limited information about their wider set of interfirm relationships.

1.6.2.2 The borrowing capacity of the smallest firms were most affected by the reform. The effect of the reform was greatest on the smallest firms' total borrowings.

We have established that the reform had real effects on how much a firm was able to borrow. To further evaluate the welfare and policy implications, we ask: What type of firms were the most affected by the reform?\(^{19}\)

To answer this question we estimate the effect of the reform on different sized firms.

Table 1.15 estimates the effect of the reform on different deciles of firm size and whether they had an informed lending relationship in December 2004. The difference-in-difference estimates suggest that those firms in the smallest decile were the most affected by the reform.

Combining the results from section 1.5.2.3 suggests that public information was crucial for the smallest firms. Not only were the more informed lenders more likely to lend to the smallest firms, but the smallest firms with an informed loan were more likely to procure larger total borrowings. The change in total borrowing for the largest decile was not significantly different from zero. These results further support the evidence

\(^{18}\)It should be noted the although the results are robust to a group size specific time trend, the results are not robust to a group size fixed effect interacted with a date fixed effect.
that the regulatory reform had significantly different effects for different sized firms. With the largest effects for the smallest firms.

1.7 Conclusion

Asymmetric information is a focal issue when studying credit markets. These information asymmetries may exist between a borrower and a lender, and between different lenders. To investigate the implications of adverse selection among differentially informed lenders, we use a reform by the State Bank of Pakistan.

In April 2006, the State Bank of Pakistan instituted a credit registry reform which exogenously reduced public information about a firm’s creditworthiness, and did so in a way that varied across firm-bank pairs.

First, we present a model of differentially informed lenders with adverse selection, which generates clear predictions on how the reform is expected to affect a bank's willingness to lend. The model predicts that those banks with greater private information about a firm would be more likely to continue to lend to that firm. Further, it may lead those lenders with the least information about a firm to stop lending.

Then, to test the model’s predictions we utilize the natural experiment that the SBP’s reform generated. We show empirically that the absence of public disclosure of group affiliation led to a reallocation of corporate borrowing. Those lenders who had greater private information were 5.4% more likely to renew a borrower’s loan. Remarkably, this is true even for banks that had a pre-existing relationship with the firm, suggesting that the strength of prior relationships does not eliminate the problem of imperfect information.

Not only do we see a reallocation of corporate borrowing, we also demonstrate that those firms who borrowed from an informed lender were able to borrow 12.3% more than firms who borrowed from less informed lenders. Consequently, we show that the source of a firm’s credit, as well as the total credit it could procure, were affected by the State Bank of Pakistan’s reform.

The model suggests that the reduction in public information leads to a negative welfare effect. Firstly, the asymmetry of information reduces the contestability of the market, and secondly, the reduction in information amplifies adverse selection, subsequently increasing the misallocation of credit. Finally, when the severity of adverse selection becomes worst, the uninformed lenders may stop lending.

Ultimately, complex interfirm relationships can alter a borrower’s incentives and ability to repay a loan – necessitating that a lender analyze the firm’s creditworthiness and consider the firm’s interfirm relationships. Public information about these relationships is key to reducing information asymmetries, both, between a borrower and a lender, and between lenders themselves.
1.8 Figures and Tables

Figure 1-4: Multiple overlapping groups

In the figure above there are three different firms and two directors but three different groups. Since a group is defined with respect to a firm (see definition 1), firm F1 only has a direct link to firm F2. Therefore, F1's group is only firms F1 and F2.
Figure 1-5: The network of firm connections - using a cutoff of 25% equity

This is a network that shows interfirm relationships which only show connections between firms if they have at least one mutual director who owns more than 25% in both firms. Those firms with more interfirm relationships are colored in a darker red.

The most visually striking aspect of this network is the lack of a single large component, and how the entire network is much less densely connected than the network compiled by merely mutual overlapping directors.
The graph displays the number of connections each firm possess in December 2004 where we define two firms to be linked if they share a mutual director who owns 25% or more of the firm’s equity.

This graph details the size of each component, where we define two firms to be linked if they share a mutual director who owns 25% or more of the firm’s equity. You can see the largest component is only 14 firms.
Figure 1-8: The histogram of log total firm borrowings

This is the histogram of the log total firm borrowings in December 2004 for group firms. Since, no loan below 500,000 Pu. Rupees was reported to the credit registry, there is a sharp cut-off at 6.2, which is approximately the log of 500.

Figure 1-9: The histogram of log loan size

This is the histogram of the log loan size in December 2004 for group firms. Since, no loan below 500 000 Pu. Rupees was reported to the credit registry, there is a sharp cut-off at 6.2, which is approximately the log of 500.
This is the histogram of the log loan size in December 2004 separated by informed and uninformed banking relationships. Since, no loan below 500,000 Pk. Rupees was reported to the credit registry, there is a sharp cut-off at 6.2, which is approximately the log of 500. We provide robustness tests to ensure the cutoff of small loans is not driving in our results.
This is the histogram of total firm borrowing in December 2004 separated by the firms we are using for identification in section 1.5. We are only identifying the effects of the policy from the set of firms who borrow from both an informed and uniformed lender, therefore, the right panel demonstrates the distribution of total firm borrowings who have both an informed and uninformed lending relationship. As you would expect the distribution of total borrowings for firms who have both types of relationship are larger. This is partly mechanical effect as they must have at least two loans. The larger total borrowings at baseline would suggest these firms are also larger in general.
This is the monthly coefficient for informed loans, when using ‘firm×date’, ‘firm×bank’ and ‘bank×date’ fixed effects. The light blue lines are point-wise 95% confidence intervals. It is clearly evident there was a dramatic increase in the likelihood of a loan not being renewed for the same firm when the borrowing relationship was an uninformed banking relationship.
This is the monthly coefficient for informed loans, when using 'firm×date', 'firm×bank' and 'bank×date' fixed effects. The light blue lines are point-wise 95% confidence intervals. There does not seem to be a pre-trend in the difference in log loan sizes but there is some evidence that the loan sizes increased after the regulation change.
This graph shows the effect of the policy according to different definitions of informativeness. In particular, we create new dummies according to the interfirm relationship a bank observes.

Those relationships where the bank observes two firms with solely overlapping directors has little to no effect on the likelihood a loan is renewed. Whereas if a lender observes two firms with an interfirm relationship where there is a mutual director who owns more than 40% of the company, those firms’ loans are 7.6% more likely to be renewed than if the lender only observed one of those firms. This is the visual analogue of column 2 in table 1.8.

The 95% confidence interval is depicted with the straight lines, and the estimated coefficient is the small blue box. All standard errors are clustered at the level of the component.
Figure 1-15: The effect of the policy on the likelihood of a loan being renewed by different number of firm’s who borrow from the same bank.

This graph shows as we increase the amount of firms who borrow from the same bank the effects of the policy are greater. This is the visual analogue of column 2 in table 1.9, where we are plotting the estimated coefficient and the standard errors from a regression of renewed loan on the number of firm connections borrowing at bank $b$ and the usual three second-order fixed effects.

The 95% confidence interval is depicted with the straight lines, and the estimated coefficient is the small blue box.

45
Figure 1-16: The effect of the policy on the likelihood of a loan being renewed by decile of firm size.

The graph shows the effects were greatest on the smallest firms (using December 2004 total borrowings). This is the visual analogue of column 2 in table 1.10, where we are plotting the estimated coefficient and the standard errors from a regression of renewed loan on decile (by firm total borrowings in December 2004) interacted with informed loans and the usual three second-order fixed effects.

The deciles are constructed using the total firm borrowings in December 2004, since those firms who borrow small amounts don’t have multiple loans, the estimate for the bottom decile is not identified.

The 95% confidence interval is depicted with the straight lines, and the estimated coefficient is the small blue box.
Figure 1-17: The effect of the policy on the likelihood of a loan being renewed by decile of group size.

The graph shows the effects were greatest on firms in the smallest groups (using December 2004 total borrowings). This is the visual analogue of column 1 in table 1.11, where we are plotting the estimated coefficient and the standard errors from a regression of renewed loan on decile (by group total borrowings in December 2004) interacted with informed loans and the usual three second-order fixed effects. These results suggest that those firms in the very largest groups were unaffected by the policy change. This is consistent with the suggestion that knowledge about the largest groups is already well known in the banking sector.

The deciles are constructed using the total group borrowings in December 2004, since those firms who borrow small amounts don't have multiple loans, the estimate for the bottom decile is not identified.

The 95% confidence interval is depicted with the straight lines, and the estimated coefficient is the small blue box. Standard errors are clustered at the level of the component.
Figure 1-18: The monthly coefficients for the difference in total borrowings between borrowers with informed lending relationships and those who do not.

This is the monthly coefficient for informed lender loans, when using 'firm', 'group size specific time trend', 'date FE×province FE', 'date FE×business sector FE' controls. The light blue lines are point-wise 95% confidence intervals. This graph shows that firms with an informed loan were more more likely to receive relatively larger loans following the reform than group firms who did not have a loan from an informed lender.
Figure 1-19: The effect of the reform on a firm’s total borrowings by decile of firm borrowings in December 2004.

This graph shows the effects were greatest on the smallest firms (using December 2004 total borrowings). This is the visual analogue of column 2 in table 1.15, where we are plotting the estimated coefficient and the standard errors from a regression of change in total firm borrowings on an informed lender interacted by each decile (by firm total borrowings in December 2004) and various controls.

The 95% confidence interval is depicted with the straight lines, and the estimated coefficient is the small blue box.
Table 1.1: Summary statistics for the entire dataset in December 2004

<table>
<thead>
<tr>
<th></th>
<th>Corporate Loans</th>
<th>Consumers and Sole Proprietors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of borrowers</td>
<td>11,395</td>
<td>86,053</td>
</tr>
<tr>
<td>Percentage of total lending</td>
<td>60%</td>
<td>40%</td>
</tr>
<tr>
<td>Overdue rates in December 2004</td>
<td>4.50%</td>
<td>7.11%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>S.D.</th>
<th>Mean</th>
<th>Median</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loan size (in '000 Pk. Rupees)</td>
<td>42,154</td>
<td>4,403</td>
<td>216,586</td>
<td>3,653</td>
<td>866</td>
<td>28,566</td>
</tr>
<tr>
<td>Number of lending partners</td>
<td>1.73</td>
<td>1.00</td>
<td>1.79</td>
<td>1.06</td>
<td>1.00</td>
<td>0.29</td>
</tr>
</tbody>
</table>

In December 2004, the credit registry maintained details on the entire credit market if a loan was greater than 500,000 Pk. Rupees. The table above demonstrates that largest lending was to corporate firms, and there was a wide dispersion in loan sizes.

Table 1.2: Summary statistics for group firms in December 2004

<table>
<thead>
<tr>
<th></th>
<th>Uninformed Loans</th>
<th>Informed Loans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of group firms</td>
<td>3,695</td>
<td></td>
</tr>
<tr>
<td>Number of loans</td>
<td>7,250</td>
<td></td>
</tr>
<tr>
<td>Number of loans</td>
<td>5,867</td>
<td>1,383</td>
</tr>
<tr>
<td>Overdue at baseline</td>
<td>2.77%</td>
<td>2.02%</td>
</tr>
<tr>
<td>Loan size (in '000s Pk. Rupees)</td>
<td>56,343</td>
<td>49,791</td>
</tr>
<tr>
<td>Log loan size</td>
<td>9.28</td>
<td>9.43</td>
</tr>
<tr>
<td>Public banks</td>
<td>7.0%</td>
<td>7.6%</td>
</tr>
<tr>
<td>Private domestic commercial bank</td>
<td>56.5%</td>
<td>66.4%</td>
</tr>
<tr>
<td>Non-Bank Fin. Corp. (NBFC)</td>
<td>28.2%</td>
<td>17.3%</td>
</tr>
<tr>
<td>Foreign bank</td>
<td>8.3%</td>
<td>8.7%</td>
</tr>
</tbody>
</table>

A loan is defined at a firm-bank pair. Therefore, if a firm has multiple loans from the same lender, it is classified as a single loan.

An informed loan is defined as a loan where at least two firms in the same group borrow from the same lender. In the following regressions, the paper shows how a firm who has both a informed and uninformed loan respond to the change in information.
Table 1.3: Summary statistics for group firms who have both an informed and uninformed loan in December 2004

<table>
<thead>
<tr>
<th></th>
<th>Uninformed Loans</th>
<th>Informed Loans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of group firms</td>
<td>449</td>
<td></td>
</tr>
<tr>
<td>Number of loans</td>
<td>1,784</td>
<td></td>
</tr>
<tr>
<td>Number of loans</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overdue at baseline</td>
<td>1.55%</td>
<td>1.33%</td>
</tr>
<tr>
<td>Loan size (in '000s Pk. Rupees)</td>
<td>45,667</td>
<td>65,162</td>
</tr>
<tr>
<td>Log loan size</td>
<td>9.18</td>
<td>9.55</td>
</tr>
<tr>
<td>Public banks</td>
<td>5.4%</td>
<td>5.0%</td>
</tr>
<tr>
<td>Private domestic commercial bank</td>
<td>48.9%</td>
<td>63.0%</td>
</tr>
<tr>
<td>Non-Bank Fin. Corp. (NBFC)</td>
<td>37.9%</td>
<td>23.1%</td>
</tr>
<tr>
<td>Foreign bank</td>
<td>7.3%</td>
<td>8.9%</td>
</tr>
</tbody>
</table>

A loan is defined at a firm-bank pair. Therefore, if a firm has multiple loans from the same lender, it is classified as a single loan.

Since the estimation results in section 1.5 are estimated from firms who have both an informed and uninformed lending relationship, this is the set of firms which identify our main coefficient of interest.
Table 1.4: The effect of the policy on the likelihood of a loan being renewed between informed and uninformed lending relationships

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informed</td>
<td>0.00213</td>
<td>0.0213</td>
<td>-0.00830</td>
<td>0.00453</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0150)</td>
<td>(0.0135)</td>
<td>(0.0135)</td>
<td>(0.0126)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post*Informed</td>
<td>0.0562***</td>
<td>0.0734***</td>
<td>0.0722***</td>
<td>0.0767***</td>
<td>0.0517**</td>
<td>0.0543**</td>
</tr>
<tr>
<td></td>
<td>(0.0121)</td>
<td>(0.0244)</td>
<td>(0.0241)</td>
<td>(0.0242)</td>
<td>(0.0217)</td>
<td>(0.0218)</td>
</tr>
<tr>
<td>Observations</td>
<td>269625</td>
<td>269625</td>
<td>269625</td>
<td>269625</td>
<td>269625</td>
<td>269625</td>
</tr>
<tr>
<td>Firm&amp;Date FE</td>
<td>Yes</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Date*Firm FE</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Bank FE</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Firm*Bank FE</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Bank*Date FE</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
* p < 0.1, ** p < 0.05, *** p < 0.01

Each regression shows the likelihood of a informed loan to be renewed to be 5-7 pp more likely. All the specifications 2-6 include a date×fim fixed effect, therefore, we identify the likelihood of a informed loan to be renewed solely from the set of borrowers who have both a informed and uninformed loan. All standard errors are clustered at the level of the component.
Table 1.5: The effect of the policy on the loan sizes between informed and uninformed lending relationships

<table>
<thead>
<tr>
<th></th>
<th>Loan Size</th>
<th>Loan Size</th>
<th>Loan Size</th>
<th>Loan Size</th>
<th>Loan Size</th>
<th>Loan Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informed</td>
<td>0.167**</td>
<td>0.449***</td>
<td>0.216***</td>
<td></td>
<td>0.227*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0761)</td>
<td>(0.0944)</td>
<td>(0.0728)</td>
<td></td>
<td>(0.0709)</td>
<td></td>
</tr>
<tr>
<td>Post*Informed</td>
<td>0.196***</td>
<td>0.111</td>
<td>0.120</td>
<td>0.125</td>
<td>0.0954</td>
<td>0.0916</td>
</tr>
<tr>
<td></td>
<td>(0.0480)</td>
<td>(0.0869)</td>
<td>(0.0854)</td>
<td>(0.0830)</td>
<td>(0.0786)</td>
<td>(0.0775)</td>
</tr>
<tr>
<td>Observations</td>
<td>269625</td>
<td>269625</td>
<td>269625</td>
<td>269625</td>
<td>269625</td>
<td>269625</td>
</tr>
<tr>
<td>Firm&amp;Date FE</td>
<td>Yes</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Date*Firm FE</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Bank FE</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Firm*Bank FE</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Bank*Date FE</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

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

Each regression shows the likelihood of an informed loan to be 9-13% to be larger following the reform. All the specifications 2-6 include a date×firm fixed effect, therefore, we identify the difference in loan size solely from the set of borrowers who have both an informed and uninformed loan. All standard errors are clustered at the level of the component.
Table 1.6: The effect of the policy on the loan sizes between informed and uninformed lending relationships restricting to only those loans that lasted until June 2008 or December 2008.

<table>
<thead>
<tr>
<th></th>
<th>Loan Size</th>
<th>Loan Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post*Informed</td>
<td>0.120</td>
<td>0.0698</td>
</tr>
<tr>
<td></td>
<td>(0.107)</td>
<td>(0.120)</td>
</tr>
<tr>
<td>Observations</td>
<td>74131</td>
<td>74417</td>
</tr>
<tr>
<td>Date cutoff</td>
<td>June 2008</td>
<td>Dec 2008</td>
</tr>
<tr>
<td>Date*Firm FE</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Firm*Bank FE</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Bank*Date FE</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

* p < 0.10, ** p < 0.05, *** p < 0.01

This set of regressions restricts attention to only those firm-banks pairs which last until June 2008 or December 2008. The results are very similar to those from table 1.5, suggesting the effect of the policy is not being driven by firms dropping out of the sample. All standard errors are clustered at the level of the component.
Table 1.7: The effect of the policy on the loan sizes between informed and uninformed lending relationships - censoring at baseline.

<table>
<thead>
<tr>
<th>Loan Renewed</th>
<th>Loan Renewed</th>
<th>Loan Renewed</th>
<th>Loan Renewed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post*Informed</td>
<td>0.0543**</td>
<td>0.0495**</td>
<td>0.0356</td>
</tr>
<tr>
<td></td>
<td>(0.0218)</td>
<td>(0.0228)</td>
<td>(0.0248)</td>
</tr>
<tr>
<td>Observations</td>
<td>269625</td>
<td>258814</td>
<td>213599</td>
</tr>
<tr>
<td>Cutoff</td>
<td>$8500</td>
<td>$12,750</td>
<td>$17,000</td>
</tr>
<tr>
<td>Date*Firm FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Firm*Bank FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Bank*Date FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
* p < 0.1, ** p < 0.05, *** p < 0.01

The SBP had a policy of only obliging banks to report a loan’s details if the firm’s total loan outstanding was greater than 500,000 Rupees ($8,500). This could lead to a bias in our results since loans which were initially just above the cutoff may remain active but partially repaid to be subsequently below 500,000 Rupees and incorrectly categorized as a non-renewed loan. This table omits firms which had a loan amount close to the cutoff in December 2004. Column 1 is the baseline results (Table 1.4), column 2 omits firms who had a loan outstanding below $12,750 in December 2004 (50% larger than the SBP cutoff), column 3 omits firms who had a loan outstanding below $17,000 in December 2004 (100% larger than the cutoff), and column 4 omits firms who had a loan outstanding below $21,250 in December 2004 (150% larger than the cutoff). All the estimates are similar to those in column 1 suggesting the cutoff is not leading to a noticeable bias in the results. All standard errors are clustered at the level of the component.
Table 1.8: The effect of greater shareholding on the effect of the policy on loan sizes and likelihood of renewing the loan.

<table>
<thead>
<tr>
<th></th>
<th>Loan Size</th>
<th>Loan Renewed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post*Informed 0%</td>
<td>0.0397</td>
<td>0.0248</td>
</tr>
<tr>
<td></td>
<td>(0.0716)</td>
<td>(0.0202)</td>
</tr>
<tr>
<td>Post*Informed 10%</td>
<td>0.0980</td>
<td>0.0440*</td>
</tr>
<tr>
<td></td>
<td>(0.0936)</td>
<td>(0.0234)</td>
</tr>
<tr>
<td>Post*Informed 25%</td>
<td>0.170*</td>
<td>0.0555**</td>
</tr>
<tr>
<td></td>
<td>(0.0993)</td>
<td>(0.0275)</td>
</tr>
<tr>
<td>Post*Informed 40%</td>
<td>0.0218</td>
<td>0.0767**</td>
</tr>
<tr>
<td></td>
<td>(0.122)</td>
<td>(0.0328)</td>
</tr>
<tr>
<td>Observations</td>
<td>269625</td>
<td>269625</td>
</tr>
<tr>
<td>Date*Firm FE</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Firm*Bank FE</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Bank*Date FE</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
* p < 0.1, ** p < 0.05, *** p < 0.01

Construction of the 'informed 40%' variable is the same as the procedure in the paper for constructing the benchmark bank-firm pairs, except using a 40% cutoff for directorial shareholding. We construct the 'informed 25%' variable to be those firm-bank pairs who would be defined as informed using the usual definition with a 25% director shareholder cutoff - except omitting those firm-bank pairs that were already labeled as 'informed 40%'. A similar procedure is used to construct the 10% and 0% informed variables. All standard errors are clustered at the level of the component.
Table 1.9: The effect of the greater number of firms borrowing from the same institution on loan sizes and likelihood of renewing the loan.

<table>
<thead>
<tr>
<th></th>
<th>Loan Size</th>
<th>Loan Renewed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post*2 Informed firms</td>
<td>0.0554</td>
<td>0.0431*</td>
</tr>
<tr>
<td></td>
<td>(0.0804)</td>
<td>(0.0228)</td>
</tr>
<tr>
<td>Post*3 Informed firms</td>
<td>0.212</td>
<td>0.111***</td>
</tr>
<tr>
<td></td>
<td>(0.143)</td>
<td>(0.0356)</td>
</tr>
<tr>
<td>Post*3+ Informed firms</td>
<td>0.446*</td>
<td>0.0913*</td>
</tr>
<tr>
<td></td>
<td>(0.229)</td>
<td>(0.0511)</td>
</tr>
</tbody>
</table>

Observations: 269625 269625

Date*Firm FE: Yes Yes
Firm*Bank FE: Yes Yes
Bank*Date FE: Yes Yes

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Construction of the '# informed firms' variable is done through adding the total number of firms in a firm's group who borrow from the same lender. Therefore, these regressions suggest a bank was more willing to renew a loan, the greater the number of group firms the lender lent to in December 2004. All standard errors are clustered at the level of the component.
Table 1.10: The effect of the policy by decile of firm size

<table>
<thead>
<tr>
<th>Post<em>Informed</em>2nd Decile by Firm Size</th>
<th>Loan Size</th>
<th>Loan Renewed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.461**</td>
<td>0.543***</td>
</tr>
<tr>
<td></td>
<td>(0.232)</td>
<td>(0.127)</td>
</tr>
<tr>
<td>Post<em>Informed</em>3rd Decile by Firm Size</td>
<td>0.247</td>
<td>0.166</td>
</tr>
<tr>
<td></td>
<td>(0.242)</td>
<td>(0.108)</td>
</tr>
<tr>
<td>Post<em>Informed</em>4th Decile by Firm Size</td>
<td>-0.0207</td>
<td>-0.122</td>
</tr>
<tr>
<td></td>
<td>(0.207)</td>
<td>(0.122)</td>
</tr>
<tr>
<td>Post<em>Informed</em>5th Decile by Firm Size</td>
<td>0.112</td>
<td>0.0607</td>
</tr>
<tr>
<td></td>
<td>(0.212)</td>
<td>(0.0783)</td>
</tr>
<tr>
<td>Post<em>Informed</em>6th Decile by Firm Size</td>
<td>0.356</td>
<td>0.109</td>
</tr>
<tr>
<td></td>
<td>(0.211)</td>
<td>(0.0738)</td>
</tr>
<tr>
<td>Post<em>Informed</em>7th Decile by Firm Size</td>
<td>0.0125</td>
<td>0.0262</td>
</tr>
<tr>
<td></td>
<td>(0.215)</td>
<td>(0.0686)</td>
</tr>
<tr>
<td>Post<em>Informed</em>8th Decile by Firm Size</td>
<td>0.134</td>
<td>0.158***</td>
</tr>
<tr>
<td></td>
<td>(0.146)</td>
<td>(0.0519)</td>
</tr>
<tr>
<td>Post<em>Informed</em>9th Decile by Firm Size</td>
<td>0.102</td>
<td>0.0225</td>
</tr>
<tr>
<td></td>
<td>(0.174)</td>
<td>(0.0446)</td>
</tr>
<tr>
<td>Post<em>Informed</em>10th Decile by Firm Size</td>
<td>-0.00955</td>
<td>-0.00843</td>
</tr>
<tr>
<td></td>
<td>(0.152)</td>
<td>(0.0355)</td>
</tr>
</tbody>
</table>

Observations: 269625

Standard errors in parentheses

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

Each decile is constructed by computing the total firm borrowings in December 2004. The effect on the smallest decile is not identified in the data since the smallest borrowers in December 2004 only had one loan and the identification relies on a borrower having at least two loans. All standard errors are clustered at the level of the component. The results for column 2 are shown in figure 1-16.
Table 1.11: The effect of the policy by decile of group size

<table>
<thead>
<tr>
<th>Post<em>Informed</em>2nd Decile by Group Size</th>
<th>Loan Renewed</th>
<th>Loan Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.191</td>
<td>-0.112</td>
</tr>
<tr>
<td></td>
<td>(0.242)</td>
<td>(0.397)</td>
</tr>
<tr>
<td>Post<em>Informed</em>3rd Decile by Group Size</td>
<td>0.0394</td>
<td>-0.158</td>
</tr>
<tr>
<td></td>
<td>(0.139)</td>
<td>(0.227)</td>
</tr>
<tr>
<td>Post<em>Informed</em>4th Decile by Group Size</td>
<td>-0.0168</td>
<td>-0.106</td>
</tr>
<tr>
<td></td>
<td>(0.109)</td>
<td>(0.252)</td>
</tr>
<tr>
<td>Post<em>Informed</em>5th Decile by Group Size</td>
<td>0.187**</td>
<td>0.460***</td>
</tr>
<tr>
<td></td>
<td>(0.0864)</td>
<td>(0.164)</td>
</tr>
<tr>
<td>Post<em>Informed</em>6th Decile by Group Size</td>
<td>-0.00495</td>
<td>0.0781</td>
</tr>
<tr>
<td></td>
<td>(0.0690)</td>
<td>(0.214)</td>
</tr>
<tr>
<td>Post<em>Informed</em>7th Decile by Group Size</td>
<td>0.158***</td>
<td>0.0999</td>
</tr>
<tr>
<td></td>
<td>(0.0589)</td>
<td>(0.195)</td>
</tr>
<tr>
<td>Post<em>Informed</em>8th Decile by Group Size</td>
<td>0.123**</td>
<td>0.297*</td>
</tr>
<tr>
<td></td>
<td>(0.0605)</td>
<td>(0.177)</td>
</tr>
<tr>
<td>Post<em>Informed</em>9th Decile by Group Size</td>
<td>0.0249</td>
<td>-0.0769</td>
</tr>
<tr>
<td></td>
<td>(0.0438)</td>
<td>(0.159)</td>
</tr>
<tr>
<td>Post<em>Informed</em>10th Decile by Group Size</td>
<td>0.0155</td>
<td>0.0967</td>
</tr>
<tr>
<td></td>
<td>(0.0368)</td>
<td>(0.152)</td>
</tr>
</tbody>
</table>

Observations 269625
Date*Firm FE Yes
Firm*Bank FE Yes
Bank*Date FE Yes

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Each decile is constructed by computing the total group borrowings in December 2004. The effect on the smallest decile is not identified in the data since the smallest borrowers in December 2004 only had one loan and the identification relies on a borrower having at least two loans. The results show that the effect of the reform was largest on those firms who belonged to the smallest group. This is consistent with the view that there is common knowledge about a firm’s group for the very largest groups.

All standard errors are clustered at the level of the component. The results for column 2 are shown in figure 1-17.
This is suggestive evidence that those firms who had been overdue in the past were more likely to renew their loans at an informed banking relationships. This suggests that those firms for may be classified as the most risky were also the ones most likely to have their loans renewed by an informed lender. All standard errors are clustered at the level of the component.
Table 1.13: The difference in the likelihood of a loan going overdue

<table>
<thead>
<tr>
<th></th>
<th>Overdue</th>
<th>Overdue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informed</td>
<td>0.0163***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00518)</td>
<td></td>
</tr>
<tr>
<td>Post*Informed</td>
<td>-0.0287***</td>
<td>-0.00719</td>
</tr>
<tr>
<td></td>
<td>(0.00666)</td>
<td>(0.0141)</td>
</tr>
<tr>
<td>Observations</td>
<td>209962</td>
<td>209962</td>
</tr>
<tr>
<td>Firm&amp;Date FE</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td>Date*Firm FE</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Bank FE</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td>Firm*Bank FE</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Bank*Date FE</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

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

Caution should be taken when taking inference from these regressions. As described in the text, this is not estimating the causal impact of the regulation change on the likelihood of a loan becoming overdue, since in table 1.4, we showed uninformed loans were more likely to be not renewed. Column 1 does not include firm interacted with date fixed effects. Therefore, column 1 suggests the loans that were renewed by uninformed lenders were more likely to default. Column 2 which includes the firm interacted with date fixed effect suggests that those firms for whom the informed and uninformed lender renewed, there was no differential in default.

All standard errors are clustered at the level of the component.
Table 1.14: Difference in total loan sizes between firms who have an informed lender and those who do not.

<table>
<thead>
<tr>
<th></th>
<th>Total Loans</th>
<th>Total Loans</th>
<th>Total Loans</th>
<th>Total Loans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post*Informed lender</td>
<td>0.130***</td>
<td>0.134***</td>
<td>0.113**</td>
<td>0.123**</td>
</tr>
<tr>
<td></td>
<td>(0.0447)</td>
<td>(0.0458)</td>
<td>(0.0464)</td>
<td>(0.0476)</td>
</tr>
<tr>
<td>Constant</td>
<td>9.845***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0470)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>161744</td>
<td>155496</td>
<td>149248</td>
<td>145684</td>
</tr>
<tr>
<td>Firm FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Date FE</td>
<td>Yes</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Group Size Specific Time Trend</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Business Sector FE*Date FE</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Province FE*Date FE</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
* p < 0.10, ** p < 0.05, *** P < 0.01

The difference-in-difference estimates suggest that those firms with an informed lender were able to borrow 11%-14% more following the reform than those firms with no informed lender. Column 1 shows the estimates with firm fixed effects and a group size specific time trend. Column 2 shows the estimates when including business sector interacted with date fixed effects. Column 3 shows the estimates when including province interacted with date fixed effects. Lastly column 4 shows the estimates where we include all interacted fixed effects. We are missing the data on the province and business sector for a small fraction of the firms, therefore, they are omitted from those regressions which include province or business sector fixed effects. All standard errors are clustered at the level of the component.
Table 1.15: The difference in total loan borrowings by firm decile and whether they had an informed lending relationship.

<table>
<thead>
<tr>
<th>Post<em>Informed</em>1st Decile</th>
<th>Total Loans</th>
<th>Total Loans</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.764***</td>
<td>0.756***</td>
</tr>
<tr>
<td></td>
<td>(0.118)</td>
<td>(0.137)</td>
</tr>
<tr>
<td>Post<em>Informed</em>2nd Decile</td>
<td>0.253**</td>
<td>0.290**</td>
</tr>
<tr>
<td></td>
<td>(0.103)</td>
<td>(0.121)</td>
</tr>
<tr>
<td>Post<em>Informed</em>3rd Decile</td>
<td>0.310***</td>
<td>0.318***</td>
</tr>
<tr>
<td></td>
<td>(0.0957)</td>
<td>(0.101)</td>
</tr>
<tr>
<td>Post<em>Informed</em>4th Decile</td>
<td>-0.274***</td>
<td>-0.311***</td>
</tr>
<tr>
<td></td>
<td>(0.104)</td>
<td>(0.119)</td>
</tr>
<tr>
<td>Post<em>Informed</em>5th Decile</td>
<td>0.217**</td>
<td>0.173</td>
</tr>
<tr>
<td></td>
<td>(0.102)</td>
<td>(0.111)</td>
</tr>
<tr>
<td>Post<em>Informed</em>6th Decile</td>
<td>0.0678</td>
<td>0.0708</td>
</tr>
<tr>
<td></td>
<td>(0.0988)</td>
<td>(0.101)</td>
</tr>
<tr>
<td>Post<em>Informed</em>7th Decile</td>
<td>-0.0511</td>
<td>-0.0282</td>
</tr>
<tr>
<td></td>
<td>(0.129)</td>
<td>(0.128)</td>
</tr>
<tr>
<td>Post<em>Informed</em>8th Decile</td>
<td>-0.0714</td>
<td>-0.100</td>
</tr>
<tr>
<td></td>
<td>(0.134)</td>
<td>(0.136)</td>
</tr>
<tr>
<td>Post<em>Informed</em>9th Decile</td>
<td>0.267**</td>
<td>0.257**</td>
</tr>
<tr>
<td></td>
<td>(0.105)</td>
<td>(0.105)</td>
</tr>
<tr>
<td>Post<em>Informed</em>10th Decile</td>
<td>0.0340</td>
<td>0.0728</td>
</tr>
<tr>
<td></td>
<td>(0.132)</td>
<td>(0.143)</td>
</tr>
<tr>
<td>Constant</td>
<td>9.845***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0458)</td>
<td></td>
</tr>
</tbody>
</table>

Observations: 161744  145684
Firm FE: Yes  Yes
Date FE: Yes  N/A
Group Size Specific Time Trend: Yes  Yes
Business Sector*Date FE: No  Yes
Province*Date FE: No  Yes

Standard errors in parentheses

* p < 0.10, ** p < 0.05, *** p < 0.01

These regressions interact a dummy variable for each decile (by total firm borrowings in December 2004) with the post×informed lender variable. Hence, we are estimating the effect of the policy separately for each decile. The estimates suggest that the largest effects were on the smallest firms, which is similar to the results shown in section 1.5.2.3. Figure 1-19 plots the results in column 2. All standard errors are clustered at the level of the component.
1.9 Appendix

1.9.1 Proof of Proposition 1

A Perfect Bayesian Equilibrium of the game is:

\[
\begin{align*}
&\text{(informed lender)} \ k_i^H = 0, \ k_i^H = k_i^H* > 0 \\
&\text{(uninformed lender)} \ k_N = k_N^* \geq 0 \\
&\text{(firm)} \ s_f^i = (\text{Acc, Acc})
\end{align*}
\]

Proof. To prove the exist a PBE we will first solve for the firm's weakly dominant strategy, and then we shall proceed to solve for the PBE under different parameter values.

To solve for the firm's weakly dominant strategy recall the firm's utility function:

\[
U_i(k, R) = \frac{X_i}{1 + k} (A - R)k
\]

The firm's utility function is strictly increasing in capital, \( k \), for all non-negative \( k \) since we assume the productivity parameter \( A \) is strictly greater than the interest rate \( R \). Therefore, the firm's weakly dominant strategy is to accept all loan offers from both lenders.

Now let us consider the informed lender's optimal strategy if the uninformed lender is offering \( k_N \) and the firm accepts all loan offers. The informed lender wishes to maximise:

\[
\pi^i_i(k_i^i, k_N, X_i) = \max \left[ \frac{X_i}{1 + k_i^i + k_N} R - \rho \right] k_i^i
\]

Since we assumed that \( \frac{RX_L}{\rho} < 1 \), we know that the informed lender's strictly dominant strategy having observed a low borrower type will be \( k_i^i = 0 \) since expected profits will always be strictly negative if she ever offers a non-zero loan size.

Now let us consider the case where the informed lender observes a high borrower type and let us assume that the following condition holds:

\[
\gamma \left( \frac{RX_L}{\rho} \right) + (1 - \gamma) \left( \frac{RX_H}{\rho} \right)^{0.5} < 1 \quad (1.2)
\]

Then there exists a PBE:
To verify this is a PBE, the informed lender’s marginal profit condition is:

\[
\frac{\partial \pi}{\partial k^I} = RX_H \frac{(1+k^N)}{(1+k^N+k^H)^2} - \rho = 0
\]  

(1.3)

And since the profit function is concave in \( k^I \) the local maximum is sufficient for the lenders’ optimal strategy given the other lender’s strategy. If the uninformed lender chooses \( k^N = 0 \), then the solution to equation 1.3 is \( k^H = k^H_{\max} \). To check this is a PBE, we need to check the uninformed lender’s optimal strategy.

The marginal profit condition for the uninformed lender is:

\[
\frac{\partial \pi}{\partial k^N} = \gamma RX_L \frac{1}{(1+k^N)^2} + (1-\gamma)RX_H \frac{(1+k^H)}{(1+k^N+k^H)^2} - \rho
\]

If we substitute in the conjectured optimal informed lender’s strategy, this becomes:

\[
\frac{\partial \pi}{\partial k^N} \bigg|_{k^N=k^N_{\max}} = \gamma RX_L \frac{1}{(1+k^N)^2} + (1-\gamma)RX_H \frac{(RX_H)}{(k^N+RX_H)^{0.5}} = 0
\]  

(1.4)

Similar to the informed lender’s problem, the uninformed lender’s profit is concave in \( k^N \). Therefore, if we can show the uninformed lender’s marginal profit is negative at \( k^N = 0 \) we know the uninformed lender’s optimal strategy given the informed lender’s strategy is to offer \( k^N = 0 \).

Plugging in \( k^N = 0 \) into equation 1.4 gives:

\[
\frac{\partial \pi}{\partial k^N} \bigg|_{k^N=0,k^H=k^H_{\max}} = \gamma RX_L + (1-\gamma)RX_H^{0.5} - \rho
\]

Using the assumption in equation 1.2, we know this is strictly less than zero. Therefore, the uninformed lender’s optimal strategy given the informed lender’s strategy is to offer \( k^N = 0 \).

To complete the proof for all feasible values of \( X_H \) and \( X_L \) we will solve the uninformed and informed lender’s maximization problems assuming the constraints that each lender does not have a negative loan size or a negative expected profits do not bind. We will then show for those parameter values where the constraints do bind the equilibrium will be such that \( k^N = 0, k^I = 0, k^H = k^H_{\max} \).

Recall the informed lender’s marginal profit condition:
Recall the uninformed lender’s marginal profit condition:

\[
\frac{\partial \pi}{\partial k_I} = RX_H \frac{(1 + k_N)}{(1 + k_N + k_I^H)^2} - \rho
\]  
(1.5)

Assuming no constraints bind, we can set equations 1.5 and 1.6 equal to zero and solve for \( k_I^H \) and \( k_N \) as a function of the parameters (the first order conditions are sufficient since both lenders’ profits are concave in their respective capital offers). First we show that the solution of these two equations imply that \( k_I^H > 0 \).

To reduce notation let us define \( Y \equiv 1 + k_N \), \( Z_H \equiv \frac{RX_H}{\rho} \) and \( Z_L \equiv \frac{RX_L}{\rho} \).

Then we can rewrite the two marginal profit conditions (FOCs) as:

\[
\gamma Z_L \frac{1}{Y^2} + (1 - \gamma) Z_H \frac{(1 + k_I^H)}{(k_I^H + Y)^2} = 1
\]  
(1.8)

Plugging in \((k_I + Y)^2\) from equation 1.7 into equation 1.8. We have the equation:

\[
\gamma Z_L \frac{1}{Y^2} + (1 - \gamma) Z_H \frac{(1 + k_I^H)}{Z_H Y} = 1
\]  
(1.9)

Simplifying and rearranging equation 1.9:

\[
Y^2 - (1 - \gamma)(1 + k_I^H)Y - \gamma Z_L = 0
\]  
(1.10)

Solving for \( Y \):

\[
Y = \frac{(1 - \gamma)(1 + k_I^H) \pm \sqrt{(1 - \gamma)^2(1 + k_I^H)^2 + 4\gamma Z_L}}{2}
\]  
(1.11)

We are interested in \( k_N > 0 \) (the maximum), so we can restrict ourselves to the positive root of \( Y \).

Rearranging equation 1.7:

\[
Y^2 + (2k_I^H - Z_H)Y + (k_I^H)^2 = 0
\]  
(1.12)

Subtracting 1.10 from 1.12 gives:

\[
(2k_I^H - Z_H + (1 - \gamma)(1 + k_I))Y + (k_I^H)^2 + \gamma Z_L = 0
\]  
(1.13)

We can substitute \( Y \) from equation 1.11 into 1.13. Equation 1.13 is continuous in \( k_I^H \), the left hand size
of 1.13 is less than zero if \( k_H^I = 0 \), and the left hand side of 1.13 tends to infinity as \( k_H^I \) goes to infinity, therefore there must exist a positive \( k_H^I \) such that both FOCs are satisfied. Since, there exists a \( k_H^I \) which satisfy the two FOCs, then there must also exist a corresponding \( k_N^I \).

However, the choice of \( k_N^I \), which satisfies the FOC may not be feasible. In particular, at those parameter values \( k_N^I \) may be less than zero. We now show that \( k_N^I \) is always greater than zero from the FOC conditions.

Recall equation 1.11:

\[
Y = \frac{(1 - \gamma)(1 + k_H^I) \pm \sqrt{(1 - \gamma)^2(1 + k_H^I)^2 + 4\gamma Z_L}}{2}
\]

We want to show that \( Y \equiv 1 + k_N^I \) is always greater than one. Since, \( Y \) is increasing in \( k_H^I \), and we have shown \( k_H^I \) is always greater than zero, then if we show this equation holds when \( k_H^I = 0 \), it must hold for all parameter values. Therefore:

\[
Y = \frac{(1 - \gamma)(1 + k_H^I) + \sqrt{(1 - \gamma)^2(1 + k_H^I)^2 + 4\gamma Z_L}}{2} > 1
\]

Substituting in \( k_H^I = 0 \) and rearranging, we obtain:

\[
\sqrt{(1 - \gamma)^2 + 4\gamma Z_L} > (1 + \gamma)
\]

Square both sides (both sides are positive) and rearrange:

\[
Z_L > \frac{1}{2}
\]

From our assumptions on the parameter space, this is always true. Therefore both the \( k_H^I \) and \( k_N^I \) that satisfy the two FOCs are greater than zero.

Finally to complete the proof, we must show what happens when the uninformed lender’s profit condition is not satisfied. We need the following preliminaries: (i) the informed lender’s best response function is decreasing in \( k_N^I \) for all feasible values of \( k_N^I \), (ii) the uninformed lender’s best response function is decreasing in \( k_H^I \) and (iii) in equilibrium \( k_H^I > k_N^I \).

Proof of preliminary (i)

Totally differentiating equation 1.5 with respect to \( k_N^I \) and rearranging:

\[
\frac{dk_H^I}{dk_N^I} = \frac{Z_H^I}{2(1 + k_N^I)^{0.5}} - 1 < 0
\]

Therefore, the optimal \( k_H^I \) is decreasing for all non-negative values of \( k_N^I \) since we assumed \( Z_H < 4 \).

Proof of preliminary (ii)

Totally differentiating the uninformed lender’s FOC condition (equation 1.8) with respect to \( k_H^I \) gives:
\[-2\gamma Z_L \frac{1}{Y^3} \frac{dY}{dk_f^H} + \frac{(1 - \gamma) Z_H}{(k_f^H + Y)^3} \left[ (k_f^H + Y)^2 - 2 \left( 1 + \frac{dY}{dk_f^H} \right) (k_f^H + Y)(1 + k_f^H) \right] = 0 \]

Rearranging and simplifying:

\[
\frac{dY}{dk_f^H} = (k_f^H + Y)^2 - 2(k_f^H + Y)(1 + k_f^H)
\]

\[
\frac{dk_N}{dk_f^H} = (k_f^H + Y)(k_N - 1 - k_f^H)
\]

Therefore, if \( k_f^H \) is greater than \( k_N \) then the uninformed lender’s choice of \( k_N \) is decreasing in \( k_f^H \).

Proof of preliminary (iii)

Follows from the proof of proposition 2.

Therefore, now we have shown that \( k_N \) is decreasing in \( k_f^H \), and \( k_f^H \) is decreasing in \( k_N \). We know if the FOC suggest a solution with negative expected profits for the uninformed lender, the uninformed lender will always set \( k_N = 0 \). This follows from (i) the informed lender’s optimal response will be to increase her choice of \( k_f^H \) if the uninformed lender’s choice of \( k_N \) decreases, and (ii) if the informed lender’s choice of \( k_f^H \) increases, the uninformed lender will always wish to decrease her choice of \( k_N \) – ultimately, this leads to \( k_N = 0 \).

Finally, we need to check the informed lenders expected profits are always greater than zero. There are two cases: (i) if the uninformed lender offers a positive loan size and (ii) the uninformed lender does not enter.

Case (i):

If the uninformed lender has a positive loan size then the following condition must hold:

\[
RX_H \frac{1}{(1 + k_N + k_f^H)^2} - \rho \geq 0 \quad (1.15)
\]

Recalling the informed lender’s profit function:

\[
E \left[ \pi_f^H(k_f^H, k_N) \right] = RX_H \frac{k_f^H}{(1 + k_N + k_f^H)^2} - \rho k_f^H
\]

If equation 1.15 is satisfied then the informed lender’s profits must be greater than zero.

Case (ii):

The informed lender profits must similarly be greater than zero since \( \frac{RX_H}{\rho} \) is assumed to be greater than 1, therefore, the informed lender can always make a positive profit.

To sum, we have shown there exists a solution to first order conditions when we do not restrict profits to be non-negative or loan size to be positive. Then we have shown a PBE there exists a PBE for those parameter values such that the first order conditions give a non-feasible solution.
Further, it can be shown there is no mixed strategy PBE in this model.

1.9.2 Proof of Proposition 2

The informed lender will lend more in expectation than the uninformed lender:

$$\Delta \equiv (1 - \gamma)k_H^* - k_N^* > 0$$

Proof. There are two possible cases, (i) $k_N^* = 0$ and (ii) $k_N^* > 0$.

Case (i): $k_N^* = 0$

Consider the informed lender's FOC:

$$RXH \frac{(1 + k_N)}{(1 + k_N + k_I)} - \rho = 0 \quad (1.16)$$

Using $k_N^* = 0$ and the assumption that $\frac{RX_H}{\rho} > 1$ (that it is profitable to offer loans to the high type), then the equilibrium $k_H^* > 0$ and $\pi_I^* > 0$. It follows $\Delta > 0$.

Case (ii): $k_N^* > 0$

In this situation both the lenders' FOCs must hold. We will first solve for the equilibrium $k_I^*$ and $k_N^*$ which satisfy the two FOCs, and then using a proof by contradiction we will show that $(1 - \gamma)k_I^* - k_N^* > 0$.

To reduce notation let us define $Y_1 \equiv 1 + k_N$; $Z_H \equiv \frac{RX_H}{\rho}$ and $Z_L \equiv \frac{RX_L}{\rho}$.

Recall equation 1.11

$$Y = \frac{(1 - \gamma)(1 + k_I) \pm \sqrt{(1 - \gamma)^2(1 + k_I)^2 + 4\gamma Z_L}}{2} \quad (1.17)$$

We are interested in $k_N > 0$ (the maximum), so we can restrict ourselves to the positive root of $Y$.

We want to show that $\Delta \equiv (1 - \gamma)k_I - k_N \equiv 1 + (1 - \gamma)k_I - Y > 0$. We will complete the proof by contradiction. Assume $\Delta \leq 0$ then:

$$1 + (1 - \gamma)k_I \leq Y \leq \frac{(1 - \gamma)(1 + k_I) \pm \sqrt{(1 - \gamma)^2(1 + k_I)^2 + 4\gamma Z_L}}{2} \quad (1.18)$$

In equation 1.18 we have substituted $Y$ from equation 1.17.

Rearranging and simplifying equation 1.18:

$$(1 - \gamma)k_I + (1 + \gamma) \leq \sqrt{(1 - \gamma)^2(1 + k_I)^2 + 4\gamma Z_L}$$

If we square both sides:
\[(1 - \gamma)k_I + (1 + \gamma))^2 \leq (1 - \gamma)^2(1 + k_I)^2 + 4\gamma Z_L\]

Simplifying:
\[4\gamma \leq 4\gamma Z_L - 4\gamma k_I \quad (1.19)\]

Since by assumption \(1 > Z_L\) and \(k_I \geq 0\) then 1.19 is false. Thereby completing the proof. \(\square\)

1.9.3 Proof of Proposition 3

If the uninformed continues to lend, the uninformed lender will have greater rates of default than the informed lender:
\[\Delta D \equiv E[D^N - D^I | k_N > 0] > 0\]

Proof. We restrict attention to this equilibrium where the uninformed lend continues to lend (since the uninformed lend cannot have a defaults if she does not lend).

In equilibrium, the informed lender's default rate is:
\[E(D^I) = \left(1 - \frac{X_H}{1 + k_I^* + k_N^*}\right) \quad (1.20)\]

The uninformed lender's default rate is:
\[E(D^N | k_N^* > 0) = \gamma \left(1 - \frac{X_L}{1 + k_N^*}\right) + (1 - \gamma) \left(1 - \frac{X_H}{1 + k_H^* + k_N^*}\right) \quad (1.21)\]

Therefore to complete the proof we need to show that, equation 1.21 is greater than equation 1.20:
\[\gamma \left(1 - \frac{X_L}{1 + k_N^*}\right) + (1 - \gamma) \left(1 - \frac{X_H}{1 + k_H^* + k_N^*}\right) > \frac{X_L}{1 + k_N^*} < \frac{X_H}{1 + k_H^* + k_N^*} \quad (1.22)\]

Recalling that the informed lender must make non-negative profits, then the RHS of inequality 1.22 must be greater than or equal to \(\frac{B}{\rho}\). From our initial assumptions \(X_L\) must be less than equal \(\frac{B}{\rho}\), \(k_N^*\) must be greater than zero. Therefore inequality 1.22 holds. \(\square\)

1.9.4 Proof of Proposition 4

As we reduce \(X_L\), the expected difference between how much the informed and uniformed lender offer is increasing.

\[\Delta \equiv (1 - \gamma)k_I - k_N \text{ is decreasing in } X_L\]
Proof. A change in $X_L$ only affects the informed lender through its effect on the uninformed lender’s choice of $k_N$ (formally, partially differentiating the informed lender’s profit maximization condition with respect to $X_L$ is zero).

Partially differentiating the uninformed lender’s FOC condition with respect to $X_L$:

$$\gamma R - \frac{1}{(1+k_N)^2} - 2\gamma RX_L \frac{1}{(1+k_N)^3} \frac{\partial k_N}{\partial X_L} - 2(1 - \gamma)RXH \frac{(1+k_H)}{(k_H + Y)^3} \frac{\partial k_N}{\partial X_L} = 0$$

Therefore it follows:

$$\frac{\partial k_N}{\partial X_L} \geq 0$$

Using the proof of proposition 1, we know that the informed lender’s choice of $k_I$ is decreasing in $k_N$, similarly the uninformed lender’s choice of $k_N$ is decreasing in $k_I$. Consequently, if $X_L$ rises, holding $k_H$ constant, we know that $k_N$ rises, but then this itself causes $k_H$ to fall, which in turn causes $k_N$ to rise more and so on. \(\square\)
A copy of the information provided by the SBP about a prospective borrower in 2004
A copy of the information provided by the SBP about a prospective borrower in 2012.
The Presidents/Chief Executives

All Banks/ DFIs/NBFCs

Dear Sirs/Madam,

GROUP LIABILITIES IN THE CIB REPORTS

The definition of “Group” for the purpose of CIB report as notified vide Circular No. SBP/CIB-23/94 dated August 22, 1994 has been reviewed. It has been decided that in the CIB report, the grouping of borrowers shall now be done on the basis of following criteria:

a) The names of the “Group” companies shall be reported by the financial institutions according to the definition of “Group” as contained in Definition No. 14 of the Prudential Regulations (PRs) for Corporate /Commercial Banking. Thus, the onus for correct formation of the group as per definition given in the Prudential Regulations will be on Banks/DFIs/NBFCs.

b) Banks/DFIs/NBFCs will ensure that while determining the group relationships in terms of criteria prescribed in the PRs, they should not consider the foreign national directors, directors of companies under liquidation, and nominee directors of the following entities/agencies:
   - Foreign Controlled Entities.
   - Banks/DFIs.
   - Public Sector Enterprises.
   - Federal/Provincial Government.
   - Private Sector Enterprises’ nominee directors on the Board of Public Sector Enterprises.

c) The definition of the Group as contained in PRs shall, however, be not applicable in the context of Government owned / controlled entities notwithstanding the fact these are listed or unlisted.

2) Since reflection of negative information in the credit report of any party adversely impacts its relationships with its lending institutions, therefore, Banks/DFIs/NBFCs are advised to be very careful while reporting the names of group entities in the CIB data. In case any party disputes the group relationship, the reporting Bank/DFI/NBFC should be able to defend its position with documentary evidence.

3) The above changes in the grouping criteria of companies have necessitated collection of certain additional data from the financial institutions. Therefore, existing formats of CIB data collection (viz. CIB-I, II and III) circulated vide BSD Circular No.4 dated 25th February, 2003 have been revised. The revised formats for CIB I, II and III are enclosed.
as Annexure-I. Banks/DFIs/NBFCs are advised to start collecting the additional information called for in the above formats from their clients and form the groups as per definition of the "Group" given in Prudential Regulations. The SBP will start collecting the monthly CIB data on the revised formats after revising its data capturing software application, for which banks/DFIs/NBFCs will be advised in due course.

Please acknowledge receipt.

Yours faithfully

(Jameel Ahmad)

Director
Chapter 2

Efficient Public Good Provision in Networks: Revisiting the Lindahl Solution
2.1 Introduction

Questions concerning the provision of public goods are central to development research. A large literature describes the problems of public good provision, especially in rural environments. For example, it has been argued that individuals with similar levels of private consumption but differing levels of public goods such as clean drinking water or effective medical care will experience vastly different qualities of life (Besley and Ghatak [2006]).

Public goods and externalities have been a fundamental topic in economics since Pigou [1920]'s landmark contribution in which he described how taxation can improve welfare. In a later article, Hardin [1997], suggested that public goods such as common lands were overutilised by the community since private individuals did not internalise the social cost when deciding how many cattle to graze. In recent years, Seabright [1993] and Dasgupta [2009] have argued that social institutions and norms have some capacity for aligning private and social incentives in the absence of explicit property rights.¹

This paper is focused on situations where the provision of public goods (or abatement of public bads) confers multilateral and heterogeneous benefits across players and has heterogeneous costs for each player. We can think of many such examples in a development setting - for example, Foster and Rosenzweig [1995] and Bandiera and Rasul [2006] consider the case of farmers learning about technology as a public good which has positive externalities across farmers.

Consider a typical village. Soil types and wealth levels vary greatly leading to different varieties of crops being grown. A farmer trialing a new hybrid seed would only offer learning advantages to other villagers who would plant the same seed. Also the magnitude of the benefit from learning would be proportional to the amount of land available to each farmer. The ability to observe the information may depend on the geographic and social distance between farmers. This suggests that the benefits will be heterogeneous across farmers depending on the capacity to update given the new information. The model outlined in section 2.2 is flexible enough to allow these heterogeneous benefits across farmers.

To give a concrete example, suppose farmer A grows only soybeans and wheat, farmer B grows soybeans and rice, and farmer C grows wheat, rice and soybeans. If farmer A trials new soybean seeds, it will lead to benefits to both farmers B and C. But, if farmer C trials wheat varieties it will only lead to benefits for farmer A. The concept of heterogeneous and multilateral benefits explicitly modelled in the paper may lead to interesting effects. For instance, changes in farmer B's experimentation with rice cultivation may affect farmer C's choice of cultivation of wheat, which in turn may affect farmer A's choice of soybean cultivation - even though farmer A does not directly benefit from farmer B's initial experimentation with rice. This will be explicitly considered in the equilibrium concept.

Water resources in farming villages offer yet another example of public goods in development which have

---

¹ For instance, Dasgupta [2009] describes punishment, social punishment within communities "His tribe traditionally practiced a form of punishment that involved spearing the thigh muscle of the errant party. When I asked him what would happen if the party obliged to spear an errant party were to balk at doing so, the young man's reply was that he in turn would have been speared. When I asked him what would happen if the person obliged to spear the latter miscreant were to balk, he replied that he too would have been speared!"
these multilateral and heterogeneous benefits. A shortage of water - a crucial input into farming - will have differing impacts on farmers according to the crop, soil and acreage owned by the farmer. Farmers have access to many water resource tools such as land levelling, irrigation and even choice of crop.

During periods of water shortage, each farmer will have the ability to manage their usage of water at differing costs of implementation (for instance, a farmer on very hard soil will find it very costly to level). There will be both heterogeneous benefits (a farmer with water intensive crops will benefit greatly from a reduction in the water shortage) and multilateral benefits. Ostrom [1992] and Bardhan [2000] have described in detail the heterogeneous effects of water access both at a local and international level.

The above examples demonstrate how public goods can be important in villages and have differential impacts and costs across the village. This leads to two key policy questions this paper is interested in trying to answer:

1. How will a given intervention in a village change the provision of public goods?
2. If we wanted to increase the provision of certain public goods, what are the optimal interventions for addressing this?

To answer these fundamental policy questions we need a theory of public good provision in a development setting.

Besley and Ghatak [2006] describe how the collective action problem is difficult to overcome both at a governmental level and a decentralised level. However, they do describe theoretical conditions under which some of the difficulties in public good provision may be mitigated. This can occur in particular, if: (i) interactions are more likely to be repeated since those who refuse can more easily be punished, (ii) information is good so that individuals’ actions to assist in public good provision can be observed and (iii) there is a strong social structure that can be used to ostracize individuals or can be used to withdraw other forms of cooperation. These conditions are more likely to be satisfied in close-knit villages.

The framework of this paper will assume theoretical conditions under which public good provision may be successful. Based on such a framework, we will study the levels of public goods provided when the benefits of public goods are heterogeneous across agents and there is a network of benefit flows.

This article studies the Lindahl equilibrium - a classical solution in a public goods game which is on the Pareto frontier. Each agent exerts costly effort to produce a public good which confers heterogeneous benefits to different individuals and the cost of producing these public goods varies across individuals. We use insights from Elliot and Golub [2013a,b] who give a simple characterization of Lindahl equilibrium in terms of a single network centrality condition.\(^2\)

This paper’s first contribution characterises how the network architecture affects the amount of investment in the public good by each individual and subsequently solves for each individual’s welfare conditional on

\(^2\)The Lindahl equilibrium enjoys several other robustness properties. As shown in Elliot and Golub [2013a], it is robust to coalitional deviations in a repeated game setting. Elliot and Golub [2013b] extend their characterization of Lindahl equilibria to demonstrate that it is the unique solution in a repeated game with communication, when marginal (but not inframarginal) costs and benefits are observed.
her place in the network. The paper uses a flexible parametric setting, making the characterization of the Lindahl equilibrium more explicit and interpretable than in Elliot and Golub [2013a].

Second, the paper demonstrates the Lindahl equilibrium is unique. Banerjee et al. [2007] argue that "strong enough coordination mechanisms can make almost any group outcome implementable. We believe a micro-founded theory of such coordination is required to make this approach interesting and sharpen its predictive power, and we are not aware of any such theory". We fill this gap in the collective action literature since the Lindahl equilibrium is unique, efficient and has bargaining foundations. Since our equilibrium is unique, we are able to offer comparative statics of how changes in the network architecture affects both public good provision and the welfare of each agent in a Lindahl equilibrium. For example: if one individual's costs of public good provision falls, how does the Lindahl equilibrium change? We are able to answer how and who the policy maker should target to increase certain agents' or groups of agents' welfare. Further,

Third, the paper highlights where public good provision may be suboptimal and suggests policy implications to correct such inefficiencies. The paper focuses on the case where individuals may have suboptimal incentives to reduce their marginal cost of effort. The paper does not explicitly model an investment stage but does offer suggestive evidence for the failure to fully internalise the benefits from cost reductions.

The context of public goods in development is already an extensively studied subject, but groups into two main themes.

Political competition for public goods in the presence of finite resources and differing preferences represents one pillar of past research. Previous papers have shown that benefits of public goods do not accrue symmetrically to everyone in a community. This can arise as a result of competition or differing preferences. As Hardin [1997] states, 'successful collective action often entails suppression of another group's interest.' Esteban and Ray [2001] theoretically demonstrate how group size can affect the probability of the successful implementation of a project. Therefore, explicitly modelling the heterogeneous benefits that accrue across individuals is important.

A second consistent theme in the literature is the optimal allocation of decision rights with respect to public good provision. Is it optimal to allocate decision rights centrally at the governmental level - where there may be imperfect information - or decentralise at the local level, where some of the positive externalities across villages may be disregarded? Oates [1972] and Besley and Coate [2003] discuss the importance of decision rights and the respective costs over where to allocate decision rights. This paper adds to this literature by characterising the equilibrium public good provision if the decision rights are decentralised at the village level when there are heterogeneous and multilateral benefits.

Although the paper’s main focus is public goods provision, by relabeling certain variables, we can consider positive spillovers within teams. For instance, assume each individual produces a single service which requires

\[3\] The water resource problem described earlier would be a classic question over where to allocate the decision rights. For example if individual villages could decide over whether to build a dam and other water conservation methods, they may ignore the benefits or costs that accrue to neighbouring villages. Whereas if the government were to decide whether to build a dam it may lack the necessary local knowledge over the optimal location of the dam. The insights from the model presented in this paper will attempt to model the extent of the public good provision within the village setting.

79
different inputs by each agent. The absence of any input will potentially reduce the productivity of the final service and subsequently the value. Hence, the model could be extended to consider the case where each agent provides a service which has heterogeneous benefits across the network, and subsequently characterise how much of the service is provided and what level of services is sustainable.

The remainder of the paper is organized as follows. In section 2.2, we describe the model and the assumptions. Section 2.3 provides details on how individual characteristics determine relative utility. Section 2.4 determines how positive individual shocks affect the network, and section 2.5 determines how an individual is affected by changes to other individuals. Section 2.6 concludes.

2.2 Model

There is a set of agents $N = \{1, 2, \ldots, n\}$ each of whom simultaneously chooses an an effort $a_i \in \mathbb{R}_{>0}$ at a cost of $c_i f(a_i)$ where we assume $f : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ is a continuously differentiable function such that $f'(\cdot) > 0$ and $f''(\cdot) > 0$. Therefore, we assume each agent’s cost function is strictly convex in her own effort and there is an individual specific cost $c \in \mathbb{R}_{>0}$ which is possibly heterogeneous across agents.

Each individual potentially benefits from the efforts of his neighbours. In particular we assume an agent benefits from a weighted sum of his peers’ efforts. Therefore, each individual’s single period utility function $U_i : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ is:

$$U_i(a_t) = \sum_j R_{ij}a_{j,t} - c_i f(a_{i,t}) \quad \forall i$$

where $R_{ij} \in \mathbb{R}_{\geq 0}$ for all $i \neq j$ and $R_{ii} = 0$ for all $i$. The repeated game is one in which each agent takes a possibly random effort $a_{i,t}$ in each of infinitely many discrete periods $t$ and payoffs are given by $u_i = \sum_{t=0}^{\infty} \delta^t U_i(a_t)$. The game is one of complete and symmetric information.

We do not place any restriction on $\sum_j R_{ij}$ being the same across $i$; thereby allowing for different individual and total benefits across individuals. We denote the $n \times n$ matrix of benefits as $R$.4

We can think of $R$ as a network of benefit flows, which is depicted in Figure (2-1). The thickness of the arrows demonstrates the importance of agent $j$’s effort for agent $i$’s utility ($R_{ij}$). Figure (2-1) demonstrates how the benefits can be both heterogenous and multilateral across agents. For instance an increase in agent $C$’s effort would confer benefits to agents $A$ and $H$, with the benefits $A$ receives greater than that of agent $H$.

In the case of farmer learning, we are allowing that some farmers may benefit more from other farmers’ experimentation. Additionally, by using different cost parameters, we are allowing it to be more costly for some farmers to invest in learning about new crops - for instance, we may think that learning about new crops is risky, and it is more costly for poorer farmers to insure themselves.

As a technical aside, we note that the above functional form actually represents a more general setting.

---

4It can be noted that each individual’s utility function can be normalised such that $c_i = 1$; however, it will be easier for building intuition for future results to retain the non-normalised utility function.
Consider the following utility function:

\[ U_i(a) = \sum_j R_{ij} h_j(a_j) - g_i(a_i) \quad \forall i \]  

(2.1)

Where \( h_j(.) \) are continuously differentiable concave functions and \( g_i(.) \) are continuously differentiable convex functions. Consider the following transformation of the effort space: \( \tilde{a}_i = h_i^{-1}(a_i) \) and \( \tilde{g}_i(a) = g_i(h_i^{-1}(a_i)) \). This transformation maps equation (2.1) into the basic model with linear benefits, without losing convexity in the cost function. Therefore, the utility function under consideration is both relatively general and tractable.

2.2.1 Assumptions on stage game utilities

A few further technical assumptions are necessary to ensure the existence of a Lindahl (centrality-stable) equilibrium.

Assumption 1 (Sufficiently costly effort). \( f'(a) > \max_i \sum_{j} \frac{R_{ij}}{c_i} \) for some \( a \in \mathbb{R}^+_0 \).

Assumption 1 (Sufficiently cheap effort). \( f'(a) < \min_i \sum_{j} \frac{R_{ij}}{c_i} \) for some \( a \in \mathbb{R}^+_0 \).

Assumption 3 (Connectedness of the benefit flow). If \( a \in \mathbb{R}^+_0 \) and \( M \) is a nonempty proper subset of \( N \), then there exist \( i \in M \) and \( j \notin M \) so that \( R_{ij} > 0 \).

Assumption 1 is required to ensure that we are able to bound the efforts of each agent at an equilibrium effort level.

Assumption 2 is required to ensure that individuals are willing to provide some positive effort.

Assumption 3 is not a particularly restrictive assumption; it is used to ensure that a key matrix is irreducible. If this assumption were to fail, we could conduct the analysis separately on each component.
Key definitions

Following Elliot and Golub [2013a] we define the Jacobian $J(a)$ at effort levels $a$ to be the $n \times n$ matrix whose $(i, j)$ element is:

$$J_{ij}(a) = \frac{\partial U_i}{\partial a_j}(a) = \begin{cases} R_{ij} & \text{if } i \neq j \\ -c_i f'(a_i) & \text{if } i = j \end{cases}$$

**Definition** (Elliot and Golub [2013a]). An effort vector $a \in \mathbb{R}^n$ is centrality-stable if $a \neq 0$ and $J(a)a = 0$.

The intuition behind this condition is that if we define each individual's contribution as $J_{ii}(a)a_i = -c_i f'(a_i)a_i$ (i.e. the marginal cost of their effort multiplied by the amount of the effort undertaken) this will equal a weighted sum of other agents' efforts, weighted by how much those efforts benefit agent $i$.

Therefore, at a centrality-stable outcome:

$$-J_{ii}(a)a_i = \sum_j J_{ij}(a)a_j$$

and

$$c_i f'(a_i)a_i = \sum_j R_{ij}a_j$$

Elliot and Golub demonstrate there is a Lindahl interpretation of this result. They show that in this model, centrality-stable outcomes are exactly the Lindahl equilibria.

### 2.2.2 Lindahl equilibrium

**Definition 4.** A Lindahl equilibrium is a pair $(a, P)$, where $a \in \mathbb{R}^n_{\geq 0}$, and $P$ is a matrix of prices, such that the following conditions hold:

(i) $P_{ij} \geq 0$ for $i \neq j$ (where $P_{ij}$ can be interpreted as the price $i$ pays for $j$'s effort)

(ii) $\forall j P_{jj} = -\sum_{i \neq j} P_{ij}$ ($j$'s wage is the sum of the prices each $i$ pays $j$)

(iii) $a^* \in \arg \max_a U_i(a)$ subject to $\sum_j P_{ij}a_j \leq 0$

Elliot and Golub [2013a] show that $a$ is a centrality-stable point if and only if there is a matrix of prices $(P)$ such that $(a, P)$ is a Lindahl equilibrium. These prices are constructed as follows: $P_{ij} = \gamma_i J_{ij}(a) = \gamma_i R_{ij}$ for all $i \neq j$, where $\gamma$ is a row vector such that $\gamma J(a) = 0$. Therefore, in this context each agent $i$ 'pays' a price to individual $j$ proportional to the marginal benefit ($R_{ij}$) he receives from individual $j$'s effort. Agent $i$ is 'paid' a price from each farmer in proportion to the marginal benefit ($R_{ji}$) she receives from agent $i$'s effort.

A further point to note regarding this solution concept: it is robust to coalitional deviations. More precisely, with patient players (sufficiently high $\delta < 1$) it can be enforced by an equilibrium in which, if an agent or coalition of agents does not put in the equilibrium level of effort, then the deviating coalition will
not receive any future benefits from those outside the coalition.

Therefore, the centrality-stable equilibrium concept has many appealing features. Not only does it offer a tractable equilibrium condition, it is robust to coalitional deviations, and is able to be sustained in a theoretical Lindahl equilibrium.

**Proposition 5.** Under the assumptions in section 2.2.1, the following statements hold:

1. If \( a \in R'^n_0 \) is centrality-stable, then \( a \) is sustainable.\(^5\)
2. There exists a centrality-stable \( a \in R'^n_0 \)
3. If \( a \in R'^n_0 \) is sustainable, then \( a \) is Pareto-efficient
4. The centrality-stable \( a \in R'^n_0 \) is unique and given by:\(^6\)

\[
\sum_k R_{ik} a_k = c_i f'(a_i) a_i \quad \forall i \\
Ra = c \circ f'(a) \circ a
\]  

**Proof.** Statements (1)-(3) hold directly from Elliot and Golub [2013a], and statement (4) is proven using a contraction mapping theorem. The proof is provided in appendix 2.7.1. \(\square\)

By proving that uniqueness holds in our environment we are able to firstly consider how the equilibrium changes as we alter the parameters of the model. Secondly, as we have proven uniqueness using a contraction mapping theorem we are able to efficiently compute the equilibrium for any functional form which satisfies our assumptions (Stokey et al. [1989]).

### 2.3 Individual Characteristics

Understanding how the heterogeneous and multilateral marginal benefits (which can be thought of as a network of benefit flows) affect different outcomes is complex. One way of characterizing the impact of the network is to look at two almost identical agents - clones - and compare how the network and the equilibrium affects their welfare.

**Definition 5.** Agent \( i \) is an unconnected clone of agent \( j \) if the following conditions hold:

1. Individuals \( i \) and \( j \) receive the same marginal benefits from all other agents in the network (\( R_{ik} = R_{jk} \) for all \( k \neq i \) or \( j \))
2. Individuals \( i \) and \( j \) have no reciprocal connections (\( R_{ji} = R_{ij} = 0 \))

\(^5\)An effort vector \( a \) is sustainable if there is a \( \delta < 1 \) so that if \( \delta > \delta \), there is a strong Nash equilibrium \( \sigma \) of the repeated game, which is also a subgame-perfect Nash Equilibrium, in which the infinite repetition of \( a \) occurs on the path of play.

\(^6\)The mathematical operator, \( \circ \), is called the Hadamard product. It is a binary operation that takes two matrices of the same dimensions, and produces another matrix where each element \( ij \) is the product of elements \( ij \) of the original two matrices.
Figure (2-5) in appendix (2.8.1) shows an example of identical clones.

Using the definition of clones allows us to examine how the Lindahl equilibrium affects similar individuals in the village.

**Proposition 6.** Under the assumptions in section 2.2.1; if individual \(i\) is an unconnected clone of agent \(j\) and has a higher cost parameter \((c_i > c_j)\), then individual \(i\) will choose a strictly lower effort \((a_i < a_j)\) in the unique Lindahl equilibrium.

The proof of this result follows from observing that each agent receives the same marginal benefits. Accordingly, the individual with the lower cost parameter will choose the greater effort level (formally shown in appendix 2.7.2) in the unique Lindahl equilibrium.

**Definition 6.** The elasticity of total cost with respect to effort is defined as:

\[
e(\alpha) = \frac{\partial c_f(\alpha)}{\partial \alpha} \frac{a}{c_i f(\alpha)} = \frac{\partial f(\alpha)}{\partial \alpha} \frac{a}{f(\alpha)}
\]

The elasticity of total cost with respect to effort is the percentage change in total costs divided by the percentage change in effort levels.

If the elasticity of total cost is increasing in the level of effort then the marginal cost \((\frac{\partial f(\alpha)}{\partial \alpha})\) divided by the average cost \((\frac{f(\alpha)}{a})\) is increasing in effort.

**Proposition 7.** Under the assumptions in section 2.2.1, if the elasticity of total costs is decreasing in effort and individual \(i\) is an unconnected clone of agent \(j\) with a higher cost parameter \((c_i > c_j)\), then agent \(i\) will be better off \((U_i(\alpha) > U_j(\alpha))\) in the unique Lindahl equilibrium.

The proof follows from Proposition (6) and the fact that each individual receives the same benefits through the network, therefore the difference in each individual's welfare will be a function of her equilibrium effort level and cost function. The proof is in appendix 2.7.3.

The effect of elasticity decreasing in \(a\) can also be shown:

**Proposition 8.** Under the assumptions in section 2.2.1, if the elasticity of total costs is increasing in effort and individual \(i\) is an unconnected clone of agent \(j\) with a higher cost parameter \((c_i > c_j)\), then agent \(i\) will be worse off \((U_i(\alpha) < U_j(\alpha))\) in the unique Lindahl equilibrium.

**Proof.** Proof is symmetric to the previous theorem and is omitted. \(\square\)

In the case where agent \(i\) is an unconnected clone of agent \(j\), we can summarize the above results as follows:

\[
c_i > c_j \Leftrightarrow a_i < a_j \Leftrightarrow \begin{cases}  
U_i < U_j & \text{if } \epsilon(\alpha) \text{ is increasing in } a \\
U_i > U_j & \text{if } \epsilon(\alpha) \text{ is decreasing in } a
\end{cases}
\]
Figure 2-2 helps to build the intuition for this result. The graph shows a function which exhibits decreasing elasticity in \( a \). Note individual \( i \)'s cost for a given effort level is always greater than that of agent \( j \). If agent \( i \) is an unconnected clone of agent \( j \) then the total benefits must be the same for the two individuals. Following from equation (2.2) this implies:

\[
\begin{align*}
c_i f'(a_i) a_i &= c_j f'(a_j) a_j \\
\end{align*}
\]  

Equation (2.3) requires that the marginal cost multiplied by the effort level must be equal across unconnected clones in the unique Lindahl equilibrium. Hence area 'A' and area 'B' in figure 2-2 must be equal at the equilibrium effort levels \( a_i \) and \( a_j \). Although agent \( i \) has a higher cost parameter \( c_i \), and consequently higher marginal costs at the equilibrium effort \( a_i \), agent \( i \)'s total costs \((c_i f(a_i))\) are lower than agent \( j \)'s. Since we have already shown that the total benefits for \( i \) and \( j \) must be the same, agent \( i \) must be strictly better off than agent \( j \) in the unique Lindahl equilibrium.

The intuition behind this result stems from the equilibrium condition being determined by marginal costs and benefits. If \( c(a) \) is decreasing in \( a \), the proportional rise in marginal costs is less than that of average costs. It follows that if agent \( i \) has a higher cost parameter \((c_i)\) than agent \( j \), then her marginal costs will always be greater than that of agent \( j \). Agent \( i \) chooses a lower equilibrium effort \( a_i \) which has higher marginal costs - but the proportional rise in marginal costs is less than the rise in average costs. Thus, agent \( i \)'s total costs are lower and agent \( i \) is better off.

**Definition 7.** Agent \( i \) is a connected clone of agent \( j \) if the following conditions hold\( ^7 \):

1. Individuals \( i \) and \( j \) receive the same marginal benefits from all other agents in the network \((R_{ik} = R_{jk} \quad \text{for all} \quad k \neq i \text{ or } j)\)
2. Individuals \( i \) and \( j \) have equal reciprocal connections \((R_{ji} = R_{ij} = K > 0)\)

Figure (2-6) in appendix (2.8.2) shows an example of connected clones.

**Proposition 9.** Under the assumptions in section 2.2.1, suppose the elasticity of total costs is decreasing or constant in effort and individual \( i \) is a connected clone of agent \( j \) with higher individual costs \((c_i > c_j)\), then agent \( i \) will be better off \((U_i(a) > U_j(a))\) in the unique Lindahl equilibrium.

The proof of this result follows from the proposition (7) and is formally provided in appendix 2.7.4. There are two effects when there are reciprocal benefits between agents; (i) the direct effect - person \( j \) exerts higher effort than person \( i \), hence person \( i \) receives more benefits from his neighborhood and (ii) the effect from the elasticity of the cost function leading to lower total costs. Therefore, the direct effect just enhances the effect mentioned in proposition (7).

**Remark 1.** Suppose the elasticity of total costs is increasing in effort and individual \( i \) is a connected clone of agent \( j \) with higher individual costs \((c_i > c_j)\), this does not imply that agent \( i \) will be worse off \((U_i(a) < U_j(a))\). The only difference between a connected and unconnected clones is whether there are reciprocal benefits between the clones.
\(U_j(a)\) in the unique Lindahl equilibrium. There are two effects when agents are connected clones: (i) the direct effect that person \(j\) exerts greater effort and (ii) agent \(i\)'s total costs are higher due to the elasticity increasing in \(a\). Hence, agent \(i\) receives more benefits but also incurs higher costs, therefore whether agent \(i\) is better off overall is ambiguous.

Propositions (7) - (9) describe how an individual's characteristics affect both her equilibrium effort level and utility. It is not necessarily true that individuals who have cheaper technology are better off and in particular it depends on the technology environment. Therefore, if we have two different farmers who receive the same benefit from their neighbours' investments, the farmer for whom it is more costly to invest in learning about new methods may be better off.

**Definition 8.** Agent \(i\) is an identical clone of agent \(j\) if the following conditions hold:

1. The cost parameter for individuals \(i\) and \(j\) are the same: \(c_i = c_j = c\)
2. Individuals \(i\) and \(j\) receive the same marginal benefits from all other agents in the network \((R_{ik} = R_{jk}\) for all \(k \neq i \text{ or } j)\)
3. Individuals \(i\) and \(j\) confer the same marginal benefits to all other agents in the network \((R_{ki} = R_{kj}\) for all \(k \neq i \text{ or } j)\)
4. Individuals \(i\) and \(j\) have equal reciprocal connections \((R_{ji} = R_{ij} = K \geq 0)\)

Figure (2-7) in appendix (2.8.3) shows an example of identical clones.

**Corollary 1.** Under the assumptions in section 2.2.1, suppose the elasticity of total costs is decreasing in effort and individual \(i\) is an identical clone of agent \(j\). If agent \(i\)'s cost was to fall then agent \(j\) would be better off than agent \(i\) \((U_j(a) > U_i(a))\) in the unique Lindahl equilibrium.

The proof follows from propositions (7) and (9). If the elasticity of total costs is decreasing in effort, we know from propositions (7) and (9) that the agent with higher costs will be better off. When the agents have the same cost and confer the same marginal benefits to all other agents in the network we also know that a reduction in either of the cost functions will lead to the same equilibrium effort vector \((a)\) in the unique Lindahl equilibrium. Therefore, if agent \(i\) could choose between reducing her own cost or farmer \(j\)'s cost, she would always choose to reduce farmer \(j\)'s cost.

Consequently, farmer \(i\) has a greater incentive to reduce \(c_j\) than her own cost parameter \(c_i\). This is a startling result and follows from: when the elasticity of total costs is decreasing in effort for a given effort level of their neighbours, each farmer would prefer to have a higher cost parameter. We have not modelled

---

8 We have made the additional assumptions ‘individuals \(i\) and \(j\) confer the same marginal benefits to all other agents in the network \((R_{ki} = R_{kj}\) for all \(k \neq i \text{ or } j)\)’ and ‘\(c_i = c_j\)’ in our result because although we know that the farmer with higher costs would be better off than the farmer with lower costs, we do not know how the other individuals effort levels are affected in equilibrium. Therefore by assuming the two additional assumptions above we can explicitly say that regardless of whose costs are reduced everyone else's equilibrium effort level would be the same.

---

86
a two stage game, therefore, this result is more suggestive of how even if we have Pareto optimality in the
stage game, initial incentives to invest in cost reducing technology may be distorted.

2.4 Network-wide Characteristics

The paper can also examine how changes in the network can affect individual effort levels and welfare. Firstly
a result pinning down the indirect utility function for each agent will be useful for explaining the intuition
and the proofs.

**Definition 9.** An agent’s indirect utility function can be characterised by \( U_i(c_i, a_i^*(R, c)) = U_i(R, c) \) such that \( Ra = cf(a) \)

**Lemma 1.** An agent’s indirect utility function is strictly increasing in the equilibrium \( a_i \)

The proof is provided in appendix 2.7.5. Ultimately, if an farmer’s effort level increases and \( c_i \) stays the
same, then it must be because farmer i’s benefits have increased. The increase from the benefits will always
be greater than the rise in costs.

**Definition 10.** Define the cost vector \( c \) to be less than \( c' \), \( c < c' \), if \( c_i < c_i' \) for all \( i \) and \( c_j < c_j' \) for at least
some one \( j \).

**Definition 11.** Define the matrix \( D(a) \) to be a positive diagonal matrix such that \( D_{ii} \) is the marginal cost for
agent \( i \) to increase her effort level at the equilibrium action \( a \): \( D_{ii}(a) = \frac{\partial}{\partial a_i} [c_i f'(a_i) a_i] = c_i f''(a_i) a_i + c_i f'(a_i) \).

**Definition 12.** Define the matrix \( M(a) \) to be the weighted distance between individual \( i \) and \( j \) at the
equilibrium effort level. \( ^{10} \) Element \( M_{ij}(a) \) is a weighted sum of the number of paths between agent \( i \) and \( j \),
weighted by the marginal benefit \( (R_{ij}) \) of an increase in \( a_j \) for individual \( i \) divided by the marginal cost for
agent \( j \) to increase his effort \( (D_{jj}) \). Therefore element \( M_{ij} \) is useful to describe how important individual \( j \)
is for \( i \) taking into account all possible paths between \( i \) and \( j \) in the unique Lindahl equilibrium.

**Proposition 10.** Under the assumptions in section 2.2.1, an increase in the cost vector \( c' > c \) will necessarily
lead to a decrease in the effort levels for all agents at the the unique Lindahl equilibrium. In particular each
agent’s change in effort level for a small change in the cost vector is equal to:

\[
(a' - a) \simeq -D^{-1}(a)M(a)(c' - c) \circ f'(a) \circ a
\]

The proof is provided in appendix 2.7.6.

In equilibrium, we would expect that an individual’s action would fall as we increase his cost function.
However, it is not immediately obvious that an increase in one individual’s cost will necessarily lead to a
reduction in everyone’s effort in equilibrium.

\(^{10}\) Formally \( M(a) \equiv (I_n - RD(a)^{-1})^{-1} \) and it is shown in Appendix (2.7.6) this can be written as \( M(a) \equiv \sum_{k=0}^{\infty} (RD(a)^{-1})^k \)
Intuitively, if farmer $i$'s costs rise, for a given effort vector of his peers $a_{-i}$, farmer $i$ will reduce his effort level.\footnote{This follows from proposition (6).} If farmer $i$ reduces his effort, any farmer $j$ who receives benefits directly from farmer $i (R_{ji} > 0)$ will reduce his effort since his benefits are reduced. We can now extend the argument to all farmers $k$ who receive benefits from the set of farmers $j$ who are directly connected to farmer $i$. Repeating this chain we can see all farmers reduce their effort levels.

Due to the network of benefit flows, those farmers who confer benefits to farmer $i$ will also reduce their effort levels. Subsequently, there is a multiplier effect from farmer $i$'s initial reduction in effort which flows through the network (the magnitude of this multiplier effect for each agent depends on the matrix $M$). Figures (2-3) and (2-4) graphically demonstrate the intuition. Figure (2-3) shows the initial stylised network of benefit flows and figure (2-4) shows how changes in person $G$’s cost parameter will cause changes through the network. Assume that $G$’s cost parameter was to rise, for a given effort vector for the other agents, $G$’s effort will fall. Therefore, the benefits $A$ receives will fall, leading $A$ to reduce her effort. Following the chain, we know that $F$ will reduce his effort, but then $G$ will reduce her effort further, leading to a new iteration of effort reductions through the network (the multiplier effect).\footnote{In the appendix, we show that the spectral radius of $RD^{-1}$, $\rho(RD^{-1})$, is less than one. Therefore, we can conclude the effect of the cost increase is a converging series.}

Overall, farmer $i$ reduces her equilibrium effort due to two effects: (i) the direct cost effect and (ii) the indirect effect from other farmers reducing their own effort levels leading to lower benefits for farmer $i$.

**Corollary 2.** Under the assumptions in section 2.2.1, a decrease in the cost vector $c' < c$ will necessarily lead to an increase in the effort levels for all agents in the unique Lindahl equilibrium.

**Proof.** The proof is symmetric to the previous proposition and is omitted \hfill $\square$

**Definition 13.** The matrix of benefits $R'$ is greater than $R$ ($R' > R$), if $R'_{ij} > R_{ij}$ for all $ij$ pairs and $R'_{ij} > R_{ij}$ for at least one $ij$ pair.

**Proposition 11.** Under the assumptions in section 2.2.1, an increase in $R'$ from $R$ such that $R' > R$ will necessarily lead to an increase in the effort levels for all agents in the unique Lindahl equilibrium.

The proof is provided in the appendix 2.7.7.

Similar to the intuition for Proposition (10), if farmer $i$ has an increase in his benefit matrix, then farmer $i$’s effort level will increase. Therefore, those farmers who receive benefits from farmer $i$ will also increase their equilibrium effort level. Consequently, a multiplier effect through the network will lead to each agent increasing their equilibrium action in the unique Lindahl equilibrium.

By combining proposition (11) with Lemma (1), we can see an increase in the benefit matrix for any agent will necessarily lead to the increase in utility for all agents - explicitly, the benefits for farmer $i$ from other agents increasing their effort levels outweighs the costs for farmer $i$ increasing her own effort level.

In terms of policy advice, attempting to increase the proliferation of public good provision within a village may be supported by investing in improving network links between agents. For instance, if new hybrid seeds
were invented which were sustainable over many different soil types, this would increase the benefit matrix, which would lead to greater crop experimentation.

Furthermore, we expect that it is easier to share information in villages which have a greater amount of social links between farmers, therefore those villages which have greater social integration may lead to a higher level of public good provision since the benefits from investing would be greater. This suggests improving village institutions to share information may improve each farmer's welfare.

One further point; in this model we consider the benefit matrix to be exogenous, however, if each individual were to privately invest in these links, there would be a suboptimal level of link formation. Since every farmer's utility increases as we improve a link between farmers \( i \) and \( j \), this means there is a positive externality from the introduction of each link on the rest of the village. Hence, the social benefits are greater than the private benefits from link formation leading to the possibility of inefficient link formation. This further supports the possibility that even if we have Pareto efficient actions being undertaken, there may be inefficiencies in the first stage, whether through insufficient incentives to form links or implementing cost reducing technology.

**Proposition 12.** Under the assumptions in section 2.2.1, an increase in the cost vector from \( c \) to \( c' \) such that \( c > c' \) will necessarily lead to a reduction in utility for all agents at the unique Lindahl equilibrium.

The proof is provided in appendix 2.7.8.

This proof shows that even though everyone's equilibrium action falls (so the total costs for each individual falls), the reduction in costs is less than the reduction in benefits for each individual. Secondly, it demonstrates no agent has an incentive to increase their own costs.

Summarising, propositions (7) and (9) argue an individual with higher costs may be better off than an individual with lower costs and proposition (12) argues that if either of their costs were reduced both individuals would be better off.

### 2.5 How the network affects the individual

In this section we will be analysing how each individual's welfare is a function of the entire network. Recall the matrix (definition 12) \( M(a) \) introduced in section 2.4. Element \( M_{ij}(a) \) is useful to describe how important individual \( j \) is for \( i \) taking into account all possible paths between \( i \) and \( j \).

**Proposition 13.** Under the assumptions in section 2.2.1, consider an increase in costs for person \( k \), then the relative impact on person \( i \) and person \( j \) can be characterised by individual fixed effects \( (H_i(a_i)) \) and \( (H_j(a_j)) \), and network effects from person \( k \) to person \( i \) and \( j \) \( ((M_{ik}(a)) \) and \( (M_{jk}(a))) \) in the unique Lindahl equilibrium.

\[
\frac{dU_i}{dc_k} = \frac{H_i(a_i)}{H_j(a_j)} \times \frac{M_{ik}(a)}{M_{jk}(a)} \text{ for } k \neq i \text{ or } j
\]

where \( H_i(a_i) = \frac{f''(a_i)a_i}{f''(a_i)a_i + f'(a_i)} \).
The proof is in appendix 2.7.9.

Therefore, this proposition formally shows how each agent’s relative welfare is affected by a change in farmer k’s costs. Furthermore, if person i and j were exerting the same effort initially, the farmer’s whose utility would be most affected would be the farmer who is ‘closest’ to farmer k in terms of weighted paths.

**Proposition 14.** Under the assumptions in section 2.2.1, consider a change in the benefit matrix \((R_{kj})\) then the change in person i’s welfare can be characterised by an individual fixed effect \((H_i(a_i))\), a network effect from person k to person i \((M_{ik}(a))\) and the action of person j in the unique Lindahl equilibrium.

**Proof.** The change in person i’s utility is equal to:

\[
\frac{dU_i}{dR_{kj}} = \frac{c_i(f''(a_i)a_i)}{c_i(f''(a_i)a_i + f'(a_i))} M_{ik}(a)a_j \quad \text{if } k \neq j
\]

\[
= H_i(a_i) \times M_{ik}(a) \times a_j \quad \text{if } k \neq j
\]

(2.4)

Therefore, if were to consider two different possible link improvements \(R_{kj}\) or \(R_{kp}\) and compare which link addition would benefit farmer i the most, we can observe this is solely dependent on whether farmer j or farmer p puts in the most effort since:

\[
\frac{\frac{dU_i}{dR_{kj}}}{\frac{dU_i}{dR_{kp}}} = \frac{a_j}{a_p} \quad \text{for } k \neq j \text{ or } p
\]

(2.5)

Intuitively, this follows from observing when we improve a link to agent k, agent k’s direct benefit is largest when the link comes from the agent who exerts the highest effort. Since the multiplier effect described in section 2.4 is proportional to the increase in agent k’s effort, agent i prefers for the link to farmer k to induce the greatest rise in his effort and consequently the network.

We can extend this result a little further. Consider the following thought exercise: If we were to improve one link in the network from any agent k to any agent j such that \(k \neq j\), which link should we improve to maximise farmer i’s utility? Equation (2.4) shows that we only to look at two different variables, the action vector (a) and the distance between the recipient of the benefit and farmer i \((M_{ij})\). We conjecture under quite general conditions that \(M_{ij} > M_{ik}\) for all \(k \neq i\). In this case, if we wanted to maximise farmer i’s utility from improving a benefit link, we should add a benefit link from the farmer j who has the highest effort level to farmer i, where \(j \neq i\).

2.6 Conclusion

Close-knit villages are characterised by repeated interaction, and a large, complex social network which provides information and potential social sanctions. These three key features of village economies suggests
that a cooperative equilibrium which is coalitionally stable may occur. The Lindahl equilibrium in our model is a useful benchmark for these villages for four reasons: it is unique, robust to coalitional deviations, has strategic foundations, and offers a market interpretation whilst simultaneously being characterised by a single network centrality condition. The solution concept has many desirable properties whilst offering a tractable model for analysing the provision of public goods.

This article examines three key features of the network: (i) how an individual's relative welfare in the network depends on their own characteristics, (ii) how individual changes affect the network and (iii) how individual's welfare changes according to the network architecture.

The paper demonstrates how the provision of public goods within a network can change with small changes in the parameters. Additionally, the paper is able to characterise the welfare effects from these changes in the environment.

Although not explicitly modelled, the paper also shows how the optimal provision of public goods may fail. In particular, by showing a cost reduction for agent \(i\) or an improvement in the benefit matrix has positive externalities for all other individuals in the network suggest there are suboptimal incentives for investment in cost reducing technology. Future research focusing on explicitly modelling the incentives for investment in cost reducing technologies would be interesting.

The paper has many further possible extensions. The utility function in this paper is a weighted sum over other individual's actions, therefore there is no complementarity or substitution between different individuals provision of public goods. For instance, we may expect that learning about different crops are substitutes since farmers can only grow a finite number of crops. Understanding how to increase the flexibility of the model whilst still retaining tractability and ascertaining whether uniqueness still holds would be an interesting line for future research.

Another interesting question for future research arises from ascertaining what is the welfare maximising benefit matrix? Assume that there is a fixed number of benefit links and all agents had the same cost parameter, what would be the optimal configuration? We conjecture that due to the strict convexity in the cost function, the optimal configuration would have an equal number of benefit flows for each agent.

Overall, this paper has described how heterogeneous and multilateral benefits affect the equilibrium provision of public goods under assumptions which are likely to be fulfilled in a close-knit villages. We have characterised the Lindahl equilibrium and described how the equilibrium changes under certain conditions.
2.7 Proofs

2.7.1 Proof of Proposition 5

Under the assumptions in Section 2.2.1, the following statements hold:

1. If \( a \in R^n_{>0} \) is centrality-stable, then \( a \) is sustainable.

2. There exists a centrality-stable \( a \in R^n_{>0} \).

3. If \( a \in R^n_{>0} \) is sustainable, then \( a \) is Pareto-efficient.

4. The centrality-stable \( a \in R^n_{>0} \) is unique and given by:

\[
\sum_k R_{ik} a_k = c_i f'(a_i) a_i \quad \forall i 
\]

(2.6)

\[
Ra = c \circ f'(a) \circ a
\]

(2.7)

Proof. Statements (1)-(3) hold directly from Elliot and Golub [2013a], and statement (4) is proven using a contraction mapping theorem.

We want to show there is only one solution to equation (2.6). If we can show that the following system of equations is a contraction mapping, then we will have a unique solution:

\[
g(a) = a - B \varphi(a)
\]

(2.8)

where \( B \) is some \( n \times n \) matrix with constant coefficients to be chosen and \( \varphi(a) \equiv c \circ f'(a) \circ a - Ra \). Therefore, the equation \( g(a) = a \) is equivalent to \( \varphi(a) = 0 \).

Formally if \( g : X \to X \) is Lipschitz continuous with Lipschitz constant \( L < 1 \), then \( g(a) \) has a unique fixed point \( a^* \) and thus \( \varphi(a^*) = 0 \). Therefore, we must first show that we can find a mapping \( g \) from a set \( X \) to itself. Then we must show this mapping is a contraction on \( X \).

Let us assume \( a_i \) is bounded above by \( a \) such that \( f'(a) \equiv \max_i \frac{\sum_j R_{ij}}{c_i} = K \) and bounded below by \( a \) such that \( f'(a) \equiv \min_i \frac{\sum_j R_{ij}}{c_i} = K \). Next let us define the set \( X \) to be \( T^n \), where \( T = [\bar{a}, \bar{a}] \). Also, let us assume our matrix \( B \) is a diagonal matrix with strictly positive elements, \( b_i \). To show that our mapping \( g \), maps set \( T^n \) into itself, we must show \( g \) satisfies:

\[
a \geq g_i(a) \geq \bar{a} \quad \forall a \in T^n, \ i
\]

Where we denote the \( i^{th} \) element of \( g(a) \), as \( g_i(a) \).

First we show \( g_i(a) \) is strictly less than or equal to \( \bar{a} \) for all \( a \). This condition, element-wise requires:

\[
\bar{a} \geq a_i - b_i \left[ c_i f'(a_i) a_i - \sum_j R_{ij} a_j \right]
\]

92
Let us define \( \bar{R} \equiv \max_j \sum R_{ij} \), then it must hold that:

\[
\bar{a} \geq a_i - b_i \left[ c_i f'(a_i)a_i - \sum_j R_{ij} a_j \right] \\
\geq a_i - b_i \left[ c_i f'(a_i)a_i - R \right] 
\]

(2.9)

Since inequality (2.9) must hold for all possible values of \( a_i \), let us define \( \hat{a}_i \) such that:

\[
\hat{a}_i = \arg \max_{a_i \in T} a_i - b_i \left[ c_i f'(a_i)a_i - R \right] 
\]

If \( \hat{a}_i \) does not equal \( a \), then \( \hat{a} \) solves the following implicit function:

\[
1 - b_i c_i f''(\hat{a}_i) \hat{a}_i - b_i c_i f'(\hat{a}_i) = 0 
\]

(2.10)

Let us first show the case where \( \hat{a}_i \) is strictly less than \( \bar{a} \). Substituting equation (2.10) into equation (2.9) we have:

\[
\bar{a} \geq \bar{a} + b_i \bar{R} a - \left[ 1 - b_i c_i f'(\hat{a}_i) \hat{a}_i \right] \bar{a}_i 
\]

Rearranging and using \( \hat{a}_i \) is strictly less than \( \bar{a} \), it follows:

\[
1 \geq b_i (\bar{R} + c_i f''(\hat{a}_i) \hat{a}_i) 
\]

Since \( c_i f''(\hat{a}_i) \hat{a}_i \) is finite, we can always find a sufficiently small \( b_i \) such that this condition holds.

If \( \hat{a}_i \) equals \( \bar{a}_i \) then inequality (2.9) becomes:

\[
\bar{a} \geq \bar{a} - b_i \left[ c_i f'(\bar{a}) \bar{a} - \bar{R} \right] \\
1 \geq 1 - b_i \left[ c_i f'(\bar{a}) - \bar{R} \right] \\
c_i f'(\bar{a}) \geq \bar{R} 
\]

This condition is satisfied by our definition that \( f'(\bar{a}) \equiv \max_i \frac{\sum_j R_{ij}}{c_i} \). Therefore we have established \( g_i(a) \) is strictly less than or equal to \( a \) for all \( a \). Next we show that \( g_i(a) \) is strictly greater than or equal to \( a \) for all \( a \). This condition, element-wise requires:

\[
a \leq a_i - b_i \left[ c_i f'(a_i)a_i - \sum_j R_{ij} a_j \right] 
\]

(2.11)

Let us define \( R \equiv \min_j \sum R_{ij} \), then it must hold that:
Where we substitute \((Ra \leq \sum_j R_j a_j)\) into equation (2.11). The LHS of equation 2.12 is minimized at \(a_i = a\), therefore we have to merely show that:

\[
a \leq a - b_i [c_i f'(a_i)a_i - Ra]
\]

Rearranging:

\[
c_i f'(a) \leq R
\]

Which is satisfied by our initial assumption that \(f'(a) \leq \min_k \sum_{c_i} R_{ij}\). Therefore, we have shown for a sufficiently small \(b_i\), the mapping \(g\), maps from \(T^n\) into itself. Now we just need to show that \(g\) is a contraction on \(T^n\).

Differentiating equation (2.8) with respect to \(a\):

\[
g'(a) = I - B \varphi'(a)
\]

Note:

\[
\varphi'(a) = \left[ \begin{array}{cccc}
    c_1 f''(a_1)a_1 + c_1 f'(a_1) & -R_{12} & \cdots & -R_{1n} \\
    -R_{21} & \ddots & \ddots & \vdots \\
    \vdots & \ddots & \ddots & -R_{n-1,n} \\
    -R_{n1} & \cdots & -R_{n,n-1} & c_n f''(a_n)a_n + c_n f'(a_n)
\end{array} \right]
\]

Therefore, if we can find a suitable matrix \(B\) such that:

\[
\| I - B \varphi'(a) \| < 1 \tag{2.13}
\]

then we have contraction mapping.

Then we can write equation \(| I - B \varphi'(a) \| < 1\) element wise as:

\[
\left[ \begin{array}{cccc}
    1 - b_1 (c_1 f''(a_1)a_1 + c_1 f'(a_1)) & b_1 R_{12} & \cdots & b_1 R_{1n} \\
    \vdots & \ddots & \ddots & \vdots \\
    \vdots & \ddots & \ddots & b_{n-1} R_{n-1,n} \\
    b_n R_{n1} & \cdots & 1 - b_n (c_n f''(a_n)a_n + c_n f'(a_n))
\end{array} \right]
\]

Then we can choose any \(b_i\) in the set \(\left( 0, \min \left\{ \frac{1}{\tilde{c}_i}, \min_k \frac{1}{R_{ik}} \right\} \right)\) such that:
\[ |I - Bf'(a)|_{ii} = |1 - b_i (c_i f''(a_i) a_i + c_j f'(a_i))| < 1 \quad \forall i \]
\[ |I - Bf'(a)|_{ij} = |b_i R_{ij}| < 1 \quad \forall i \neq j \]

Hence if we use the sup-norm as our distance metric then for a sufficiently small \( b_i \) equation (2.13) is satisfied. Therefore, for a suitably small \( b_i \) we have shown that \( g \) is a contraction mapping on \( T^n \) and as such, there is unique fixed point \( a^* \) in the set \( T^n \) which satisfies the equilibrium equation (2.6). To complete the proof we show there is no equilibrium \( a^* > \tilde{a} \).

Formally, for all possible \( a \), if at least one \( a_i > \tilde{a} \) then \( a \) is not an equilibrium action profile. We shall prove this by contradiction.

Let us define \( a' \) such that:
\[ a' = \tilde{a} + \varepsilon \]

Assume there is some \( a_k \geq a' \) and \( a_k \geq a_j \) for all \( j \). Then the equilibrium condition requires:
\[ f'(a_k) a_k = \frac{\sum_j R_{kj} a_j}{c_k} \quad \text{for agent } k \]

Let us substitute in \( f'(a) = \max_i \frac{\sum_j R_{ij} a_j}{c_i} \).

\[ f'(a_k) a_k = \frac{\sum_j R_{kj} a_j}{c_k} \quad \text{for agent } k \]
\[ \leq \max_i \frac{\sum_j R_{ij} a_j}{c_k} \]
\[ \leq f'(\tilde{a}) a_k \quad (2.14) \]

Recalling that \( f'(a) \) is increasing in \( a \), then \( f'(\tilde{a}) \) must be strictly less than \( f'(a') \), therefore inequality (2.14) cannot hold. Thereby proving that for all possible \( a \), if at least one \( a_i > \tilde{a} \) then \( a \) is not an equilibrium action profile. A symmetric proof shows that if at least one \( a_i < \tilde{a} \) and \( a \in R^n_{>0} \) then \( a \) is not an equilibrium action profile.

This completes the proof the centrality-stable \( a \in R^n_{>0} \) is unique. \( \square \)

2.7.2 Proof of Proposition 6

Under the assumptions in section 2.2.1; if individual \( i \) is an unconnected clone of agent \( j \) and has a higher cost parameter \( (c_i > c_j) \), then individual \( i \) will choose a strictly lower effort \( (a_i < a_j) \) in the unique Lindahl equilibrium.
Proof. Recall that the equilibrium condition \( J(a)a = 0 \) implies:

\[
\sum_k R_{ik}a_k = c_if'(a_i)a_i \quad \forall i
\] (2.15)

Next, if we assume that \( i \)th and \( j \)th row of the benefit matrix are identical and there are no reciprocal connections, then the total benefits from the network must be the same for both individuals.\(^{13}\) Inserting this into equation (2.15):

\[
\sum_k R_{ik}a_k = \sum_k R_{jk}a_k \Rightarrow c_if'(a_i)a_i = c_jf'(a_j)a_j
\] (2.16)

Without loss of generality assume \( c_i < c_j \). Then using equation (2.16):

\[
\frac{f'(a_i)a_i}{f'(a_j)a_j} = \frac{c_j}{c_i} > 1
\] (2.17)

Recalling that \( f(.) \) is strictly convex, therefore it follows that \( a_i > a_j \) if and only if \( c_i < c_j \). \( \square \)

2.7.3 Proof of Proposition 7

Under the assumptions in section 2.2.1, if the elasticity of total costs is decreasing in effort and individual \( i \) is an unconnected clone of agent \( j \) with a higher cost parameter \( (c_i > c_j) \), then agent \( i \) will be better off \( (U_i(a) > U_j(a)) \) in the unique Lindahl equilibrium.

Proof. Recalling the original utility function \( U_i = \sum_k R_{ik}a_k - c_if(a_i) \) and without loss of generality assume \( c_i < c_j \), then \( i \) is worse off than \( j \) if and only if:

\[
U_i < U_j
\]

\[
\sum_k R_{ik}a_k - c_if(a_i) < \sum_k R_{jk}a_k - c_jf(a_j)
\]

Since by the assumption that two individuals receive the same benefit through the network and there are no reciprocal links we know that \( \sum_k R_{ik}a_k = \sum_k R_{jk}a_k \) then:

\[
U_i < U_j
\]

\[
-c_if(a_i) < -c_jf(a_j)
\] (2.18)

\(^{13}\) The condition that \( i \)th and \( j \)th row of the benefit matrix are identical and there are no reciprocal connections explicitly requires \( R_{ik} = R_{jk} \) for all \( k \neq i \) or \( j \) and \( R_{ij} = R_{ji} = 0 \)
Using equation (2.17) we know that
\[ c_i = \frac{c_j f'(a_j) a_j}{f'(a_i) a_i} \quad (2.19) \]

Sub in equation (2.19) for \( c_i \) in equation (2.18):
\[
\frac{c_j f'(a_j) a_j}{f'(a_i) a_i} f(a_i) > \frac{c_j f(a_j)}{f'(a_i) a_i}
\]
\[
\frac{f'(a_j) a_j}{f(a_j)} > \frac{f'(a_i) a_i}{f(a_i)}
\]

If we define the elasticity of total cost to be 
\[
\epsilon(a_i) = \frac{\partial c_i f(a_i)}{\partial a_i} = \frac{\partial f(a_i)}{\partial a_i} f(a_i)
\]
then:
\[
U_i < U_j \iff \epsilon(a_i) < \epsilon(a_j)
\]

In the previous proposition we showed \( a \) is decreasing in \( c \) hence if the elasticity is decreasing in \( a \) then the agent with higher cost will be better off.

\[ \Box \]

2.7.4 Proof of Proposition 9

Under the assumptions in section 2.2.1, suppose the elasticity of total costs is decreasing or constant in effort and individual \( i \) is a connected clone of agent \( j \) with higher individual costs \((c_i > c_j)\), then agent \( i \) will be better off \((U_i(a) > U_j(a))\) in the unique Lindahl equilibrium.

Proof. Recall that the equilibrium condition \( J(a) a = 0 \) implies:
\[ \sum_k R_{ik} a_k = c_i f'(a_i) a_i \quad \forall i \quad (2.20) \]

Next, if we assume that \( i \)th and \( j \)th row of the benefit matrix are identical and have equal reciprocal connections, then the total benefits from the network must be different only by the extent of their respective actions.\(^{15}\) Therefore, we can write a similar condition to equation (2.16):
\[
\sum_{k \neq j} R_{ik} a_k + K a_j = c_i f'(a_i) a_i
\]
\[
\sum_{k \neq i} R_{jk} a_k + K a_i = c_j f'(a_j) a_j
\]

\(^{14}\)In the proof we have passed terms across inequalities without worrying about changes in the sign of the inequality since all the terms of the expression must be greater than zero - since we have a convex cost function \((so f'(.) > 0)\) and actions must be non-negative.

\(^{15}\)The condition that \( i \)th and \( j \)th row of the benefit matrix are identical and there are equal reciprocal connections explicitly requires \( R_{ik} = R_{jk} \) for all \( k \neq i \) or \( j \) and \( R_{ij} = R_{ji} = K > 0 \)
Hence subbing out for \( \sum_{k \neq i} R_{ik} a_k \) we can combine the two equations into:

\[
\begin{align*}
    c_i f'(a_i) a_i - K a_j &= c_j f'(a_j) a_j - K a_i \\
    c_i f'(a_i) a_i + K a_i &= c_j f'(a_j) a_j + K a_j \\
    a_i [c_i f'(a_i) + K] &= a_j [c_j f'(a_j) + K]
\end{align*}
\]

(2.21)

Assume without loss of generality that \( c_i < c_j \), then it follows that \( a_i > a_j \) due to the strict convexity in \( f(.) \). This can be proven by contradiction, assume \( a_i \leq a_j \) then for equation (2.21) to hold, we need:

\[
\begin{align*}
    c_i f'(a_i) + K &\geq c_j f'(a_j) + K \\
    c_i f'(a_i) &\geq c_j f'(a_j)
\end{align*}
\]

Since we have assumed \( c_i < c_j \) it most hold that:

\[
f'(a_i) > f'(a_j)
\]

However, due to the strict convexity in \( f(.) \) then \( a_i > a_j \) leading to a contradiction. Therefore agents with higher costs put in lower effort when they have equal benefits from the network and reciprocal connections:

\[ c_i < c_j \iff a_i > a_j \]

We are interested in characterising when agent j’s utility is greater than agent i’s, in the case where \( c_i < c_j \), therefore:

\[
\begin{align*}
    U_i < U_j \\
    \sum_{k} R_{ik} a_k - c_i f(a_i) < \sum_{k} R_{jk} a_k - c_j f(a_j) \\
    \sum_{k \neq i} R_{ik} a_k + K a_j - c_i f(a_i) < \sum_{k \neq i} R_{jk} a_k + K a_i - c_j f(a_j)
\end{align*}
\]

Where in the last equation we have used \( R_{ij} = R_{ji} = K > 0 \) to define the reciprocal benefits between agents.

\[
\begin{align*}
    U_i < U_j \\
    K a_j + c_j f(a_j) < K a_i + c_i f(a_i)
\end{align*}
\]
Using equation (2.21) we can sub in for $c_j$:

$$Ka_j + \left[ \frac{c_if'(a_i)a_i + K(a_i - a_j)}{f'(a_j)a_j} \right] f(a_j) < Ka_i + c_i f(a_i)$$

With some rearranging\(^{16}\):

$$K(a_j - a_i) \left[ \frac{f'(a_j)a_j - f(a_j)}{f(a_j)f'(a_j)} \right] < c_i f(a_i) - \left( \frac{c_i f'(a_i)a_i}{f'(a_j)a_j} \right) f(a_j)$$

$$K(a_j - a_i) \left[ \frac{f'(a_j)a_j - f(a_j)}{f(a_i)f(a_j)} \right] < c_i \left( \frac{f'(a_j)a_j}{f(a_j)} \right) - c_i \left( \frac{f'(a_i)a_i}{f(a_i)} \right)$$

(2.22)

Recalling the definition of elasticity of total costs with respect to action is $\epsilon(a_i) = \frac{\partial c_i f(a_i)}{\partial a_i} = c_i f'(a_i)$, then we can write equation (2.22) as:

$$K(a_j - a_i) \left[ \frac{f'(a_j)a_j - f(a_j)}{f(a_i)f'(a_j)} \right] < c_i (\epsilon(a_j) - \epsilon(a_i))$$

$$K(a_i - a_j) \left[ \frac{f'(a_i)a_i - f(a_i)}{f(a_i)f(a_j)} \right] > c_i (\epsilon(a_i) - \epsilon(a_j))$$

Therefore the LHS is greater than zero since due to the strict convexity in $f(\cdot)$ and the assumption $c_i < c_j$, hence if elasticity of total costs with respect to actions is decreasing in the action, then the individual with lower costs is worse off.

Hence:

$$c_i < c_j \Rightarrow a_i > a_j \Rightarrow \left\{ \begin{array}{l} U_i < U_j \text{ if } \epsilon(a)\text{is decreasing in } a \end{array} \right.$$  

\(\square\)

### 2.7.5 Proof of Lemma 1

An agent's indirect $U_i(c_i, a_i(R, c))$ is strictly increasing in the equilibrium $a_i$

**Proof.** Using the equilibrium condition and the strict concavity assumption on $f(\cdot)$ we can show that $U_i(c_i, a_i(R, c))$ is strictly increasing in $a_i$.

Firstly $J(a)a = 0$ implies:

$$\sum_k R_{ik}a^*_k = c_i f'(a^*_i)a^*_i \quad \forall i \quad (2.23)$$

Recall that:

$$U_i = \sum_k R_{ik}a_k - c_i f(a_i)$$

\(^{16}\)Recall that $f(a) > 0$ for all $a > 0$ hence all terms are positive
Substituting in $\sum_k R_{ik}a^*_k$ from equation (2.23) into the utility function:

$$U_i(c_i, a^*_i) = c_i f'(a^*_i)a^*_i - c_i f(a^*_i)$$

$$U_i(c_i, a^*_i) = c_i [f'(a^*_i)a^*_i - f(a^*_i)]$$

Due to $f(.)$ being strictly convex, this function is increasing in $a^*_i$. Formally:

$$\frac{\partial U_i}{\partial a^*_i} = c_i [f''(a^*_i)a^*_i] > 0$$

2.7.6 Proof of Proposition 10

Under the assumptions in section 2.2.1, an increase in the cost vector $c' > c$ will necessarily lead to a decrease in the effort levels for all agents at the the unique Lindahl equilibrium. In particular each agent’s change in effort level for a small change in the cost vector is equal to:

$$(a' - a) \simeq -D^{-1}(a)M(a) (c' - c) \circ f'(a) \circ a$$

Proof. Consider the equilibrium action for all agents:

$$\sum_k R_{ik}a_k = c_i f'(a_i)a_i \ \forall i$$

Define a new parameter $\theta \in [0, 1]$, which $c$ depends on. This implicitly defines a function $a^* : [0, 1] \to \mathbb{R}$, where $a(\theta)$ is the fixed point when $c = c(\theta)$. The parameter $\theta$ is used to model how changes in $\theta$ affect $c(\theta)$ and in turn $a^*(\theta)$.

Writing out the equilibrium condition in the new notation gives us:

$$\sum_k R_{ik}a_k(\theta) = c_i(\theta)f'(a_i(\theta))a_i(\theta) \ \forall i$$

$$Ra(\theta) = c(\theta) \circ f'(a(\theta)) \circ a(\theta)$$

Differentiating this function with respect to $\theta$ gives:

$$Ra'(\theta) = c'(\theta)c'f'(a(\theta)) \circ a(\theta) + c(\theta)c'f''(a(\theta)) \circ a(\theta) \circ a'(\theta) + c \circ f'(a(\theta)) \circ a'(\theta)$$

Note when using Hadamard product, $c \circ f''(a(\theta)) \circ a(\theta) \circ a'(\theta)$ can be rewritten as $C[F''(a(\theta))A(\theta)a'(\theta)]$ where $C = diag(c)$ and $F''(a(\theta)) = diag(f''(a(\theta)))$ and $A(\theta) = diag(a(\theta))$. Therefore, we transform a $n \times 1$
vector into a diagonal $n \times n$ matrix.

Collecting terms in $a'(\theta)$ and rearranging gives$^{17}$:

$$
(R - CF''(a)A - CF'(a))a'(\theta) = c'(\theta) \circ f'(a) \circ a
$$

$$
a'(\theta) = - (CF''(a)A + CF'(a) - R)^{-1} c'(\theta) \circ f'(a) \circ a
$$

(2.24)

Equation (2.24) states how the action profile $a$ changes for changes in the parameter $\theta$. Define $D = CF''(a)A + CF'(a)$ and this can be further reduced to:

$$
a'(\theta) = - [D - R]^{-1} c'(\theta) \circ f'(a) \circ a
$$

$$
a'(\theta) = -D^{-1} [I - RD^{-1}]^{-1} c'(\theta) \circ f'(a) \circ a
$$

(2.25)

If the spectral radius, $\rho(\cdot)$, of $[RD^{-1}]$ is less than one, we can write equation (2.25) as:

$$
a'(\theta) = -D^{-1} \left[ \sum_{k=0}^{\infty} (RD^{-1})^k \right] c'(\theta) \circ f'(a) \circ a
$$

(2.26)

To show $\rho(RD^{-1}) < 1$, we use lemma 1 in Elliott and Golub (2012a), which shows the spectral radius of $\rho(RE^{-1}) = 1$, where $E = CF'(a)$. By noting

$$
D - E = (CF''(a)A + CF'(a)) - (CF'(a)) = CF''(a) > 0
$$

Then the spectral radius:

$$
\rho(RD^{-1}) < \rho(RE^{-1}) = 1
$$

Therefore, the change in the action profile $a'(\theta)$ can be written in the form of equation (2.26). Equation (2.26) shows that an increase in the cost function ($c'(\theta) \geq 0$ for all elements with at least one strict inequality) will lead to a reduction in the action profile for all agents due to all the terms on the RHS are positive and they are multiplied by a negative.

2.7.7 Proof of Proposition 11

Under the assumptions in section 2.2.1, an increase in $R'$ from $R$ such that $R' > R$ will necessarily lead to an increase in the effort levels for all agents in the unique Lindahl equilibrium.

Proof. Similar to the proof of Proposition 2.7.6, define a new parameter $\theta \in [0, 1]$, which $R$ depends on.

\footnote{For ease of exposition, the term $\theta$ has been dropped except from the terms which are differentiated with respect $\theta$}
This implicitly defines a function \( a : [0, 1] \to \mathbb{R} \), where \( a(\theta) \) is the fixed point when \( c = c(\theta) \). The parameter \( \theta \) is used to model how changes in \( \theta \) affect \( R(\theta) \) and in turn \( a(\theta) \). We can show \( a'(\theta) \) is increasing in \( R'(\theta) \).

To be precise, recall the original equilibrium condition:

\[
\sum_k R_{ik}(\theta) a_k^*(\theta) = c_i f'(a_i^*(\theta)) a_i^*(\theta) \quad \forall i
\]

Which we can rewrite as:

\[
R(\theta)a(\theta) = c \circ f'(a(\theta)) \circ a(\theta)
\]

(2.27)

Following the proof of Proposition 2.7.6, we can differentiate equation (2.27) and rearrange such that:

\[
a'(\theta) = \{D - R\}^{-1} R'(\theta) a
\]

Therefore, since the spectral radius, \( \rho(D^{-1}) < 1 \), we can write this as:

\[
a'(\theta) = D \sum_k (RD^{-1})^k R'(\theta) a
\]

Hence, an increase in the intensity of links or the introduction of new links (\( R'(\theta) > 0 \)) in the benefit matrix will lead to an increase in the equilibrium action for all agents.

**2.7.8 Proof of Proposition 12**

Under the assumptions in section 2.2.1, an increase in the cost vector from \( c \) to \( c' \) such that \( c > c' \) will necessarily lead to a reduction in utility for all agents at the unique Lindahl equilibrium.

**Proof.** The proof relies on using lemma 1, proposition 10 and proposition 11. First note that the change in the indirect utility, \( U_i(c_i, a_i(R, c)) \) can be written as:

\[
U_i(c_i, a_i(R, c)) = c_i f'(a_i^*) a_i^* - c_i f(a_i^*)
\]

Differentiating with respect to \( c_k \)

\[
\frac{dU_i}{dc_k} = \sum_{i=1}^{n} [f'(a_i) a_i - f(a_i)] + c_i \left[ f''(a_i) a_i \frac{da_i}{dc_k} \right]
\]

First consider the case \( i \neq k \); the sign \( \left( \frac{dU_i}{dc_k} \right) = \text{sign}(\frac{dU_j}{dc_k}) \), since the other terms are greater than zero.
Hence following proposition 10, which shows actions are decreasing in costs, then:

\[
\frac{dU_i}{dc_k} < 0 \text{ if } k \neq i
\]

To show that:

\[
\frac{dU_i}{dc_i} < 0
\]

Note that a utility function can be transformed with an affine transformation such that an individual's choice of \( a_i \) is unchanged from an increase in \( c_i \) or an appropriately weighted reduction in the \( i \)th row of the benefits matrix.

Formally, consider scaling an individual's cost function by \( \kappa \), therefore the equilibrium will be such that:

\[
\sum_k R_{ik} a_k^* = \kappa c_i f'(a_i^*)a_i^ *
\]

Further we could replicate the same equilibrium \( a^* \) by using \( R'_{ik} = \frac{R_{ik}}{\kappa} \), therefore, it follows the change in person \( i \)'s action from the increase in costs is the same as a reduction in the \( i \)th row of the benefit matrix \( R \).

Hence, following proposition 11 which showed actions decreasing in reductions in the benefit matrix and:

\[
\frac{dU_i}{dR_{ik}} = c_i \left[ f''(a_i)a_i \frac{da_i}{dR_{ik}} \right]
\]

Hence, \( \text{sign} \left( \frac{dU_i}{dR_{ik}} \right) = \text{sign} \left( \frac{da_i}{dR_{ik}} \right) > 0 \). Therefore \( \frac{dU_i}{dc_k} < 0 \).

2.7.9 Proof of Proposition 13

Under the assumptions in section 2.2.1, consider an increase in costs for person \( k \), then the relative impact on person \( i \) and person \( j \) can be characterised by individual fixed effects \( (H_i(a_i) \) and \( H_j(a_j)) \), and network effects from person \( k \) to person \( i \) and \( j \) \( (M_{ik}(a)) \) and \( (M_{jk}(a)) \) in the unique Lindahl equilibrium.

\[
\frac{dU_i}{dc_k} = \frac{H_i(a_i)}{H_j(a_j)} \times \frac{M_{ik}(a)}{M_{jk}(a)} \text{ for } k \neq i \text{ or } j
\]

where \( H_i(a_i) = \frac{f''(a_i)a_i}{f''(a_i)a_i + f'(a_i)} \)

Proof.

\[
\frac{dU_i}{dc_k} = \frac{1}{f''(a_i)a_i + f'(a_i)} \left[ f''(a_i)a_i f'(a_k) \right] \text{ for } i \neq k
\]
Where we have substituted in $\frac{d\alpha_i}{dc_k}$ from the earlier proofs. Therefore defining $H_i = \frac{f''(a_i)\alpha_i}{f''(a_i)\alpha_i + f'(a_i)}$ we can denote this as:

$$\frac{dU_i}{dc_k} = H_i M_{ik} f'(a_k) \alpha_k \text{ if } i \neq k$$

Therefore:

$$\frac{dU_i}{dc_k} = \frac{H_i}{H_j} \times \frac{M_{ik}}{M_{jk}} \text{ for } k \neq i \text{ or } j$$
2.8 Clones

2.8.1 Unconnected clones

Figure 2-5: Agents A and C are unconnected clones in this matrix, since they receive the same marginal benefits from the rest of the network and confer no benefits to each other.

The benefits which agents A and C receive are denoted in red for clarity.
2.8.2 Connected clones

Figure 2-6: Agents $A$ and $C$ are connected clones in this matrix, since they receive the same marginal benefits from the rest of the network and confer equal benefits to each other.

The benefits which agents $A$ and $C$ receive are denoted in red for clarity.
2.8.3 Identical Clones

Figure 2-7: Agents $A$ and $C$ are identical clones in this matrix, since they receive and confer the same marginal benefits from and to the rest of the network.

Agents $A$ and $C$ must the same cost parameter too (not shown in the figure). The benefits which agents $A$ and $C$ receive and confer are denoted in red for clarity.
Figure 2-2: How costs, efforts and utilities are related for unconnected clones $i$ and $j$ if $c_i > c_j$ and elasticity is decreasing in $a$. 

- **Total Costs**
  - $c_i f(a_i)$
  - $c_i f'(a_i)$

- **Marginal Costs**
  - $c_i f'(a_i)$
  - $c_j f'(a_j)$

The graph illustrates the relationship between total costs and marginal costs as a function of effort $a$. The areas $A$ and $B$ represent the cost differences between clones $i$ and $j$.
Figure 2-3: A stylised benefit flow through the network. The arrows demonstrate who confers benefits on whom.

Figure 2-4: The flow of a cost increase on person G through the network.
Chapter 3

The Economics of Platforms in a Walrasian Framework
3.1 Introduction

We are interested in economic platforms which inherently depend on attracting multiple different types of users. For instance, the quality or usefulness of a credit card will depend on the merchants who accept the card and which consumers use the card. Each side cares about the other. A mobile phone network is attractive only if it allows a user to contact her social/business contacts on the same technological platform. A dark pool for the trading of financial instruments needs to attract both buyers and sellers in somewhat proportionate numbers if it is to allow trade and coexist with other public exchanges. Likewise, a clearinghouse is a mechanism to net trades and mitigates obligations that continue beyond an end-of-day settlement period. Even traditional financial intermediaries can be thought of in this way, in the sense that they stand between savers and borrowers and transform the risk and time structure of funds.

We ask, in these types of markets, with multiple competing platforms how does one define a Walrasian equilibrium. Typically does it exist or are there inherent problems? If an equilibrium exists, is it efficient in the allocation of costs or is there a case for the regulation of prices? Finally, what is the relationship between competitive equilibria and the distribution of welfare, specifically, does one side or the other have an inherent advantage?

We use tools from standard General Equilibrium Theory in this modified environment, where there is an obvious externality: a user's willingness to pay for a product is dependent on the composition of the product's user base.

Our paper has three main results: first, building on Prescott and Townsend [2006] who analyzed firms as clubs in general equilibrium, we provide a framework which shows that platforms can internalize the above-described externality, and we characterize the equilibrium among competing platforms: which types of platforms exist, the prices paid by user types, and the fees charged by intermediaries.

Second, we prove that both the first and second welfare theorems hold in this environment; a competitive equilibrium is Pareto optimal and any optimal allocations of resources can be achieved by lump sum taxes and transfers on underlying wealth.

Third, using the framework we characterize how the equilibrium prices for each type of user (for example, buyers and sellers or consumers and merchants) and the composition of a platform's users change as we alter parameters of the underlying economic environment such as decreasing the costs of creating platforms or altering the initial distribution of wealth among the agents. Indeed, we make a distinction between a fundamental type of user versus within-a-type users that differ only in wealth – for instance, we can examine how the equilibrium changes as we alter different consumers' wealth. The latter allows us to see how the advantage of higher wealth plays out in terms of advantageous matches spills over to others' and their own utility.

Our framework builds heavily on club theory and in particular, the firms as club literature. Koopmans and Beckmann [1957] discuss the problem of assigning indivisible plants to a finite number of locations and its link to more general linear assignment/programming problems. A system of rents sustains an optimal
assignment in the sense that the profit from each plant-location pair can be split into an imputed rented to the plant and an imputed rent to the location. At these prices landowners and factory owners would not wish to change the mix of tenants or location. As Koopmans and Beckman point out, the key to this beyond linear programming is Gale et al. [1951]'s theorem which delivers Lagrange multipliers on constraints, every location has a match and the firms and location are not over- or under-subscribed. A linear program ignores the intrinsic indivisibilities, the integer nature of the actual problem, nevertheless achieves the solution.

In the well-known labor assignment model of Sattinger [1993], workers are assigned to jobs and the contribution of a worker with a mix of skills depends not only on the particular type of job being performed but the assignment of others – that is the work(er) environment. Hornstein and Prescott [1993] consider a Lucas [1978] managerial span-of control problem in which agents can choose to be workers or firms – an indivisibility – and also a second problem in which the number of hours a firm operates its plants and the numbers of workers assigned to each plant is endogenous. These environments appear to introduce a non-convexity in the production set. But with a large number of agents one can approximate the environment with a production set that has constant returns, that is when the non-convexity is small relative to the size of the economy. Essentially, the production set becomes a convex cone, as in McKenzie [1959, 1981]'s formulation of general equilibrium.

The economics underlying McKenzie's formulation – in contrast to Arrow and Debreu [1954] – makes endogenous the ownership of shares in firms, i.e. profits must be earned through entrepreneurial rents rather than through shares which are given a priori. The basic tools in Hornstein and Prescott is the use of lotteries as developed in ,Prescott and Townsend [1984] for private information environments in which incentive constraints introduce a non-convexity. The common element is that lotteries are a way at the aggregate level to assign fractions of agent types to contracts, clubs, occupations, and so on, even though assignments are discrete. Likewise, Hansen [1985] and Rogerson [1988] in macro determine the fraction of the population working overtime, a discrete choice. Pawasutipaisit [2010] assigns one male and one female type to common marriage. The firms as clubs methodology is well suited for our setting because it allows us to solve for which platforms form in equilibrium, the size of each platform, and who is part of each platform.

We show that the first and second welfare theorems hold in our economy with platforms. A competitive equilibrium exists, and in this equilibrium, platforms are able to internalise the effects of interdependencies through the composition of users. Each basic user type faces a user price for each of a (continuum) number of potential platforms, which vary in the number of own-type participants and other-type participants. In equilibrium at given prices, the solution to these decentralized problems delivers the mix and number of participants in active platforms that each user anticipated when they choose platforms. That is, in equilibrium the club or platform is populated with user types exactly as anticipated. The solution is efficient because market price for joining a platform, which a user takes as given, changes across platforms in a way which internalizes the marginal effect of altering the composition of the platform. Put different, each agent of each type (having tiny, negligible influence), is buying a bundle which include the composition and number
of total participants, i.e. the commodity space is expanded to include the intrinsic externality feature of the platform. Essentially we solve the externality problem in the way suggested by Meade [1952] and Arrow [1969].

Having established the economy is efficient, we demonstrate how the size of platforms, prices, and individuals’ utility change as we alter parameters of the environment, such as the cost of building a platform, the disutility of having too many users (congestion), and the underlying wealth (endowments of the capital good) of user types. Higher costs and congestion naturally tend to reduce the relative number of equilibrium platforms. But there are distributional aspects as higher costs make capital used to construct platforms more valuable and that favors wealthy agents who are abundantly endowed with that capital. The poor are thus hurt in this comparison. A change in the wealth distribution toward a favored type not only increases the competitively determined utility of that type, it also changes the utility of others, i.e. potentially increasing the utility of those that favored types wish to be matched with and decrease others with lower wealth who are in competition to populate platforms. We exploit in these latter comparative statics the fact that changing Pareto weights is equivalent with changing wealth, i.e. we use a programming problem to maximize Pareto weighed sums of utilities and then change the weights, tracing out all Pareto optimal. A given optimum requires lump sum taxes and transfers or equivalently a change in the initial underlying distribution of wealth.

The closest literature to our work is on two-sided markets. Two-sided markets consider platforms who sell to at least two different user groups, whose utility is dependent on who else uses the platform. Two-sided markets is an industrial organization, partial equilibrium framework. The main findings in Rochet and Tirole [2003, 2006], Weyl [2010], Rysman [2009] is that two-sided markets lead to market failure. In particular, in the two-sided market literature a key concern is how the distribution/impact of users’ fees will cover the platform’s fixed and marginal costs. There are many controversies: whether the allocation of fees alters the outcome (one definition of a two sided market is that the distribution of fees matters to the outcome); whether there are implicit subsidies; whether users are, or should be aware of what the price is covering (should payment charges be a separate part of the bill), and how to regulate the interchange fee that the issuing bank charges redemption banks (and again how fees are passed to merchants – i.e. small merchants vs block entities such as Walmart). In contrast, our work is in a general equilibrium environment with a Walrasian allocation mechanism in which net prices are appropriate and outcomes are shown to be efficient.

The literature on middlemen is also related to our work. Middlemen facilitate trade between two different agents. Rubinstein and Wolinsky [1987]’s seminal paper outlines a model where middlemen increase the efficiency of the market through the reduction of search costs for buyers and sellers. Our paper concentrates on how intermediaries are platforms which facilitate trade between different parties. In contrast to the search and middlemen literature, we allow a competitive, constant returns to scale intermediary sector with free entry, a large (continuum) number of agents of each types, and no search frictions.

Indeed we need to emphasize the limitations of what we are doing, specifically what we are not doing.
We do not consider agents having any pricing power. We do not consider the problem of establishing new products/platforms in the sense of innovation and entry into an existing equilibrium outcome and the problem of changing client expectations. Related, we do not discuss the historical development of platforms nor consider current regulatory restrictions, including well intended but potentially misguided regulations which may limit our ideal market design. Nor do we model monopolistic competition though we do allow our platforms to be configured with different composition of customers, so there is clear product differentiation (just no market power). Finally, we do not allow ever increasing economies of scale in platform size.

3.1.1 Applications

Our paper is relevant to many different environments which involve the intermediation between different agents who simultaneously have preferences about the size and composition of the intermediary’s users.

Stock Exchanges and Dark Pools

‘Dark Pools’ are collections of buyers and sellers which join a platform as a mechanism to anonymously trade bonds and stocks. There are many seemingly similar platforms (The Economist [2011] refers to at least 80 dark pools as of August 2011,) and they are increasingly a larger proportion of equity trading (the percentage of consolidated U.S. equity trading on dark pools is estimated to have rose from 6.5% in 2008 to 12% in 2011, Tabb Group [2012]). Thus the industry for exchanges is competitive and large, i.e. not concentrated; relatedly, O’Hara and Ye [2011] argue, controlling for sample selection issues, that new equity exchanges (not necessarily dark pools) have lower bid ask spreads and faster execution times.\footnote{O’Hara and Ye [2011] estimate those stocks which are listed on multiple exchanges have trading costs which are 0.33-0.34 cents lower than those which are listed on only the major exchanges.} Zhu [2013] distinguishes three types of dark pools: in the first, platforms match customer order without own account trading, in the second platforms are operated by broker dealers who operate as continuous non displayed limit order books; and in the third, dark platforms act as fast electronic market makers that instantaneous accept or reject incoming orders. In Zhu’s model, the likelihood of being to execute your trade depends on the availability of counterparties, in particular, the side with more orders – will fail to be executed. Our main point is obvious; for a platform to be successful it must attract both buyers and sellers for the potential transaction.

Clearinghouses

Clearinghouses (and depository trusts) net trades across platform participants, often at the end of each business day. A clearinghouse can increase the efficiency of trading between users by guaranteeing trades amongst its users.\footnote{In particular, the clearinghouses introduce two new contracts, whereby each party now bilaterally trades with the clearinghouse rather than each other in a process called ‘novation’.} This may lead to three main advantages: first, it can reduce a user’s exposure through netting their total obligations across multilateral counterparties; second, it can reduce systematic risk by reducing the probability of defaults propagating across counterparties; and third it can increase the speed of transactions. For a clearinghouse to effectively settle and process trades, it must attract multiple buyers

114
and sellers to its platform.

Credit Cards, Debit Cards, and other Payment Systems

Consumers and merchants use an array of different platforms to transfer cash or value: credit cards, debit cards, cash, checks, interbank transfers, mobile applications or online transfers. Further, each mechanism has a plethora of competitors, for instance, in credit cards, there are MasterCard, Visa, American Express and Discovery and in the rapidly growing space of mobile payments, there is Square, Paypal, and Levelup. Each of these payment platforms must attract merchants and consumers to use their platform.

3.2 Model

There are two types of individuals, merchants $(A)$ and consumers $(B)$. There is a continuum of measure one of each. There is variation within these types - namely, there are sub-types of merchants and consumers who differ in their endowment levels. We index each agent by $(T, s)$ where $T$ is the type (merchant or consumer) and $s$ is the sub-type. There are $I$ subtypes of merchants, $A$ (and indexed by $i$) and $J$ subtypes of consumers, $B$ (and indexed by $j$). By introducing variation in an agent’s wealth, we can analyse how changes in the economic environment both affect the composition of a platform and an agent’s utility.

There is a fraction $a_{T,s}$ of each type $T$ and subtype $s$, and there is a measure of each one of each type $T$, $\sum_s a_{T,s} = 1 \forall T \in \{A, B\}$. Clearly the fraction of each subtype $a_{T,s}$ are arbitrary real numbers on the unit interval - not integers. Each agent has an endowment of capital, denoted by $K_{T,s} > 0$.

We model utility at a reduced form level, and assume agents procure utility from being matched with other agents. Although this is not realistic per se, we presume the process of being matched with other agents facilitates trade over the platform. We do not model that underlying environment - otherwise our setup is fairly general. For instance, agents may prefer larger platforms as it increases the number of potential trades. Further, we could generalize and introduce a term for any private benefit the platform provides over and above its matching service. In short, utilities are to be thought of as indirect.

We only allow non-negative integers of merchants and consumers to join a platform. The utility of a merchant (of any subtype $i$) matched with $N_A$ merchants and $N_B$ consumers is:

$$U_{A,i}(N_A, N_B) = U_A(N_A, N_B) = \begin{cases} 0 & \text{if } N_A \text{ or } N_B = 0 \\ \left[ \frac{N_B}{N_A} \right]^{r} + N_B^r & \text{else} \end{cases}$$

Note that the baseline utility of not being on any platform is zero. This is the ‘opt-out’ option and is

---

3The average US consumer uses a portfolio of payment mechanisms; Foster et al. [2013] found the average consumer used 5.2 different payment instruments (out of a possible 9).

4Hard numbers about the growth of mobile payments is difficult to procure, although there are a couple of statistics which hint at the possible growth of mobile payments; 25% of Starbucks transactions are paid for via a Starbucks’ prepaid account (Tavilla [2012]) and Nielsen [2012] report that 9% of survey responders paid for goods and services through their phone.

5An extension would include allowing for differences in subtype preferences.

6This is a natural assumption for the opt-out utility since the lower bound for $N_A$ is one (since as soon as a merchant joins a platform, there must be at least one merchant on the platform) and if $N_A$ is positive the limit of $U_A(N_A, N_B)$ as $N_B$ goes to zero is zero ($\lim_{N_B \to 0, N_B \to 0} \left( \frac{N_B}{N_A} \right)^{r} + N_B^r = 0$).
always available.

Symmetrically, for a consumer (of any subtype $j$) it is:

$$U_{B,j}(N_A, N_B) = U_B(N_A, N_B) = \begin{cases} 0 & \text{if } N_A \text{ or } N_B = 0 \\ \left[ \left( \frac{N_A}{N_B} \right)^{\gamma_B} + N_A^{\epsilon_B} \right] & \text{else} \end{cases}$$

Where $\gamma_A$, $\gamma_B$, $\epsilon_A$, $\epsilon_B \in (0, 1)$.

This utility function exhibits two important features which Ellison and Fudenberg [2003] highlight:

Individuals will compete between agents of their own type, whereas, they prefer more of the other type. For example, in the general merchant and consumer case, we are presuming that merchants dislike more merchants, since they will have to compete with more merchants which may reduce the good’s price due to greater competition. Therefore, this is a reduced form specification for competition between the same types on a platform which Armstrong [2006] highlights.\(^7\)

(1) **Market Impact Effects:** Each type prefers more of the other type and less of it’s own.

$$U_A(N_A + 1, N_B) - U_A(N_A, N_B) = \left[ \left( \frac{1}{N_A + 1} \right)^{\gamma_A} - \left( \frac{1}{N_A} \right)^{\gamma_A} \right] N_B^{\gamma_A} < 0$$

$$U_A(N_A, N_B + 1) - U_A(N_A, N_B) = \left[ \left( \frac{N_B + 1}{N_A} \right)^{\gamma_A} - (N_B + 1)^{\gamma_A} \right] + \left[ (N_B + 1)^{\gamma_A} - N_B^{\gamma_A} \right] > 0$$

Individuals prefer to be on larger platforms as it offers greater possibilities to trade for a given ratio of participants.

(2) **Scale effects:** An individual prefers larger platforms for a given ratio

- assume $\tau > 1$, therefore:

$$U_A(\tau N_A, \tau N_B) - U_A(N_A, N_B) = (\tau - 1)N_B^{\gamma_A} > 0$$

Symmetrically, both effects also apply for the type $B$ utility function. Therefore, there are scale effects within type and across types.

In the model, agents buy personal contracts which stipulate the number of merchants and consumers on the platform.

We denote the contract by $d_T(N_A, N_B)$ where $N_A$ and $N_B$ are the number of merchants and consumers respectively in the given platform and $T$ denotes the type of individual the contract is for, whether it is merchants ($A$) or consumers ($B$). Types are observed and Type $T$ cannot buy a contract indexed by $T'$.

Further one can think of an agent of a given type $T$ and subtype $s$ as allowed to join only one platform. Thus

---

\(^7\)It could be argued that the presence of more merchants could be beneficial for a merchant, since it could lead to ‘economies of agglomeration’, however, if the reasoning of the beneficial effects by agglomerating is that it attracts more consumers, then this is still achieved by the utility function posed. The utility function models that merchants prefer less merchants for a given number of consumers.
we can create a function \( x_{TS}[dT(NA, NB)] \geq 0 \) such that \( \sum_s x_{TS}[dT(NA, NB)] = 1 \) which is an indicator\(^8\) (or more generally a probability distribution) for the assignment of an agent \((T, s)\) to contract \(dT(NA, NB)\). The set of contracts for type \(A\) is denoted as \(DA\) and similarly the set of contracts for type \(B\) is denoted \(DB\).

The consumption set of type \(A, i\) agents can be written as:

\[
X_{A,i} = \{x_{A,i}[d_A(NA, NB)] \geq 0 \forall d_A \in DA, x_{A,i}[d_B(NA, NB)] = 0 \forall d_B \in DB\}
\]

The above condition states that type \(A, i\) agents can buy any non-negative amount of contract \(d_A \in DA\), but none of the type \(B\) contracts.

Symmetrically the consumption set of type \(B, j\) agents can be written as:

\[
X_{B,j} = \{x_{B,j}[d_B(NA, NB)] \geq 0 \forall d_B \in DB, x_{B,j}[d_A(NA, NB)] = 0 \forall d_A \in DA\}
\]

Since individuals may only join a single platform, this introduces an indivisibility into an agent’s consumption space. To overcome this problem we allow individuals to purchase mixtures, or probabilities of being assigned to a platform of a certain size including the opt-out option.\(^9\) For example, consider an agent who buys two different contracts: the first contract assigns the agent to a platform consisting of four merchants and three consumers with probability one-third, and the second contract assigns the agent to a platform consisting of three merchants and one consumer with probability two-thirds. The deterministic case, where an agent buys a contract which matches them with a platform of size \((NA, NB)\) with certainty, can be seen as a special case. We do not insist that there is mixing in a competitive equilibrium but it can happen as a special case. For instance, when agents are poor there can be mixing between a given platform and an opt-out contract.

As a technical assumption, we assume there is a maximal platform of size \((\bar{NA}, \bar{NB})\), and any platform up to this size can be created. By assuming there is a maximal platform size, it bounds the possible set of platforms and hence makes the commodity space finite. This is for simplicity, since we can choose \(\bar{NA}\) and \(\bar{NB}\) arbitrarily such that this condition does not bind.

The commodity space is thus:

\[
L = \mathbb{R}^{2(\bar{NA} \times \bar{NB}+1)+1}
\]

There are contracts for every possible platform size, in turn indexed by the two types and further there is capital. Thus as we define the maximal platform size to be \((\bar{NA}, \bar{NB})\) and there is always the opt-out contract, there are \(\bar{NA} \times \bar{NB} + 1\) contracts for each type. Since there are two types, we multiply this number by two for the number of contracts available. Finally there is a market for capital.

\(^8\)Our model can be extended to multi-homing (agents join multiple platforms) by omitting the requirement that an agent is matched to only one platform \(\sum_s x_{TS}[dT(NA, NB)] = 1\).

\(^9\)A similar modelling approach is used in Prescott and Townsend [1984], Prescott and Townsend [2005], Pawasutipaisit [2010].
All contracts $d_T(N_A, N_B)$ are priced in units of the capital good and the type $T$ price for contract $d_T(N_A, N_B)$ is denoted as $p[d_T(N_A, N_B)]$ for types $A$ and $B$ ($T \in \{A, B\}$).

### 3.2.1 Agent’s Problem

In summary, agent $T$, takes prices $p[d_T(N_A, N_B)] \forall d_T \in D_T$ as given and solves the maximisation problem:

$$\max_{x_{T,s} \in X_{T,s}} \sum_{N_A, N_B} x_{T,s}[d_T(N_A, N_B)] U_T[d_T(N_A, N_B)]$$

(s.t. $\sum_{N_A, N_B} x_{T,s}[d_T(N_A, N_B)] p[d_T(N_A, N_B)] \leq \kappa_{T,s}$)

$$\sum_{N_A, N_B} x_{T,s}[d_T(N_A, N_B)] = 1$$

Where each type of individual has an endowment of $\kappa_{T,s}$ of capital and the price of capital is normalized to one, i.e. capital is the numeraire.

Equation (3.1) is the agent’s expected utility from the assignment problem. Equation (3.2) is the agent’s budget constraint.

### 3.2.2 Platforms

We assume there are intermediaries or marketmakers who create platforms and sell contracts for each type to join the platform. As is evident there is constant returns to scale for the intermediaries so for simplicity we can envision just one is needed in equilibrium. We denote $y_A[d_A(N_A, N_B)]$, as the number of contracts produced for type $A$ of size $(N_A, N_B)$ and $y_B[d_B(N_A, N_B)]$ as the number of contracts produced for type $B$ of size $(N_A, N_B)$. These are counting measures and there is nothing random. Also, these numbers are on a continuum so do not have to be integer values. Further, we denote the number of platforms of size $(N_A, N_B)$ as $y(N_A, N_B)$. Thus $N_A \times y(N_A, N_B)$ is the number of type $A$’s in total on the type of platform $y(N_A, N_B)$. Similarly, $N_B \times y(N_A, N_B)$ for type $B$.

The total number of agents of each type on a platform must be consistent with the total number of contracts offered for that type. Thus, as indicated above the intermediary must satisfy the following matching constraint:

$$\frac{y_A[d_A(N_A, N_B)]}{N_A} = \frac{y_B[d_B(N_A, N_B)]}{N_B} = y(N_A, N_B) \forall d_A \in D_A, \forall d_B \in D_B$$

This constraint states that the measure of contracts created for type $A$ of size $(N_A, N_B)$ must have the equivalent number of contracts created for type $B$ normalised by the number of agents in each platform. For example, consider 0.1 platforms are created which match three merchants and two consumers. This would
require $0.1 \times 3 = 0.3$ merchant contracts and $0.1 \times 2 = 0.2$ consumer contracts being created for a platform of size $(3, 2)$.

A platform of size $(N_A, N_B)$ requires the following amount of capital:

$$C(N_A, N_B) = \begin{cases} 0 & \text{if } N_A = 0 \text{ or } N_B = 0 \\ c_A N_A + c_B N_B + c N_A N_B + K & \text{else} \end{cases}$$

The capital requirement of a singleton platform or autarky is normalised to zero as it costs nothing to produce it is always available. The amount of capital required for a platform has a positive marginal cost for an extra agent on each side of the platform (captured by $c_A$ and $c_B$) and for the multiple of agents on both sides (captured by the interaction term $c$). Additionally, we model there can be some fixed cost, $K$, in creating a platform. For a more flexible specification we allow $c_A$ and $c_B$ to be different. We assume\(^\text{10}\) that $c_A, c_B, c \in (0, \infty)$ and $K \in [0, \infty)$.

We denote the amount of capital input purchased by the intermediary as $y_k$—this has to be sufficient to build the proposed platforms. So we can write the intermediary’s capital constraint as:

$$\sum_{N_A, N_B} y(N_A, N_B) [C(N_A, N_B)] \leq y_k \quad (3.5)$$

Hence, the intermediary’s production set is:

$$Y = \{(y, y_k) \in \mathbb{R}^{2(N_A \times N_B + 1)} | (3.4) \text{ and } (3.5) \text{ are satisfied}\}$$

It is a convex cone as in McKenzie [1959]. The intermediary takes the same Walrasian prices $p[d_T(N_A, N_B)] \forall d_T \in D_T, T \in \{A, B\}$ and maximizes profits constructing platforms and selling type specific matchings (again we normalize the price of capital to be one):

$$\pi = \max_{(y, y_k) \in Y} \sum_{N_A, N_B} \{p[d_A(N_A, N_B)] N_A + p[d_B(N_A, N_B)] N_B\} \times y(N_A, N_B) - y_k \quad (3.6)$$

Equation (3.6) states the intermediary maximizes how many platforms of a given size $(N_A, N_B)$ to produce given the prices for each position in the platform. The intermediary’s profits are equal to the number of platforms it constructs multiplied by the sum of prices of the contracts minus the cost of the capital input.

The intermediary’s first order condition for creating a platform $y(N_A, N_B)$ is:

$$C(N_A, N_B) \geq p[d_A(N_A, N_B)] N_A + p[d_B(N_A, N_B)] N_B \quad (3.7)$$

Where equation (3.7) holds with equality if there are positive number of platforms of that size $(N_A, N_B)$. If equation (3.7) is a strict inequality then no such platform exists in equilibrium. Notice this natural condition requires the payments/memberships the platform receives must cover all of the platform’s costs.

\(^{10}\)We require $c > 0$, this assumption ensures we can bound the size of the equilibrium platforms.
3.2.3 Market Clearing

For market clearing we require the following conditions to hold

$$\sum_s \alpha_{T,s} x_{T,s}[d_T(N_A, N_B)] = y_T(d_T(N_A, N_B) \equiv N_A \times y(N_A, N_B) \forall N_A, N_B, T \in \{A, B\}$$ (3.8)

$$\sum_{T,s} y_{T,s} = y_\kappa$$ (3.9)

Equation (3.8) ensures that the (decentralized) amount of demand for each contract for each type equals the (decentralized) supply of that contract. Equation (3.9) states the total endowment of capital must equal the amount of capital used by the intermediary.

3.2.4 Competitive Equilibrium

Let us define $x$ as the vector of contracts bought $x_{T,s}[d_T(N_A, N_B)]$ for all subtypes $(T, s)$, then a competitive equilibrium in this economy is $(p, x, \{y, y_\kappa\}) \in L \times X \times Y$ such that for given prices $p[d_T(N_A, N_B)]$:

1. The allocation $\{x_{T,s}[d_T(N_A, N_B)]\}$ solves the agent’s maximisation problem [i.e. $x_{T,s}[d_T(N_A, N_B)]$ solves equation (3.1) subject to equations (3.2 and 3.3)].

2. The allocation $\{y(N_A, N_B), y_\kappa\}$ solves the platform’s maximisation problem [i.e. $\{y(N_A, N_B), y_\kappa\}$ solve equation (3.6)].

3. The market clearing conditions hold [equations (3.8) and (3.9) hold].

In equilibrium, the pricing mechanism will determine the size and number of each platform and subsequently the relative proportions of merchants and consumers on each platform.

3.3 Social Planner’s Problem

First, we set up the social planner’s problem and determine the set of all Pareto optimal contracts. We show (i) a competitive equilibrium is Pareto optimal, (ii) any Pareto optimal allocation can be achieved with lump-sum transfers and taxes among agents and (iii) there exists a competitive equilibrium. So these results have two important implications (i) the decentralised problem is Pareto optimal, and hence efficient and (ii) when solving for the competitive equilibrium, we can use the simpler social planner’s problem to compute the allocation. Subsequently, we can use the Lagrangian multipliers to impute the competitive equilibrium prices and wealth associated with that allocation.

The social planner’s welfare maximising problem with Pareto weights $\lambda_{A,i}$ and $\lambda_{B,j}$ for types $(A, i)$ and $(B, j)$ respectively is:

\[\text{minimize } \sum_{i,j} \lambda_{A,i} (\phi_A^i - \xi_i^A) + \lambda_{B,j} (\phi_B^j - \xi_j^B)\]

\[\text{subject to } \sum_{i,j} \lambda_{A,i} (\phi_A^i - \xi_i^A) + \lambda_{B,j} (\phi_B^j - \xi_j^B) \leq \text{budget} \]

\[\text{and } \sum_{i,j} \lambda_{A,i} (\phi_A^i - \xi_i^A) + \lambda_{B,j} (\phi_B^j - \xi_j^B) \geq \text{utility}\]
$$\max_{x \geq 0, y \geq 0} \sum_{i} \lambda_{A,i} \left\{ \sum_{d_A(N_A, N_B)} \alpha_{A,i} x_{A,i} [d_A(N_A, N_B)] U_A(N_A, N_B) \right\}$$

$$+ \sum_{j} \lambda_{B,j} \left\{ \sum_{d_B(N_A, N_B)} \alpha_{B,j} x_{B,j} [d_B(N_A, N_B)] U_B(N_A, N_B) \right\}$$

s.t. \( \sum_{b(N_A, N_B)} x_{T,s} [d_T(N_A, N_B)] = 1 \forall T, s \) \hspace{1cm} (3.10)

\[ \sum_{s} \alpha_{T,s} x_{T,s} [d_T(N_A, N_B)] = y(N_A, N_B) \times N_T \forall d_T, \forall T \in \{A, B\} \] \hspace{1cm} (3.11)

\[ \sum_{N_A, N_B} y(N_A, N_B) [C(N_A, N_B)] \leq \sum_{T,s} \alpha_{T,s} \kappa_{T,s} \] \hspace{1cm} (3.12)

Equation (3.10) ensures that each individual is assigned to a platform, equation (3.11) ensures that the total purchase of contracts equals the number of contracts produced, and equation (3.12) ensures the total number of contracts produced is resource feasible.

3.3.1 Dual

The Pareto Problem can also be written in terms of its dual equivalent:

$$\min_{p} \sum_{T,s} \alpha_{T,s} (p_{T,s} + p_{\kappa T,s})$$

s.t. \( p_{T,s} + p_T [d_T(N_A, N_B)] \geq \lambda_{T,s} U_T(N_A, N_B) \forall i, \forall T, \forall (N_A, N_B) \) \hspace{1cm} (3.13)

\[ C(N_A, N_B) - \{ p_A [d_A(N_A, N_B)] \times N_A + p_B [d_B(N_A, N_B)] \times N_B \} \geq 0 \forall (N_A, N_B) \]

In this formulation \( p_{T,s} \) is the imputed valuation for type \( T, s \). So \( p_{T,s} \) will be higher, the scarcer the type. And \( p_{\kappa} \) is the imputed valuation for capital. The dual problem minimizes the aggregate cost of the economy (in terms of prices of each type and total capital) such that each type of agent receives a given level of Pareto weighted utility. The primal problem maximizes the Pareto weighted expected utility of each type subject to the matching and resource constraints.

The Pareto problem is well defined in both the primal and dual form therefore, by the 'strong duality property'\(^{11}\) there must exist an optimal solution \((p^{*}, \alpha^{*}, y^{*})\) such that:

\(^{11}\)See Bradley et. al. [1977] pages 142-143 for more details.
Theorem 1. If all agents are non satiated, a competitive equilibrium \((p^*, x^*, y^*)\) is a Pareto optimal allocation \((x^*, y^*)\). [First Welfare Theorem]

Proof follows from Prescott and Townsend [2005].

Theorem 2. Any Pareto optimal allocation \((x^*, y^*)\) can be achieved through a competitive equilibrium with transfers between agents subject to there being a cheaper point for all agents and agents are non-satiated.

The proof relies on using the Pareto weights \(\lambda_{T,s}\) and the dual variables from the planner problem, to claim the Pareto optimal allocation \((x^*, y^*)\) can be supported as a competitive equilibrium with transfers between agents. The proof follows from Prescott and Townsend [2005].

Theorem 3. There exists a competitive equilibrium.

To provide a more general proof of existence of a competitive equilibrium, where the distribution of wealth across individuals is taken as given, but there is no restriction on the mass of agents requires the use of a fixed point theorem. Negishi [1960] alters the Pareto weights in the economy such that the budget constraints binds for all agents at the fixed point. The proof follows from Prescott and Townsend [2005].

3.4 Results

3.4.1 Prices for joining a platform

The consumer’s maximisation problem allows us to see how an agent decides which contracts to purchase. Her first order conditions for contract \(x_{T,s}[d_T(N_A, N_B)]\) is:

\[
U_T(N_A, N_B) - \mu^P_{T,s} - \mu^B_{T,s} * p d_T(N_A, N_B) \leq 0
\]  \hspace{1cm} (3.14)

Where \(\mu^P_{T,s}\) is the Lagrangian multiplier associated with the individual being assigned to some platform and \(\mu^B_{T,s}\) is the Lagrangian multiplier associated with the agent’s budget constraint. Furthermore, for any platform the agent buys with positive probability \((x_{T,s}[d_T(N_A, N_B)] > 0)\), then the equation will hold with
equality. If the left hand side of equation (3.14) is strictly less than zero, this implies that agent will not purchase that contract.

Let us consider what equation (3.14) implies. Consider an agent who purchases positive probabilities of two different contracts $dT(N_A, N_B)$ and $dT(N_A', N_B')$. Let us define the variable $\Delta U \equiv U_T(N_A, N_B) - U_T(N_A', N_B')$ and $\Delta p \equiv p[d_T(N_A, N_B)] - p[d_T(N_A', N_B')]$, then we can state:

$$\Delta U = \mu_{T,s}^B \Delta p$$

Therefore, if an individual buys two contracts with positive probability, the difference in utility she will derive will be a constant multiplied by the change in price. Intuitively if an individual is indifferent between two contracts the change in utility must be equal to the change in price.

In general, an agent is unwilling to pay proportionally more for a contract which confers proportionally more utility ($\frac{\Delta U}{\Delta p} \neq \mu_{T,s}^B$) - this is only true when the individual’s matching constraint is not binding ($\mu_{T,s}^B = 0$). Intuitively, when an individual’s matching constraint binds, this individual would prefer to join more platforms but is constrained by the ability to only join one platform. In turn, this will ensure the percentage increase in the individual’s willingness to pay to join the platform which confers the greater utility will be more than the percentage change in utility. Since both platforms require the same assignment of type component, but one platform confers greater utility.

### 3.5 Examples

In a general equilibrium framework we can analyze both how the composition of platforms and how the resulting utilities change as we alter some of the parameters. First as a useful benchmark we examine an equilibrium where we have symmetric parameters for both sides of the market, that is the same costs, preferences and Pareto weights. We then provide a second example which varies the wealth across the agents and examine which platforms are created, what the respective utilities are and what the equilibrium prices. Third, we are interested how the equilibrium and subsequently agents’ utilities chang as we redistribute wealth within our economy. Fourth, we examine how the equilibrium utilities change as we alter the Pareto weights. We show that even if an agent’s relative Pareto weight falls, their equilibrium utility can actually rise – it depends on the General Equilibrium effects. Fifth, given that the cost of producing platforms changes over time, for instance due to technological improvement, we demonstrate how the equilibrium utilities change as we alter the fixed cost of producing platforms. We show that increasing fixed costs leads to heterogeneous effects and potentially increases inequality.
Table 3.1: Equilibrium platforms and user utility for Example 1.

<table>
<thead>
<tr>
<th>Equilibrium Platforms</th>
<th>Platforms created production $y(N_A,N_B)$</th>
<th>Cost of production $C(N_A,N_B)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>($N_A, N_B$)</td>
<td>0.5</td>
<td>8</td>
</tr>
</tbody>
</table>

Equilibrium user utility and platform choice.

| Type $T_i, s$ | Wealth $s_{T_i, s}$ | Platform joined $(N_A, N_B)$ | Price of joining $p(\lambda T_i|[N_A,N_B])$ | Pr(joining) $x_{T_i, s}(\lambda T_i|[N_A,N_B])$ | Utility on Platform $U_T(N_A, N_B)$ |
|---------------|----------------------|-----------------------------|---------------------------------------------|----------------------------------------------|-----------------------------------|
| $A, 1$        | 2                    | (2, 2)                      | 2                                           | 1                                            | 2.41                              |
| $A, 2$        | 2                    | (2, 2)                      | 2                                           | 1                                            | 2.41                              |
| $B, 1$        | 2                    | (2, 2)                      | 2                                           | 1                                            | 2.41                              |
| $B, 2$        | 2                    | (2, 2)                      | 2                                           | 1                                            | 2.41                              |

3.5.1 Example 1

Our initial example has two subtypes for each type and is symmetric – there is equal fractions of each type ($\alpha_{A1} = \alpha_{A2} = \alpha_{B1} = \alpha_{B2}$), each subtype has the same Pareto weight ($\lambda_{A1} = \lambda_{A2} = \lambda_{B1} = \lambda_{B2}$), the cost function is the same for both types ($c_A = c_B$) and the utility functions' parameters are the same ($\gamma_A = \gamma_B$ and $\epsilon_A = \epsilon_B$).\(^{12}\) In this initial example, although there are two subtypes, but in fact they are identical therefore there is no variation by subtype.

In this equilibrium only one type of platform is created. All users pay a price of two units of capital to join a platform which matches them with two users of the other type, and one more user of their own type, so the total for each type is two.

3.5.2 Example 2

Our second example varies wealth both within- and across- type. To improve intuition, let us consider a payment platform which connects merchants to consumers. There are two subtypes of merchants – Small ($A, 1$) and Big ($A, 2$); and two subtypes of consumers – Rural ($B, 1$) and Urban ($B, 2$). Each consumer would prefer to be on platform with more merchants (more places to pay) and less consumers (less congestion). Similarly, merchants want as many consumers to use the same platform but would like less rival merchants.

There is equal fractions of each type ($\alpha_{A1} = \alpha_{A2} = \alpha_{B1} = \alpha_{B2}$), the cost function is the same for both types ($c_A = c_B$) and the utility functions' parameters are the same ($\gamma_A = \gamma_B$ and $\epsilon_A = \epsilon_B$) however they vary in wealth.\(^{13}\)

\(^{12}\)The parameter values are:

$\alpha_{A1} = \alpha_{A2} = \alpha_{B1} = \alpha_{B2} = \frac{1}{2}; c_A = c_B = c = 1; K = 0; \gamma_A = \gamma_B = \epsilon_A = \epsilon_B = \frac{1}{2}$

$\lambda_{A1} = \lambda_{A2} = \lambda_{B1} = \lambda_{B2} = \frac{1}{2}$

\(^{13}\)The parameter values are:

$\alpha_{A1} = \alpha_{A2} = \alpha_{B1} = \alpha_{B2} = \frac{1}{2}; c_A = c_B = c = 1; K = 0; \gamma_A = \gamma_B = \epsilon_A = \epsilon_B = \frac{1}{2}$
Table 3.2: Equilibrium platforms and user utility for Example 2.

<table>
<thead>
<tr>
<th>Platform Size</th>
<th>Number of Platforms created</th>
<th>Cost of Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>((N_A, N_B))</td>
<td>(y(N_A, N_B))</td>
<td>(C(N_A, N_B))</td>
</tr>
<tr>
<td>((3, 2))</td>
<td>0.25</td>
<td>11</td>
</tr>
<tr>
<td>((1, 2))</td>
<td>0.25</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>Wealth</th>
<th>Platform joined ((N_A, N_B))</th>
<th>Price of joining (p(dT[N_A, N_B]))</th>
<th>(\text{Pr}(\text{joining})) (x_{T,s}(dT[N_A, N_B]))</th>
<th>Utility on Platform (U_T(N_A, N_B))</th>
<th>Expected Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Merchant, Small</td>
<td>1.37</td>
<td>((3, 2))</td>
<td>1.37</td>
<td>1</td>
<td>2.23</td>
<td>2.23</td>
</tr>
<tr>
<td>Merchant, Big</td>
<td>1.64</td>
<td>((3, 2))</td>
<td>1.37</td>
<td>0.5</td>
<td>2.23</td>
<td>2.53</td>
</tr>
<tr>
<td>((1, 2))</td>
<td>1.91</td>
<td></td>
<td></td>
<td>0.5</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>(B) Consumer, Rural</td>
<td>1.54</td>
<td>((1, 2))</td>
<td>1.54</td>
<td>1</td>
<td>1.7</td>
<td>1.7</td>
</tr>
<tr>
<td>Consumer, Urban</td>
<td>3.45</td>
<td>((3, 2))</td>
<td>3.45</td>
<td>1</td>
<td>2.96</td>
<td>2.96</td>
</tr>
</tbody>
</table>

In this equilibrium, there are two different types of platforms created. One platform is larger than the other, and is populated with more merchants; the richer urban consumers join this platform. Urban consumers have higher resultant utility since this platform is both bigger and has a more favorable ratio of merchants to consumers. Whereas, the poorer rural consumers join the smaller platform, this platform is both smaller and is populated with a less favorable ratio of merchants to consumers causing lower utility for rural consumers (and lower prices for consumers to join that platform).

The urban and rural consumers, and the small merchants all buy degenerate lotteries of contracts, where they are assigned to a particular platform with probability one. However, the big merchants buy probabilities in two different platforms, therefore 50% of them are allocated to platforms of size \((3, 2)\) and 50% are allocated to platforms of size \((1, 2)\). The respective prices for these two different contracts are 1.37 and 1.91.

The cost of the platform is borne primarily by the type which values the platform the most (consumers); therefore, in the platforms of size \((3, 2)\) consumers contribute 63% of the platform’s cost, even though, they are only 40% of the platform’s population. Further, rural consumers are wealthier than small merchants, yet they are worse off. Urban consumers are much richer than the other participants hence they are willing to sponsor larger platforms. This allows merchants to contribute less towards joining a platform. Another way to look at this is that the merchants are in scarcer supply (since consumers are so much wealthier – average consumer wealth is 2.5 and average merchant wealth is only 1.5), therefore the price schedule they face is lower than the consumers’ schedule.\(^\text{14}\)

3.5.3 Changing endowments

If we redistribute wealth in our economy, this will change the relative demand for merchants and consumers and subsequently change the relative prices to join a given platform. To examine the general equilibrium

\(^{14}\)Recall that the cost parameters \(c_A = c_B\) therefore, the asymmetric prices and allocations are solely driven by agents’ different capital endowments.
effects of redistributing wealth, we construct two placebo interventions which reallocate wealth within our economy whilst holding the total resources constant.

Figure (3-1) shows the effects of redistributing wealth within our economy while holding total resources constant (with the same cost and preferences as in the previous examples). The left panel shows the effects on utility from redistributing wealth across types\(^{15}\), in particular between \((B, 1)\) and \((A, 2)\) – i.e. \(\kappa_{B,1} + \kappa_{A,2} \approx 2.4\). The right panel shows the effects on utility from redistributing wealth within type\(^{16}\), in particular between \((B, 1)\) and \((B, 2)\).

Recall our payment platform example which connects merchants to consumers. There are two subtypes of merchants – Small \((A, 1)\) and Big \((A, 2)\); and two subtypes of consumers – Rural \((B, 1)\) and Urban \((B, 2)\). As we increase the wealth of rural consumers \((\kappa_{B,1})\), the utility of urban consumers \((B, 2)\) falls (holding both urban consumer’s wealth \((\kappa_{B,2})\) and total wealth \((\sum_{T,s} \kappa_{T,s})\) constant).

As we increase rural consumer’s wealth \((\kappa_{B,1})\), the demand to join platforms with merchants rises. Subsequently the price for consumers to join platforms for a given number of merchants will also rise. Therefore, since urban consumer’s wealth \((\kappa_{B,2})\) is a constant and they now face higher prices, their utility must fall. Symmetrically, a similar result holds when we increase the wealth of small merchants on big merchants.

As you would expect the utility of merchants and consumers are increasing in their respective wealth.\(^{17}\)

Further, we can consider how the equilibrium changes as we adjust the endowments within a type (consumers), and hold the endowments of the other types (merchants) fixed. There is no effect on merchants’ utilities since any reduction in purchasing power by one of the consumer subtypes is compensated by an equal change in the other slightly richer consumer subtype. This example is shown in the right panel of figure (3-1).\(^{18}\)

### 3.5.4 Altering Pareto weights

We can also consider how the equilibrium changes as we adjust the Pareto weights\(^{19}\) on only one subtype \((B, 2)\).

Figure (3-2) demonstrates (for given parameters) how the resulting utilities change as the Pareto weight for type \((B, 2)\) increases. First it is clear and intuitive the utility of \((B, 2)\) monotonically weakly increases

---

\(^{15}\)We solve the model using the Pareto problem and then impute the wealth and prices which replicate the same allocation. We simulate 2880 equilibria for different Pareto weights, and then collected only the equilibria in which \(0.51 < \kappa_{A,1} < 0.59\) and \(1.01 < \kappa_{B,2} < 1.19\). We then ’join up’ all the points to plot a smooth curve.

\(^{16}\)We solve the model using the Pareto problem and then impute the wealth and prices which replicate the same allocation. We simulate 2880 equilibria for different Pareto weights, and then collected only the equilibria in which \(0.51 < \kappa_{A,1} < 0.59\) and \(1.01 < \kappa_{A,2} < 1.1\). We are approximately holding the endowment of \((A, 1)\) and \((A, 2)\) constant.

\(^{17}\)Figure (3-5) in the appendix shows the effect of the wealth changes on all agents’ utilities.

\(^{18}\)There is a tiny change in the utility of \((A, 2)\) this is due to the discrete nature of the possible platform combinations and the changes in the platforms type \((B)\) can purchase.

\(^{19}\)The parameter values are:

\[
\begin{align*}
\alpha_{A1} &= \alpha_{A2} = \alpha_{B1} = \alpha_{B2} = \frac{1}{2}; c_A = c_B = c = 1; K = 0; \gamma_A = \gamma_B = \epsilon_A = \epsilon_B = \frac{1}{2} \\
\lambda_{A1} &= \frac{1.01 - x}{3}, \lambda_{A2} = \frac{0.99 - x}{3}, \lambda_{B1} = \frac{1 - x}{3}, \lambda_{B2} = x; \text{ We introduce a tiny wedge between } (A, 1) \text{ and } (A, 2) \text{ to highlight the effects on a favored subtype.}
\end{align*}
\]
Figure 3-1: Redistributing wealth within and across type

Redistributing wealth across type:

The effect on utility from transferring wealth from \((A, 2)\) to \((B, 1)\)

How does utility change as we adjust wealth for B1

Note we do not change \((A, 1)\) or \((B, 2)\)'s wealth in these examples. The changes in \((A, 1)\) and \((B, 2)\)'s utility is due to the general equilibrium effects of redistributing wealth from \((A, 2)\) to \((B, 1)\).

Redistributing wealth within type:

The effect on utility from transferring wealth from \((B, 2)\) to \((B, 1)\)

How does utility change as we adjust wealth for B1

Note we do not change \((A, 1)\) or \((A, 2)\)'s wealth in these examples. The changes in \((A, 1)\) and \((A, 2)\)'s utility is due to the general equilibrium effects of redistributing wealth from \((B, 2)\) to \((B, 1)\).

Figure 3-2: How does the utility for each subtype change as we alter the Pareto weight for subtype B,2
with their respective Pareto weight. This is a general result, as the Pareto weight on \((B, 2)\) rises, it must be true that the utility for \((B, 2)\) weakly increases.

As figure (3-2) shows, type \((B, 1)\) is clearly disadvantaged. This is a general result and follows from the utility of subtype \((B, 2)\) rising.

**Theorem 4.** If we increase the Pareto weight on subtype \((B, 2)\) by \(\Delta\) and reduce the Pareto weight on subtype \((B, 1)\) by \(\Delta\), the utility of type \((B, 1)\) monotonically increases in \(\Delta\) and the utility of subtype \((B, 1)\) monotonically falls in \(\Delta\), where \(0 < \Delta < \lambda_{B,1}\).

This follows from the result that a subtype’s utility is monotonically increasing in their own Pareto weight and that the different subtypes compete amongst themselves to be matched with other type since subtype \((B, 1)\)'s utility never increases by being matched with type \((B, 2)\).

Recall our merchant and consumer example from before. If we increase the Pareto weight on urban consumers \((\lambda_{B,2})\), the allocation will match them in both larger platforms and with more merchants. This has two effects: first, there is less resources left for the rural consumers, and second, there are less merchants left unmatched.

The story is more complicated for the merchants. An increase in the rural consumers Pareto weight can lead to lower or higher utility for merchants. One of the merchants subtypes will always be made worse off \((A, 1)\) by the rise in \(\lambda_{B,2}\) since some platforms are comprised of relatively more merchants – favoring consumers on those platforms.

As seen in figure (3-2) it is possible for one of the merchant subtypes \((A, 2)\) to be made better off – even as their relative Pareto weight falls. This follows from the most favored consumer subtype (urban) are matched to proportionally more merchants as \(\lambda_{B,2}\) increases, this means the proportion of merchants to consumers remaining declines. Hence, those merchants who are not matched with urban consumers may be matched at favorable ratios of consumers to merchants – increasing their utility.

### 3.5.5 Changing Costs

A further important consideration is how the equilibrium changes as we adjust costs, for instance, technological innovations may decrease the costs of creating a platform. Figure (3-3) shows how the equilibrium changes for two different fixed costs for building a platform; the left panel shows the equilibrium utilities if the fixed cost is zero and the right panel shows the equilibrium utilities if the fixed cost is one unit of capital. As you would expect for a given distribution of wealth, utility is lower. However, the distribution itself changes. At the larger fixed costs of producing a platform, the distribution becomes more dispersed, and inequality between different subtypes becomes more pronounced.

The higher costs of creating platforms will cause less platforms to be created in equilibrium. To gain intuition for our result, recall our interpretation that agents are endowed with two assets: labor and capital. As we increase the costs of producing platforms of a given size, this will lead to the relative value of capital
Figure 3-3: How does utility change as we alter the platform’s cost function

Utility for each subtype as the wealth for (B,1) varies and $K = 0$

How does utility change as we adjust wealth for $B_1$

Utility for each subtype as the wealth for (B,1) varies and $K = 1$

How does utility change as we adjust wealth for $B_1$

to labor becoming larger. Therefore, those agents who are endowed with more capital are less hurt by the rise in costs, leading to greater inequality.

For example, consider the equilibrium with $\kappa_{B,1} = 1.09$, with the introduction of the fixed cost, the utility of subtype $(A, 1)$ falls by over 25% whereas for the richest subtype $(A, 2)$ the fall in utility is about 3%. A similar qualitative result holds for subtypes $(B, 1)$ and $(B, 2)$.

A further way to demonstrate how changes in a fixed cost have distributional impacts is to vary the fixed cost of building a platform whilst holding each agent’s wealth constant. Figure (3-4) shows the equilibrium utility for each subtype when we vary the fixed cost of building a platform for given parameters.\(^{20}\) As you can see the richest subtype $(A, 2)$ is barely affected by the rise in platform costs. Yet, the poorest subtype $(A, 1)$’s utility, falls by about 50% as we increase the fixed cost of building a platform from 0.2 units of capital to 2 units of capital. Further supporting the evidence that changes in fixed costs of building platforms have heterogenous effects and in particular the most adverse effects occur on the poorest agents.

Do these distributional impacts suggest a rationale to regulate prices? No – our environment is Pareto optimal, so the optimal government intervention would be to introduce lump-sum taxation on the rich and transfers to the poor. This would increase the poorest’s utility, achieving a more equitable division of utility while maintaining a Pareto optimal allocation.

\(^{20}\)The economy’s parameters are $\alpha_{A1} = \alpha_{A2} = \alpha_{B1} = \alpha_{B2} = \frac{1}{2}$; $c_A = c_B = c = 1$; $\gamma_A = \gamma_B = \epsilon_A = \epsilon_B = \frac{1}{2}$; $\kappa_{A1} = 0.5, \kappa_{A2} = 1.5, \kappa_{B1} = 0.8, \kappa_{B2} = 1.1$. For computational simplicity, we allow the equilibrium wealth levels to be close to the desired wealth levels ($\kappa_{A1} = 0.5, \kappa_{A2} = 1.5, \kappa_{B1} = 0.8, \kappa_{B2} = 1.1$). We only plot the equilibrium utilities for those equilibriums such that the maximum difference between the desired wealth endowment and the plotted capital endowment is less than 0.1 units of capital for each subtype.
3.6 Conclusion

There are many economic platforms which must cater to multiple, differentiated users, who in turn, care about who else the platform serves. There are many examples: credit cards, clearinghouses, dark pools to name but a few. Attempting to model these interdependent preferences—especially in a general equilibrium framework—is difficult. However, these preferences lead to interesting dynamics, which can only be characterized within a general equilibrium framework.

Our paper has three main contributions.

Our first contribution is methodological. Contrary to the literature on two-sided markets, we model an economy with competing platforms in a general equilibrium framework. By building on the work of Prescott and Townsend [2006], we show that a competitive equilibrium has different implications than previous papers in two-sided markets. Further, our framework is relatively general; we can analyze an economy with many (i.e. more than two) types of users, who may have heterogeneous preferences; an economy with heterogeneous costs for servicing different users or an economy with inherent differences within a type’s wealth.

Second, our economy incorporates that an individual’s utility may be contingent on the actions of others. In this modified economy we show that these interdependencies do not lead to a network externality (contrary to Rochet and Tirole [2003, 2006])—in particular, the potential externality is ‘priced in’—in a manner suggested by Meade [1952] and Arrow [1969]. Thereby, our environment does not suggest a rationale to regulate prices on a platform—if a social planner wishes to implement a more equitable allocation, a social
planner should redistribute wealth and not regulate prices. Further, in contrast to the prevailing work in two-sided markets we demonstrate that it is not the sole presence of interdependent utilities that leads to inefficiencies. In our environment a further market imperfection would be required, such as market power or regulated prices.

Third, we demonstrate how changes in one agent’s wealth (or Pareto weight) has interesting general equilibrium effects both within- and across-type. For instance, consider a payment platform for consumers and merchants, where there are two subtypes of consumers – rural and urban consumers. An increase in rural consumer’s wealth will lead to decreases in urban consumer’s welfare and ambiguous effects on the merchants’ welfare. This follows from our assumption that agents do not like to be on a platform with more of their own type, therefore as we increase the rural consumer’s wealth, they will sponsor platforms with more merchants (and less consumers). Further, the rise in rural consumer’s wealth will lead them to pay a greater fraction of the costs of being on a platform – thereby potentially improving some merchants’ welfare.

We should make clear the limitations of our framework. First, our model is purely static, therefore we exclude any coordination failures (Caillaud and Jullien [2003]’s “Chicken and the Egg problem”) and any possibility of innovation in platform design. Second, no platforms or agents have any pricing power in our model, which as Weyl [2010] shows may interact with the agent’s preferences over other agent’s actions to exacerbate or minimize market failures. Third, our sole source of platform differentiation arises from the size and composition of a platform’s users. Fourth, we do not allow ever increasing economies of scale in platform size.

Ultimately, modelling and understanding platform economies is difficult. However, using techniques from the ‘Firms as Clubs’ and General Equilibrium literature we show even with interdependent utilities – our economy’s competitive equilibrium is Pareto optimal. We hope this paper ignites a discussion on how to model and analyze multiple, competing platforms.
3.7 Appendix

3.7.1 Computation

Attempting to compute the Pareto problem can be difficult due to the large commodity space and the number of constraints, therefore, we transform the above Pareto problem by removing the club constraints therefore, allowing us to use simplex algorithms which are quicker and easier to handle the large commodity and constraint space.

For ease of explanation let us assume there is only two subtypes of merchants and consumers, i.e. $i \in \{1, 2\}$ and $j \in \{1, 2\}$.

First we eliminate the club constraints recall equation (3.11), this constraint can be rewritten in matrices for each contract $d_T(N_A, N_B)$ as

\[
\begin{bmatrix}
\alpha_{A,1} & \alpha_{A,2} & 0 & 0 & -N_A \\
0 & 0 & \alpha_{B,1} & \alpha_{B,2} & -N_B
\end{bmatrix}
\begin{bmatrix}
x_{A,1}[d_A(N_A, N_B)] \\
x_{A,2}[d_A(N_A, N_B)] \\
x_{B,1}[d_B(N_A, N_B)] \\
x_{B,2}[d_B(N_A, N_B)] \\
y(N_A, N_B)
\end{bmatrix}
= \begin{bmatrix} 0 \\ 0 \end{bmatrix}
\] (3.15)

Since, $x_{T,s}[d_T(N_A, N_B)]$ and $y(N_A, N_B)$ must be non-negative, with equation (3.15) let us define a polyhedral cone, with a single extreme point at the origin. Therefore, using the Resolution Theorem of Polyhedrons, the systems of equations can be represented as the set of all non-negative linear combinations of its extreme rays. Scaling each $y(N_A, N_B) = 1$, the extreme rays of this cone are:

\[
\begin{align*}
&\left( \frac{N_A}{\alpha_{A,1}}, 0, \frac{N_B}{\alpha_{B,1}}, 0, 1 \right) \\
&\left( \frac{N_A}{\alpha_{A,1}}, 0, 0, \frac{N_B}{\alpha_{B,2}}, 1 \right) \\
&\left( 0, \frac{N_A}{\alpha_{A,2}}, N_B, 0, 1 \right) \\
&\left( 0, \frac{N_A}{\alpha_{A,2}}, 0, \frac{N_B}{\alpha_{B,2}}, 1 \right)
\end{align*}
\]

Let $y^{(i,j)}(N_A, N_B)$, the quantity of each ray, where $i$ is the subtype A agent, $j$ is the subtype B agent.

Therefore, we can define the set of \( \{x_{T,s}[d_T(N_A, N_B)], y(N_A, N_B)\} \) that satisfies (3.15) as:

\[
\{x_{T,s}[d_T(N_A, N_B)], y(N_A, N_B)\} = [y^{(1,1)}(N_A, N_B)] \left( \frac{N_A}{\alpha_{A,1}}, 0, \frac{N_B}{\alpha_{B,1}}, 0, 1 \right) + \\
\ldots + [y^{(2,2)}(N_A, N_B)] \left( 0, \frac{N_A}{\alpha_{A,2}}, 0, \frac{N_B}{\alpha_{B,2}}, 1 \right)
\]

132
Where $y^{(i,j)}(NA, NB) \geq 0$, $i = 1, 2$ and $j = 1, 2$. Intuitively, each ray is a different composition of types of agents to fulfill the contract, for example $y^{(1,1)}(NA, NB)$ corresponds to the measure of platforms which are fulfilled by agents $(A, 1)$ and $(B, 1)$. There are four extreme rays hence a linear combination of these four rays is able to replicate any combination of types of agents. In general, if there are $I$ types of $A$ and $J$ types of $B$ then there will be $I \times J$ extreme rays for each contract.

Furthermore, we have the following relations:

$$x_{A,i}(NA, NB) = \sum_{i'} \frac{y^{(i,j)}}{\alpha_{A,i}} NA$$

$$x_{B,j}(NA, NB) = \sum_{i} \frac{y^{(i,j)}}{\alpha_{B,j}} NB$$

$$y(NA, NB) = \sum_{i,j} y^{(i,j)}(NA, NB)$$

Hence, we are now ready to redefine the Pareto problem in terms of our new definitions which satisfy the matching constraints.

$$\max_{y^{(i,j)}(NA, NB) \geq 0} \sum_{i} \lambda_{A,i} \left[ \sum_{j} \sum_{(NA, NB)} y^{(i,j)}(NA, NB) \times NA \times UA(NA, NB) \right] +$$

$$+ \sum_{j} \lambda_{B,j} \left[ \sum_{i} \sum_{(NA, NB)} y^{(i,j)}(NA, NB) \times NB \times UB(NA, NB) \right]$$

Such that each agent is assigned to a platform with probability one (the counterpart to equation (3.10)).

$$\sum_{j} \sum_{(NA, NB)} \frac{y^{(i,j)}(NA, NB)}{\alpha_{A,i}} NA = 1 \forall i,$$

$$\sum_{i} \sum_{(NA, NB)} \frac{y^{(i,j)}(NA, NB)}{\alpha_{B,j}} NB = 1 \forall j$$

Such that the resource constraint is satisfied (the counterpart to equation (3.12)).

$$\sum_{(NA, NB)} \left[ \sum_{i,j} y^{(i,j)}(NA, NB) \times C(NA, NB) \right] \leq \sum_{T,s} \alpha_{T,s} \kappa_{T,s}$$

Therefore, the advantage of writing the Pareto problem in the above formulation reduces the constraint set, in this example, there are only five constraints, however, the number of variables is very large.

We can use a linear programming solver to compute the reformulated Pareto programme.
3.7.2 Additional Graphs

Figure 3-5: Redistributing wealth within and across type – the effects on all agents’ utilities

Redistributing wealth across type:
the effect on utility from transferring wealth from (A, 2) to (B, 1)

Redistributing wealth within type:
the effect on utility from transferring wealth from (B, 2) to (B, 1)

These figures are identical to figure (3-1), except we also include the agents whose wealth changes. In the left panel, the x-axis redistributes wealth from subtype (A, 2) to subtype (B, 1). As you would expect this increases (B, 1)’s utility and simultaneously reduces (A, 1)’s utility.

In the right panel, the x-axis redistributes wealth from subtype (B, 2) to subtype (B, 1). As you would expect this increases (B, 1)’s utility and simultaneously reduces (B, 2)’s utility.
Bibliography


