CHANGES IN SOCIAL NETWORK STRUCTURE IN RESPONSE TO EXPOSURE TO FORMAL CREDIT MARKETS

ABHIJIT BANERJEE†, ARUN G. CHANDRASEKHAR‡, ESTHER DUFLO§, AND MATTHEW O. JACKSON⋆

ABSTRACT. We study how the introduction of microfinance changes the networks of interactions in two settings: (1) among 16476 households in 75 villages around Karnataka, India and (2) among XXX households in 102 neighborhoods in Hyderabad, India. We then develop a new dynamic model of network formation to explain the empirical findings. In the Karnataka dataset, none of the villages were exposed to microfinance by 2006, 43 villages were through 2010, and we have a two-wave panel of network data collected in 2006 and 2012. The Hyderabad dataset is from a randomized controlled trial (RCT) in which half the neighborhoods were randomized to have a microfinance entry. In both cases, we compare changes in networks in villages/neighborhoods exposed to microfinance relative to those not exposed. Networks exposed to microfinance experience a significantly greater loss of links—for credit relationships as well as advice and other types of relationships—compared to those not exposed. Microfinance not only results in decreases in relationships among those likely to get loans, but also decreases in relationships between those unlikely to get loans. These patterns are inconsistent with models of network formation in which people have opportunities to connect with whomever they wish, but are consistent with a model that emphasizes chance meetings that depend on relative efforts to socialize coupled with conditional choices of whom to connect with, as well as externalities in payoffs across relationships between pairs and triples of people.

JEL CLASSIFICATION CODES: D85, D13, L14, O12, Z13

KEYWORDS: Social Networks, Network Change, Network Formation, Network Evolution, Microfinance, Market Exposure, Favor Exchange, Social Capital

1. INTRODUCTION

That formal institutions can crowd out informal ones, possibly to the detriment of many or even a majority of the population, is an old idea (Arrow, 2000; Putnam, 2000) but the empirical evidence is scant and with a few exceptions, rarely goes beyond simple correlations.

Date: June 4, 2019.

We thank Siwan Anderson, Patrick Francois, Rachel Kranton, and participants at various seminars for suggestions. Financial support from the NSF under grants SES-1156182, SES-1155302, and SES-1629446, and from the AFOSR and DARPA under grant FA9550-12-1-0411, and from ARO MURI under award No. W911NF-12-1-0509 is gratefully acknowledged. We thank Devika Lakhote, Varun Kapoor, Stephen Nei, and Mounu Prem, as well as the CMF at IFMR, Tanay Balantrapu, Gowri Nagraj, and Manaswini Rao for excellent research assistance. and helpful discussions.

†Department of Economics, MIT; NBER; JPAL.

‡Department of Economics, Stanford University; NBER; JPAL.

§Department of Economics, MIT; NBER; JPAL.

⋆Department of Economics, Stanford University; Santa Fe Institute; and CIFAR.

A challenge is to find detailed data on networks of informal relationships with variation in access to formal institutions. Using a unique data set with these features, we provide a detailed look at how a formal institution crowds out informal relationships in a specific context, and develop a new model of network formation that may help answer these questions more broadly. In particular, we see effects that reach well-beyond the people who are involved in the formal institution, and on relationships that have no direct connection to the institution, and argue that the observed patterns are informative about the theory of network formation.

Specifically, we examine how the introduction of microfinance affects the evolution of social networks in villages in rural India. Social networks are an important source of credit, insurance, information, advice, and other economic and non-economic benefits.[1] One would expect that people who get microfinance loans would change their networks of informal borrowing and lending. What is interesting and perhaps less obvious, is that the that exposure to microfinance induces widespread changes in social and economic networks, well beyond the people directly involved in microfinance borrowing and beyond relationships of borrowing and lending. We argue, based on this evidence, that network formation exhibits several key features: social effort is needed to maintain current relationships, forming new connections involves significant chance in meetings ("undirected search"), and different types of relationships are formed at the same time (multiplexing).

We make use of two datasets in our analysis. First, we use a panel that we collected (Banerjee, Chandrasekhar, Duflo, and Jackson, 2013, 2014b) over 6 years in 75 villages. In Wave 1, collected in 2006, no villages were exposed to microfinance. By Wave 2, collected in 2012, 43 villages were exposed to microfinance and the remaining 32 were not. Using detailed panel network data that we collected from the two waves (covering 16,476 households), we examine differences in the evolution of the networks between the microfinance and non-microfinance villages.

These 75 villages were selected because, in 2007, no microfinance was available in any of these villages, but a microfinance institution, Bharatha Swamukti Samsthe (BSS), intended to start operating in all of them. Between 2007 and 2010, BSS entered 43 of these 75 villages. We call these MF villages. The world financial crisis halted BSS's expansion and so they stalled expansion, and the remaining 32 villages were not exposed to BSS during our study. We call these non-MF villages. We take advantage of this variation to estimate the impact of microfinance on the villages, using a difference-in-difference strategy.

Second, we use a cross-sectional dataset that we collected (Banerjee, Duflo, Glennerster, and Kinnan, 2015a) in 102 slum neighborhoods in Hyderabad, India. This data comes from a randomized controlled trial (RCT). In it, 52 neighborhoods were randomized to receive

[1] See, e.g., Udry (1994); Fafchamps and Lund (2003); Karlan, Mobius, Rosenblat, and Szeidl (2009); Beaman and Magruder (2012); Ambrus, Mobius, and Szeidl (2014); Blumenstock, Eagle, and Fafchamps (2016); Munshi and Rosenzweig (2016); Blumenstock and Tan (2016); Breza (2016).

entry by a microfinance institution, Spandana, whereas the remaining 52 were randomized to have no entry and serve as a control.

The advantage of the Karnataka panel is that it consists of high-quality network data. We know details of link patterns between households as well as the nature of the link (e.g., financial, informational, social). However, the setting does not involve an RCT and therefore we rely on a difference-in-differences assumption in the analysis.

The Hyderabad dataset resolves this issue: whether Spandana enters a network is randomized. Therefore, differences across networks can cleanly attributed to exposure to microcredit. The downside in this dataset is that we do not have complete network data. Instead, what we collected is data sufficient to study network structure through the perspective of using Aggregated Relational Data (ARD) (CITEARD paper). That is, we collected data on how many links each respondent in the sample has and how many links to members of several categories the individual has. We also collected support (how many of one's friends have yet another friend in common). ARD allows us to identify parameters of a network formation model and then estimate what the neighborhood network might have looked like (the distribution of the network given the observables). CITE ARD PAPER has shown this to be an effective way of identifying network effects, almost as well as if the researcher had full network data when looking at a number of network features.

We find that the introduction of microfinance is associated with a 11% decline in the probability of a link between any two households ($p = 0.077$) in a MF village compared to a non-MF village. This is robust to controlling for a rich array of baseline variables (chosen via double post-LASSO (Belloni, Chernozhukov, and Hansen, 2014a,b)).

We then investigate how the changes in networks are distributed across two types of households: those who end up taking up microfinance loans and those who do not. A priori, we would expect a financial innovation to have different effects on those who end up borrowing from the MFI relative to those who do not. One might expect that those who become microfinance clients would curtail their financial relationship with others – in particular they might be less willing to lend small sums to their friends, because a microfinance loan requires weekly repayments, which takes discipline and induces financial stress (Field, Pande, Papp, and Park, 2012). Alternatively, it could also be that microfinance clients end up with extra money that they would lend to others. On the other side of the ledger, their own need to borrow from friends, including those not involved in microfinance, could potentially go up or down – down because they can get microfinance loans to cover large expenses, or up because they need short-term help to meet their repayment obligations (Field, Pande, Papp, and Park, 2012; Feigenberg, Field, and Pande, 2013). Such changes could affect their willingness to maintain friendships, including with those who do not take up microfinance.

Although microfinance could change relationships between those who get microfinance loans as well as their relationships with others, prima facie (without any sort of externality or spillover), one might not expect any changes in the relations between pairs of households neither of whom take up microfinance.

To explore how the network changes are distributed across people who take up microfinance and those who do not, we need to be able to compare to the equivalent groups in the non-microfinance villages. That is, we need to identify those who are more likely to and less likely to have become microfinance clients if they had the chance in the non-microfinance villages. To this end, we use a random forest model to classify households in MF and non-MF villages in terms of whether they would have a high (H) or low (L) likelihood of joining microfinance if were offered in their village, using eligibility rules and network position as source of identification.[2] This allows us to compare what happens to Hs in microfinance versus non-microfinance villages, as well as to Ls.

We begin by looking at the rate that two HHs that were linked in Wave 1 are linked in Wave 2 in microfinance villages compared to non-microfinance villages. We find an (statistically insignificant) decrease of 1.6pp relative to a mean of 44.1%. Next we look at LH links as access to microfinance could lead Hs to decrease their linking to others. In fact, we see a larger decrease in LH relationships (7pp) than in HH relationships in microfinance villages compared to non-microfinance villages and the difference is significant ($p = 0.00$).

The surprising result is that LL links also differentially decline in microfinance villages. An LL link that exists in the Wave 1 in a microfinance village is 7pp less likely to exist in Wave 2 than in a non-microfinance village; this decline is significantly greater than the decline in HH links ($p = 0.003$). If anything, one would have expected LL relationships to increase in microfinance villages, since they are losing access to Hs and still have needs to borrow and lend.

A priori, one potential explanation of this comes from a model of network formation that predicts correlation in adjacent links (Jackson, Rodriguez-Barraquer, and Tan, 2012). In that setting a relationship between two Ls could be 'supported' by an H. The idea is that the relationship between two Ls can only be sustained of their mutual friendship with some other household. That friend-in-common can help support incentives to borrow and lend between the Ls because if one of the Ls fails to help the other, then that L also risks losing the relationship with the mutual friend, for instance an H. The theory predicts that we should see triangles of relationships that are mutually dependent (or even more elaborate

[2]This can be thought of as a machine-learning version of standard propensity scoring. We use network position because the MFI's procedure was to enter a village by contacting certain leaders first (e.g., teachers, leaders of self-help groups, shopkeepers), network distance to such leaders influences the odds that the household hears about microfinance (Banerjee, Chandrasekhar, Duflo, and Jackson, 2013), even after controlling for their own network position. Eligibility rules are obviously useful.

structures called quilts). So this provides one potential explanation for the disappearance of the LL links: if they appear in HLL triangles, then when HL links disappear, then the LL links also disappear.

We do find some support for 'support'. In general, links that are connected in triangles are significantly correlated in their appearance/disappearance. For instance, the probability that an LLL triad survives is over twice what would be predicted using the probability that three independent adjacent LL links survive.

This is however not the end of the story. When we examine the disappearance of triangles, we see that it is the LLL triangles that are most likely to disappear in MF villages compared to non-MF villages. In MF villages, the probability that an LLL triad survives is 10.6pp less than the probability that an HHH triad survives ($p = 0.001$) using the corresponding triads in non-MF villages as controls. Furthermore, the probability that at least one link in the Wave 1 triad disappears after microfinance exposure, in MF villages relative to non-MF, is highest for LLL, and the difference with LLH and LHH triads is positive and significant ($p = 0.009$ and $p = 0.010$, respectively). Thus, the greatest effect of microfinance on network links is among the people least likely to be involved in microfinance, and even in parts of the network that are not directly dependent upon relationships with Hs. This spillover extends beyond what is predicted by the support model of Jackson et al. (2012), and is also inconsistent with a variety of other network formation models (as we discuss in below and in Appendix C).

To provide an explanation for this phenomenon, we develop a new model of network formation that can explain why links between the L's would decline at the fastest rate. The model we build comes from a simple idea, which fits with our anecdotal impressions of how socializing works in the villages. In the model, old relationships are maintained and new ones are formed when people socialize in an "undirected" way. A stylized interpretation is that people show up at the town square, or a local tea shop, to 'hang out' and socialize. Seeing their current friends keeps those relationships intact, and meeting new people sometimes results in new relationships. People who do not show up at the town square lose old relationships and form fewer new ones.

The key externality is that the returns to socializing depend on who else is socializing. L types care about how H types socialize and the value of HL links, and changes in H types' socialization in response to microfinance then changes incentives for L types to socialize - which changes the incidence of LL links. In particular, the availability microfinance changes H types access to credit, and as a result could either increase or decrease their willingness to borrow and lend. For instance, because microcredit requires regular repayments, small loans to friends may become more burdensome and therefore H types who take out microfinance loans may cut back on their lending. This is, in fact, consistent with what we see: the probability of lending declines for H types in microfinance villages and while their probability of borrowing from a network member stays the same (or possibly declines). As such, the value of links to H types for both H and L types declines. As Hs become less willing to lend, this can have a larger negative impact on Ls rather than Hs, which leads to less socializing by Ls. As Ls socialize less there is a larger relative drop in LL links. A simple extension of the model to account for the formation of triads (triangles) rather than just dyads lead to similar results for LLL relationships.

This model matches the patterns we observe the data village exposed to microfinance. To explain why other standard models cannot deliver these results, we explore four versions of alternative models of network formation that do not have "undirected" search. These are described in Online Appendix C.

Another prediction of this class of models is that there should be spillovers across different types of relationships, since it is the same town square where people also form other relationships such as advice relationships. We see almost exactly the same patterns of disappearance in relationships of advice as in borrowing/lending relationships.

Finally, as a check on what the model is predicting, we also examine changes in where people borrow from: friends, family, MF, self-help groups, or money-lenders. We find that H households, as expected, borrow significantly more from microfinance than L households in microfinance villages. And consistent with the disappearance of the LL links, the L households, after the introduction of microfinance, borrow relatively less from friends in Mf villages compared to non-MF. For instance, for an L with only one L friend in the baseline, we see a Rs. 2582 ($p = 0.02$) decline in loans from friends relative to their counterparts in non-MF villages and in fact the decline in loans from friends exceeds the corresponding decline for the Hs (Rs. 1214, which is significantly less than Rs. 2582, $p = 0.09$).

Our research on how exposure to formal credit affects social and economic networks is related to some important recent and ongoing work. Feigenberg et al. (2013) find that participation in microcredit creates tighter social relationships among group members. Binzel, Field, and Pande (2013), Banerjee, Breza, Duflo, and Kinnan (2014a), and Comola and Prina (2014) explore whether and in what ways interventions that affect household finance decisions also end up affecting those households' networks. Binzel et al. (2013) look at network effects in a randomized roll-out of bank branches in India. Their focus is on whether individuals are less likely to make transfers to their friends in a non-anonymous dictator game after being exposed to the financial institution. Comola and Prina (2014) study how individuals' social networks change when randomly assigned to receive a savings account in Nepal. Their focus is on looking at post-period expenditure spillovers, taking into account network change due to the exposure to the savings account. Another, recent and independent, study that examines how policy interventions affect network structure is Heß, Jaimovich, and Schündeln

(2018). Unlike our study, they look at a community driven development initiative (CDD). The CDD involved providing a very large disbursement—one half of the annual income per capita— *per household* in each treatment village. The villagers had to collectively decide through the CDD which projects to execute. Heß et al. (2018) collected a cross-section of network data in 2014 and, like us and prior studies, document declines in network density and closure. The CDD is subject to political maneuvering and so the researchers examine how the scope for elite capture affects the network structure. A key difference between the CDD versus the microcredit/savings contexts is that the injection of the former is massive and at the community level whereas the micro-loans are both smaller in size and only apply to a small subset of the community; so the general equilibrium effects on network structure come from very different sources and for different reasons.[3]

Our study contributes to and extends this line of inquiry. Our main distinction is the emphasis on spillovers on non-participants and other types of relationships. We also use the evidence of those spillovers to argue for a particular model of network formation. Our Karnataka data is also exceptionally rich: we have a panel of nearly 16,476 households with 71% of links observed in Wave 1 and 99% observed in Wave 2. And our Hyderabad data combines a large-scale RCT with novel data that allows us to use ARD techniques to estimate effects of treatment on the network.

A main lesson from our paper is that presence of significant and widespread interdependencies and spillovers in network formation across both types of people and types of relationships. The nature of these interdependencies suggests that some form of undirected search is needed to explain the data. This bears out a long-standing prediction of theoretical models of network formation. Networks are not designed but result from the decentralized decisions of individuals and, as the evidence in this paper shows, a shift in the incentives of one group of people to form links can have substantial (negative) effects on other parts of the network and groups that they ignore when choosing their own behavior.[4]

Our work also contributes to the literature on network formation in two ways, both by introducing a new model and showing why the various features of that model are needed (at least in this application). Previous models of network formation that involve explicit choice in the process[5] can be thought of as coming in three flavors:

(i) models in which people have the opportunity to connect with whomever they want, subject to reciprocation (e.g., Jackson and Wolinsky (1996); Dutta and Mutuswami

[3]Nonetheless, our model could still be useful in understanding the effects in such larger and more pervasive interventions.

[4]Of course, this does not mean that microcredit should be discouraged, but only any welfare analysis needs to be take into account the potential for spillovers.

[5]There is also a large literature of network formation that involves no strategic choice, but just a stochastic model of network formation/evolution (e.g., see Jackson (2008) for some description and references). Those models are not equipped to match the data here.

(1997); Bala and Goyal (2000); Currarini and Morelli (2000); Jackson and Van den Nouweland (2005); Bloch, Genicot, and Ray (2008); Herings, Mauleon, and Vannetelbosch (2009); Jackson, Rodriguez-Barraquer, and Tan (2012); Boucher (2015)...);

(ii) models in which there are exogenously random meetings and then conditional upon meeting people choose with whom to connect (e.g., Watts (2001); Jackson and Watts (2002); Christakis, Fowler, Imbens, and Kalyanaraman (2010a); König, Tessone, and Zenou (2014); Mele (2017)...); and

(iii) models in which people put in some efforts to socialize which then results in some random meetings, but then relationships are formed as a result of those efforts without further choice (e.g., Currarini, Jackson, and Pin (2009, 2010); Cabrales, Calvó-Armengol, and Zenou (2011); Canen, Jackson, and Trebbi (2017)).

The empirical patterns that we observe here require models that are richer than (i) and (ii), and require some externalities in the efforts to search and meet. Models in class (iii) that only involve those efforts and no further choice also have problems generating the differential patterns that we see among HH, HL, and LLs. The model that we introduce has features of all three of these classes: effort is needed to meet others and affects the relative rates at which people are randomly met, but then choice is involved conditional upon meeting. In addition, our model has two other features that help us to match the data. One is that effort not only is needed to meet new people, but also to maintain existing relationships - as the patterns we observe in the data exhibit similarities both in terms of which relationships are retained and which new ones are formed. The second is that socializing affects the opportunities to form multiple types of relationships at the same time.

The combination of all five of these features – efforts to socialize, rates of meetings dependent on relative efforts, mutual choice required to form relationships conditional upon meeting, effort needed to maintain relationships, and multiple types of relationships formed at the same time – allow us to capture all of the nuances and rich patterns that we observe in the data. In an appendix, we discuss why dropping any one of these features would fail to capture some aspects of the data.

The remainder of the paper is organized as follows. In Section 2, we describe the setting, network data collection, and the classification of households into H and L types using a random forest algorithm, and sample statistics. Section 3 presents our main empirical results. Motivated by the data, in Section 4 we develop a new dynamic model of network formation that is consistent with it and discuss why four standard models from the literature are inconsistent with the data. Section 5 concludes.

2. Setting, Data and Sample Statistics

2.1. **Setting.**

2.1.1. *Karnataka.* In 2006, the microfinance organization, BSS provided us with a list of 75 villages in Karnataka in which they were planning to start lending operations. The villages were spread across 5 districts of the state of Karnataka in India. Prior to BSS's entry, these villages had minimal exposure to microfinance.

Six months prior to BSS's entry into any village, in 2006, we conducted a baseline survey in all 75 villages. This survey consisted of a village questionnaire, a full census that collected data on all households in the villages, and a detailed follow-up survey fielded to a subsample of adults.

By the end of 2010, BSS had entered 43 villages. The choice of 43 villages was not randomly assigned by us, but chosen by the bank. We have anecdotal reasons to believe that the choice was not systematic: BSS clearly planned to enter all of the villages but slowed down and ultimately stopped expanding during the Andhra Pradesh microcredit crisis.

There is no measurable difference in network structures or demographics between the microfinance villages on non-microfinance villages, as seen in Table 1. The villages that they entered are similar to the ones that they did not on almost all dimensions, the only measurable difference is that the MF villages have, on average more households than the non-MF villages.

2.1.2. *Hyderabad.* In 2006, a large lender, Spandana, was randomized to enter 52 of 104 neighborhoods in Hyderabad, India. After two years, the remaining 52 neighborhoods received access in mid-2008. The study of the short and long run impacts of randomized access to microfinance are studied in Banerjee et al. (2015a).

It is important to note that the Andhra Pradesh microcredit crisis also impacted Spandana. Indeed Hyderabad was in Andhra Pradesh (currently serving as the interim capital of Andhra Pradesh but now in Telengana after the state was split in 2014). So al of the respondents faced simultaneous withrdawal of microcredit in response to the 2010 Andhra Pradesh ordinance to halt microcredit loans. A follow-up data collection was done in 2012, with a sample of 5744 households. At the time the treatment neighborhoods were exposed to microcredit for 6 years (4 years active lending) and the control neighborhoods were exposed for 3.5 years (with 1.5 years active lending).

Table XXX shows that treatment and control neighborhoods are balanced on a number of characteristics. FIXME EXPAND.

2.2. **Data.**

2.2.1. *Karnataka.* To collect the network data,[6] we asked adults to name those with whom they interact in the course of daily activities. In Wave 1, collected in 2006, we have the full village census (enumerating every individual in every household in every village) and network data from 46% of households per village. In Wave 2, collected in 2012, in addition to the census, we have network data from 89.14% of the 16,476 households based on interviews with 65% of all adult individuals aged 18 to 50. This means that we have network data in Wave 1 on 70.8% of the links and in Wave 2 on 98.8% of the links when we build the undirected, unweighted graph that we study. For most of the analysis, we concentrate on households that are present in both waves.

We have data about 12 different types of interactions for a given survey respondent: (1) whose houses he or she visits, (2) who visits his or her house, (3) relatives they socialize with, (4) non-relatives they socialize with, (5) who gives him or her medical help, (6) from whom he or she borrows money, (7) to whom he or she lends money, (8) from whom he or she borrows material goods (e.g., kerosene, rice), (9) to whom he or she lends material goods, (10) from whom he or she gets important advice, (11) to whom he or she gives advice, (12) with whom he or she goes to pray (e.g., at a temple, church or mosque).

Using these data, we look at the financial network (a union of (6-9) above) to begin with as well as the informational network ((10-11) from above). After demonstrating that links across both categories change and change in similar ways, we proceed by aggregating the network data as follows. We construct one network for each village, at the household level where a link exists between households if any member of either household is linked to any other member of the other household in at least one of the 12 ways. We assume that individuals can communicate if they interact in any of the 12 ways, so this is the network of potential communications. The resulting objects are undirected, unweighted networks at the household level. More specialized networks are discussed later.

2.2.2. *Hyderabad.* Collecting complete network data was infeasible across the 102 neighborhoods in Hyderabad, so instead we collected Aggregated Relational Data (ARD).

Specifically, about 55 nodes in every neighborhood was surveyed. These respondents were asked ARD questions. First, they were asked how many links they had within the neighborhood (eliciting their degree). Second, they were asked questions of the form "How many individuals from your list of friends do you know who have trait X?" This gives degree-by-trait. In the census instrument all households were surveyed about all of the k ARD traits. Third, we asked for each friend, with how many do you have a friend in common, to have support.

[6]The Wave 1 data are described in detail in Banerjee, Chandrasekhar, Duflo, and Jackson (2013) and publicly available at `http://economics.mit.edu/faculty/eduflo/social`. The Wave 2 data will be available upon publication.

The value of ARD is discussed in CITE ARD PAPER. The basic idea is that ARD by knowing for each node in a sample, how many nodes of a given trait that it is linked to, one can identify a simple model of network formation where every node links to every other node as a function of "latent distance." Basically, by knowing how many nodes of each type a given respondent knows, we can triangulate the locations of which nodes are located where in a latent space, knowing their characteristics, and therefore have an estimate of the distribution of network formation. CITE ARD PAPER shows that we can replicate the results of SAVINGS MONITORS as well as if we saw the entire network data and also because we directly measure support as well, we can show that the ARD estimate leads to the identical conclusions.

We list the ARD questions used in Appendix XXX.

2.3. **Sample Statistics.**

2.3.1. *Karnataka.* Table 1 presents the sample statistics. Panel A presents summary statistics for the Wave 1 data. The networks are sparse: the average degree is 14.8 in microfinance villages and 13.4 in the non-microfinance villages, despite having about 200 households on average. This difference is statistically significant, though not conditional on controlling for village size (number of households). The average clustering coefficients (the percent of time two of a household's friends are themselves friends) are 0.28 and 0.31 respectively, and the difference is not statistically significant. Finally, these networks have short distances: the average closeness (the mean of the inverse of path lengths, with 0 taken for nodes on different components) is 0.366 and 0.378, respectively and the difference again is not statistically significant across microfinance and non-microfinance villages.[7]

Panel B looks at the Wave 2 data. We find that the average degrees are higher 6 years later, though comparable across samples (17.3 and 17.5 for microfinance and non-microfinance villages). The average closeness has increased across both samples to 0.448 and 0.484, respectively, though the increase is higher in the non-microfinance sample.

We also asked, in both Wave 1 and Wave 2, for households to give us a list of all outstanding loans that they have taken, the sources of these loans (e.g., family member, friend, microfinance institution, self-help group, money lender) and their terms. We use this to create a panel to study changes in borrowing patterns.

In our analysis we look at all households who existed in Wave 1 (and in Wave 2 as well). This involves those who remained and those who split. 11% migrated out, though this is not differential by microfinance exposure, and 4.8% Wave 2 households in-migrated or split off

[7]In order to deal with the fact that we sampled data in Wave 1, we compute average degree among the sampled households in Wave 1. We compute the clustering coefficient among the subgraph induced by restricting to sampled households in Wave 1, since that is centered around the true parameter. It is also worth noting that the correlation among the different link types (specifically multiplexing of information and financial links) is 0.638.

from existing households (as children reach adulthood), again not differential by microfinance exposure.

2.3.2. *Hyderabad.* FIXME!!!

2.4. **Classifying Nodes as H and L.** In order to study heterogeneity in effects by propensity to participate in microfinance, we need to identify which households would have had taken out microfinance loans in the non-microfinance villages or neighborhoods, had BSS or Spandana entered those villages. To do this, we make use of determinants of access to microfinance in a random forest model, as well as detailed household demographics.

Let us begin with the Karnataka setting. One obvious determinant is from the BSS rules: only households with an female in the age range 18-50 were eligible for microfinance. Also, certain households were identified by BSS as a "leader" household and were specifically informed about the product. Another criterion we use is based on the argument in Banerjee et al. (2013), that being closer in the network to BSS leaders makes it more likely that the household would have heard of the microfinance opportunity and would have taken it up.[8] We estimate the random forest model based on household demographic and network characteristics from the microfinance villages on a training sample of 7199 households and then validate the method on a testing sample of 2399 households. The details of the estimation algorithm and implemented choices are presented in Appendix B.

We can then apply the classifier to both microfinance villages and non-microfinance villages to classify each household as H or L (high or low likelihood of joining microfinance).

As mentioned above, our choice of characteristics is motivated by BSS's village entry strategy. The variation used to classify whether a household is of type H or L comes from the process by which the microfinance institution selected initial households to start the diffusion project which are typically different from intrinsic features of household demand as argued in (Banerjee et al., 2013). The features are as follows: (1) a dummy for the household being a BSS leader, which are households with an individual that the microfinance institution would approach when entering a village; (2) a dummy for whether the household has a female of eligible age (below 50), which was a requirement to be able to participate in microfinance; (3) the average closeness (mean of inverse of network distance) to leaders, which is relevant because as in Banerjee et al. (2013) those who are closer to leaders should be more likely to hear of microfinance; (4) the average closeness (mean of inverse distance) to same-caste leaders, because interactions within-caste are more likely and therefore should influence the likelihood of being informed; and (5) the share of same-caste leaders in the village.

Turning to the Hyderabad setting, the strategy is similar. FIXME DESCRIBE!

[8]Our theory is that hearing about it from someone provides both information about the product and some understanding of how it works. In fact, in a small follow up survey we find that even two years after the expansion about 17% of households in microfinance villages had even not heard of BSS.

An advantage of using random forests is that they naturally allow for non-linearities and potentially complex interactions between characteristics that could drive microfinance take-up. Alternatives such as logistic regression would not be able to handle such interactions without typically introducing a very high dimensionality of interaction terms, perhaps limited by variable selection techniques that impose sparsity. Essentially, logistic regressions have a log-odds function that is linear in parameters, so a classification imposes a hyperplane whereas random forests allows for complex partitioning of the data space to classify observations. Disadvantages of random forest include that the actual mapping from characteristics to classification are less interpretable with random forests than with logistic regression models and that if the true underlying data generating process has log-odds that are linear in parameters, then the random forest may overfit. Given that we are using this simply to classify households based on likelihood to take up microfinance, this disadvantage does not matter (think of this as a propensity scoring analysis).

A second advantage comes from its value in identification. Random forests allow for classification via a complicated non-linear function of the network and relation to leadership positions, which subsequently allows us to control smoothly for network position. Therefore, unobservables correlated smoothly with network parameters are unlikely to drive the results.

Table 2 presents some summary statistics. There are notable differences betwen H and L households. Although none of these features were used in the estimation, we find that H households are much more likely to be SCST, have smaller houses in terms of room count, much less likely to have a latrine in the household, and much less likely to have an RCC (reinforced concrete cement) roof, all of which suggests that they tend to be poorer. Finally, we see that H households and L households have comparable degrees, but the composition exhibits homophily: H types have a lower number of links to L types and a higher number of links to H types. Finally H households are more eigenvector central in the network, which again is consistent with the idea that these households should be more likely to hear about microfinance. In Section 3.5 we show that indeed H types borrow considerably more than L types in microfinance villages. H types borrow Rs. 1731 ($p = 0.00$) whereas the L types borrow Rs. 320.8 (not significantly different from zero) indicating that the classification performs well.

FIXME:

- ADD SUMMARY ABOUT HYDERABAD
- ADD A CLASSIFICATION CONFUSION MATRIX FOR BOTH CASES
- HYDERABAD HAS TO HAVE THE BORROWING BY CLASSIFICATION TYPE

Nonetheless, we present our main results in Appendix **??** using logistic regression to classify households into H and L types. The results are concordant with our main results.

3. Changes in Networks

How does exposure to microfinance change networks? We begin with a discussion of how links of various types are affected, and then we examine how triads (triangles) are affected as they embody a basic form of spillovers.

3.1. **Effect on the total number of links.** We first look at how introducing microfinance affects the overall structure of the village social networks in a difference-in-differences framework:

$$y(\mathbf{g}_{vt}) = \alpha + \beta \text{Microfinance}_v \times \text{Post}_t + \gamma \text{Microfinance}_v + \eta \text{Post}_t + \delta' X_v + \epsilon_{vt},$$

where $y(\cdot)$ computes the density of the network $\mathbf{g}_{vt}$ for village v in period t, the average closeness (the mean of the inverse distance between all pairs), or clustering. The density is the percentage of links a random household has to all other households in the village, so it measures how well-connected the village is on average. The distance in the network is the (minimum) number of steps through the network it takes to get from one household to another. In models where favors, transactions, or information travels through the network, higher distance or lower closeness means that the movement of such phenomena through the network is slower. Finally, clustering is the share of a household's connections that are themselves connected. Economic models of network formation identify clustering as an important feature to sustain cooperation.

Table 3 presents the results. Columns 1-2 present the result for network density, columns 3-4 for clustering, and 7-9 for closeness. The first column in each category (columns 1, 4, 7) present a simple difference in differences specification. The second column in each specification (2, 5, 8) adds to that a vector of baseline controls as well as the controls interacted with treatment. These controls are the share of upper caste households, number of households in the village, share of households in self-help groups, share Hindu, share with a latrine in the house, share that own the household, share that have electricity and share that are leaders. We add these because differences in the size of the village, caste composition, or the wealth distribution could potentially have networks evolving differently even without introduction of microfinance. While the entry of BSS does not seem to correlate with much of anything beyond village size, we include these controls to ensure that they do not drive the results. Finally, the third column in each specification (3, 6, 9) includes village fixed effects as well as controls for the baseline value of the outcome variable interacted with Post, to allow for differential time trends by baseline network feature. Because we only have 150 observations but many controls (up to 18 controls and their interactions before adding the fixed effects), we use the double post-LASSO procedure (Belloni and Chernozhukov, 2009; Belloni, Chernozhukov, and Hansen, 2014a,b) to select the controls.

We find that exposure to microfinance leads to a drop in density by about 1.2-1.3pp relative to a a mean of 11.9% in non-microfinance villages in Wave 1 (columns 1-3, $p = 0.077$ in column 3 for example). This is a 10.8% drop in density.

In columns 4-6, we find that there is no detectable effect of microfinance on the clustering of the village. This is true irrespective of whether controls are used.

Columns 7-9 present the result for closeness, the average of the inverse of path lengths in the village. Without controls we find a significant decrease in the average closeness (column 7, $p = 0.02$), though this loses significance in columns 8 and 9 with the inclusion of controls ($p = 0.19$, $p = 0.21$, respectively). Having microfinance decreases the average closeness by 0.022, which corresponds to a 0.53 standard deviation effect relative to the Wave 1 (column 1). The point estimates in columns 2 and 3 are comparably large, roughly 0.38 of a standard deviation.

Taken together, the results point to large changes in aggregate network structure such as density, though some of the estimates are noisy. Our main empirical results, however, involve looking at the difference in how connections change for potential borrowers and non-borrowers within MF and non-MF villages and comparing the two, where we have higher precision.

3.2. **How are links affected by microfinance?** In this subsection, we explore how microfinance exposure affects the formation of links across types of households – our Hs and Ls.

Bilateral links can be of three types: HH, LH, and LL. We first ask whether a link that was present in the first wave of data *before* microfinance is still present in the second wave of data *after* microfinance, and how this depends on whether the village was exposed to microfinance and the type of the link. Let $g_{ij,v,t}$ be an indicator for whether a link is present between households i and j in village v in wave t. Letting LH_{ij} be an indicator for pair consisting of one low type and one high type, and analogously for LL_{ij} etc., the regression we run takes the form

$$\begin{aligned} g_{ij,v,2} = \alpha + \beta MF_v + \beta_{LH} MF_v \times LH_{i,vj} + \beta_{LL} MF_v \times LL_{ij,v} \\ + \gamma_{LH} LH_{ij,v} + \gamma_{LL} LL_{ij,v} + \delta' X_{ij,v} + \epsilon_{ij,v,2}, \end{aligned}$$

where $X_{ij,v}$ includes a vector of flexible controls (a polynomial) for centrality of both nodes, demographic variables (caste and a number of wealth proxies including number of rooms, number of beds, electrification, latrine presence, roofing material), all variables that are used in the random forest classification, and then interactions of all of these network, demographic, and classification variables with microfinance.

The idea behind identification is that the classification type, H or L, is a non-linear function of a subset of the features described above. As such, we can still smoothly control for them and allow the control to vary by whether the village is exposed to microfinance or not. This allows us to control for the potentially differential effect of microfinance exposure on households that are demographically distinct and located differently in the network when we estimate the coefficient of interest. The coefficients of interest capture whether being in a microfinance village differentially affects the evolution of a link among types classified as HH, HL, and LL, conditional on all the characteristics above.

We also present regressions without any controls whatsoever to demonstrate that the results are robust to the presence or absence of these detailed controls. As argued in Altonji, Elder, and Taber (2005) this provides some support for the view that unobservables are not spuriously driving the results.

Next we do the opposite, focusing on the formation of new links where there were none. Specifically we start with the set of ij such that $g_{ij,v,1} = 0$ (so the link doesn't exist in the first period), and ask about the formation probability in Wave 2 as a function of exposure to microfinance.

Table 5 presents the results. In columns 1-3 we focus on the set of links existing in Wave 1 and in columns 4-6 we focus on the set of unlinked nodes in Wave 1. Columns 1 and 4 include no controls whatsoever. Columns 2 and 5 flexibly control for the average centrality of the nodes and interaction with microfinance. By controlling smoothly for network position, and the interaction with microfinance, we control for the varying effect of microfinance on households which are rather different. In columns 3 and 6 we further include a full set of household demographics that involve the variables utilized in the prediction exercise as well as their interactions with microcredit. Note that the classification dummy is a highly non-linear combination of all of these factors which we can control for smoothly and even allow the effect of microcredit to vary heterogenously by these factors. Furthermore, as we show below, since the results without controls and with a large set of flexible controls do not look very different, this supports the perspective that unobservables do not spuriously drive the result (Altonji et al., 2005), and that even if we had further controls we would not expect to have different results.

The key coefficients for testing the hypotheses are the Microfinance×LL and Microfinance×LH coefficients, as well as "Microfinance" itself in columns 1 and 4. The remaining columns present interactions with a large vector of covariates rendering the individual coefficients unuseful for interpretation. Within the set of links in Wave 1 we find that being in a microfinance village increases the rate at which links of all types break). HH links (the base category – so the coefficient of microfinance) do not have a significant change in MF villages relative to non-MF villages. However LH links decrease by 6.7pp in microfinance villages as compared to a base of 47.6% in non-microfinance villages ($p = 0.00$, column 1). We also see that LL links break significantly faster in MF villages than in non-MF villages: the probability of the link surviving is 6.2pp less on a base of 48.3% in non-MF villages ($p = 0.001$). LL and LH links both break significantly faster than HH links in MF villages relative to non-MF villages, but there is no statistically or economically significant difference in their rate of breakage.

We see that all results are robust both to the inclusion of flexible controls for node network position (column 2) as well as a large set of demographics and their interactions with microfinance, including predictors in the classification algorithm (column 3).

Columns 4-6 present similar results for link formation. Beginning with HH links, we see that they are 1.7pp less likely to form in microfinance villages on a base of 8.9% in non-microfinance villages ($p = 0.051$). LH links are 2.2pp less likely to form in microfinance villages on a base of 7.7 in non-microfinance villages ($p = 0.001$). Finally, LL links are 2.1pp less likely to form on a base of 7.5% ($p = 0.012$). New links are less likely to be formed in microfinance villages and this is equally true across all three types. Once again, all of these results are robust to smoothly controlling for the centrality of nodes involved as well as demographic controls and their interactions with microfinance.

For robustness, Table D.1, Panel A repeats the exercise using the graph removing any links to kin with whom the household socializes, which perhaps are less likely to change, while Panel B repeats the exercise using only the kin whom they socialize with network. We see that the results are robust to the exclusion of kin links, but reassuringly there is little action in the kin tables.

The relative changes in network structure in the microfinance villages sheds light on network formation. The fact that the LH links break might reflect the fact that the Hs are no longer interested in maintaining their links with the Ls now that they have an alternative source of credit, but the fact that LL links are equally likely to break is more surprising, especially since the Ls should have a stronger incentive to hold on to their mutual links precisely because they no longer have access to the links with the Hs.

It might seem puzzling that Ls are dropping links of all types even though they are not the ones who are getting microfinance. So, what explains the drop in L links, especially LL links?

3.3. **Triangles and Supported Relationships.** There are many reasons to expect that links are dependent. For instance, Jackson, Rodriguez-Barraquer, and Tan (2012) introduce the notion of support which correlates the presence of links based on incentives to exchange favors (including lending to each other). The idea is that two households in isolation may not be able to sustain cooperation amongst themselves, but if they both also have relationships with some other households in common, then fear of losing those relationships if they misbehave provides added incentives to maintain cooperation. If someone deviated from the expected behavior, say by not providing a loan or a favor and the link were severed, they would end up losing other relationships in response as well.

We begin by first providing evidence that support or support-like forces are likely to be important in this setting. To do this, we first want to understand the survival probabilty of an LL link that is not connected to any H. So we consider LL links and regress whether a link $g_{ij,2}$ exists in Wave 2, given that it existed in Wave 1, on whether the village had microfinance, whether the households had links in common, and interactions. Specifically,

$$\begin{aligned} g_{ij,v,2} = \alpha &+ \beta MF_v + \beta_{FIC}\text{No. Friends in Common}_{ij,v} \\ &+ \beta_{FIC,MF}\text{No. of Friends in Common}_{ij,v} \times MF_v \\ &+ \beta_H \frac{\text{No. of High FIC}}{\text{No. of FIC}}_{ij,v} + \beta_{H,MF}\frac{\text{No. of High FIC}}{\text{No. of FIC}}_{ij,v} \times MF_v + \epsilon_{ij,v,2}, \end{aligned}$$

Table 6 presents the result. A quick calculation shows that when LL are unsupported the probability of survival is 0.421. This means that the probability that a triangle survives is $(0.421)^3 = 0.074$ if relationships formed independently. Below we calculate what the probability of an LLL surviving is in the data.

In Table 7 we look at how triangles evolve and presents the results of a regression

$$\begin{aligned} y_{ijk} = \alpha &+ \beta MF_v + \beta_{LHH} MF_v \times LHH_{ijk} + \beta_{LLH} MF_v \times LLH_{ijk} + \beta_{LLL} MF_v \times LLL_{ijk,v} \\ &+ \gamma_{LHH} LHH_{ijk,v} + \gamma_{LLH} LLH_{ijk,v} + \gamma_{LLL} LLL_{ijk,v} + \delta' X_{ijk,v} + \epsilon_{ijk,v,2}, \end{aligned}$$

where y_{ijk} is either a dummy for whether the triangle ijk exists ($g_{ij,v,2}g_{jk,v,2}g_{ik,v,2} = 1$) in some specifications or whether any link in the former triangle exists ($g_{ij,v,2} + g_{jk,v,2} + g_{ik,v,2} > 0$) in other specifications and where $X_{ijk,v}$ includes flexible controls for centralities of households, demographic characteristics previously described for all households, all classification variables used in the random forest model and the interactions of all of these variables with microfinance. As before, we also present regressions without any controls to demonstrate that the results are not sensitive to the inclusion of controls and therefore are, under appropriate assumptions, less likely to be a product of spurious correlation from unobservables.

From Table 7, the probability of LLL surviving is 0.142, which is significantly greater than the 0.074 rate under independence (column 1). This suggests that the rate at which we see joint triangle survival is nearly 2 times as high than expected if links were conditionally independent.

Moreover, we also see in Table 7, that all triads (except the HHH) break faster in microfinance villages relative to non-microfinance villages, which is consistent with a theory of support.

However, LLL triads are actually the most likely to break in microfinance villages and not the least likely. This is not what would be expected if the only source of externality from microfinance was the breaking of links with type H villagers, where it would be the triads supported by H that we would expect to break apart. The rate of dissolution of LLL triads is always higher than the HHH and the difference is statistically significant ($p = 0.001$, column 3). When we look at the more general question of whether a triangle loses at least one link, the point estimate of the dissolution rate for LLL is higher than LHH or LLH, with the latter difference being significant ($p = 0.54$ and $p = 0.08$, respectively). However, when we look at whether any link survives we see that the odds that the entire triangle completely dissolved is significantly higher with LLL as compared to LLH or LHH ($p = 0.009$ and $p = 0.010$ respectively). LHHs on the other hand break up at the same rate as the LLHs. These results are robust to removing kin links (Table D.3), while looking at only the kin graph has no such effect (Table D.4).

Our data supports the idea that support matters: triangles are more than just three independent links. However, it also tells us that the exposure to microfinance does not only affect mixed triangles through the support externality–it actually affects the LLL triangles the most. Therefore, to explain what we observe we need a model where the incentive of Ls to form links with other Ls must be adversely affected by the change in incentives and behavior of Hs.

3.4. **Financial and other form of relationships.** Our network data comprises 12 different types of relationships, and for most of the analysis we consider that two households are linked if they are tied together via any of those relationships. It is interesting, however, to see if the effect are driven primarily by financial relationships (which are perhaps more likely to be differentially affected by microfinance).

In Table 4 we replicate the results of Table 5 but break out the network by link type separating two sets of graphs: ones involving financial links (which are the links that are directly affected) and ones involving information (those on which agents report seeking or giving advice about important decisions). What we see is that the effect of microfinance exposure is remarkably similar on both information links and financial links. While the two types of links are certainly correlated (the correlation is 0.638), they are far from identical. This is evidence of another kind of spillover—that on different types of relationships.

3.5. **Impact on Borrowing Patterns.** Next we examine how the amount borrowed from several sources of credit (microfinance, friend, self-help group member, family member, or money lender) depends on exposure to microfinance. This should respond in a manner corresponding to the changes in network structure.

We have data on the amount borrowed by source for the entirety of our sample. As such, we begin by regressing the amount borrowed on dummies for microfinance village, post, and household type:

$$y_{ivt} = \alpha + \beta_1 \text{MF}_v \times H_{iv} \times \text{Post}_t + \gamma_1 \text{MF}_v \times \text{Post}_t + \gamma_2 H_{iv} \times \text{Post}_t + \gamma_3 \text{MF}_v \times H_{iv} + \delta_1 \text{MF}_v + \delta_2 H_{iv} + \delta_3 \text{Post}_t + \epsilon_{ivt},$$

where again y_{ivt} is the amount borrowed from the stated source (MFI, friends, self-help group, family, moneylenders).

Table 8 presents the results. We find that L households lose Rs. 1943 ($p = 0.1$) in loans across their entire network (friends, self-help groups, and family) after being exposed to microfinance. Though the decline in loans for H is not statistically distinguishable from the decline for L, the point estimate shows that differentially due to exposure to microfinance H only lose Rs. 967 overall in loans. The starkest relative decline is among friend loans, where L lose Rs. 1046 ($p = 0.045$) whereas H lose only Rs. 420. The difference between the H and L loss is noisy ($p = 0.196$).

Next, we look in more detail at the increases and declines in borrowing amount paying attention to the number of Hs the household is linked to at baseline and the degree at baseline. This allows us to concentrate on Ls with no links to Hs and look at the pure externality effect. We estimate

$$\begin{aligned} y_{ivt} = \alpha &+ \beta_1 \text{MF}_v \times \text{No. of } H \text{ links}_{iv} \times \text{Post}_t \\ &+ \gamma_1 \text{MF}_v \times \text{Post}_t + \gamma_2 \text{No. of } H \text{ links}_{iv} \times \text{Post}_t + \gamma_3 \text{MF}_v \times \text{No. of } H \text{ links}_{iv} \\ &+ \delta_1 \text{MF}_v + \delta_2 \text{No. of } H \text{ links}_{iv} + \delta_3 \text{Post}_t \\ &+ \eta_1 \text{MF}_v \times \text{No. of links}_{iv} \times \text{Post}_t + \eta_2 \text{MF}_v \times \text{No. of links}_{iv} + \eta_3 \text{No. of links}_{iv} + \epsilon_{ivt}, \end{aligned}$$

where again y_{ivt} is the amount borrowed from the stated source (MFI, friends, self-help group, family, moneylenders), No. of links is the degree in Wave 1, and No. of H links is the baseline number of H links in Wave 1.

We are particularly interested in $\gamma_1 + \eta_1 \cdot d$, the differential effect of being in a microfinance village in the second period, without any H-type of links at baseline and just d of L-type links at baseline. This computes consists of the pure externalities among people who had nothing to do with microcredit whatsoever as they have no H links. For example, if the household had only one L link and no H links, then $\gamma_1 + \eta_1$ is the differential effect of being in the second period in a village exposed to microfinance.

Tables 9 and 10 present the results. In Table 9 we look at H respondents and in Table 10 we look at L respondents.[9]

[9]We find that both Hs and Ls gain loans from microfinance, but H gain by about 5.5 times more (Rs. 1741 relative to Rs. 320.3), which is reassuring. That L also obtain some microcredit is unsurprising since we expect there to be some classification error, and, of course, also not every H received microcredit. That there are some microfinance loans even in what we call non-microfinance villages is also unsurprising as there

Loans from friends go down for both types (not significantly for H) when exposed to microfinance. Strikingly the fall is greater for L households. For an L household with a single L friend and no H friends the effect is $\gamma_1 + \eta_1$, which corresponds to a decline of Rs. 2572.3 ($p = 0.015$) more than a comparable H household (a decline of Rs. 1194, not significantly different from zero). We can statistically reject that the decline of Rs. 2572 for L is smaller than the Rs. 1194 ($p = 0.094$).

We also find that Hs engage in complementary borrowing from moneylenders. An H household with one L link borrows Rs. 2379 more from moneylenders when exposed to microfinance ($p = 0.04$).

Taken together, the evidence suggests that exposure to microfinance may have an adverse effect on the network borrowing of the Ls which is large enough to dominate their small potential gains from microfinance. This is especially striking because the Ls are much less likely to have access to microfinance so, all else being the same, we would have expected their borrowing from friends to go down less than that of the Hs or even to go up to the extent there is re-lending (from Hs to Ls).

4. A Model

In this section we present a model of network formation that is consistent with what we see in the data. We discuss some of the ingredients of the model that are necessary to get results consistent with the data. As we discussed above, the model has to explain why LL links (and LLL triangles) disappear at the highest rate in microfinance villages compared to non-microfinance villages, even though L's do not have loans.

We present the model for links, and then describe how it can be extended to cover triangles. As the model may be useful beyond the current paper, we describe it in a general form and then specialize to the two-type (H, L) microfinance case.

4.1. **Types and Utilities.**

There are n individuals, indexed by $i, j... \in \{1, \dots, n\}$. Each agent has a type θ_i from a type set Θ. Let $v_{\theta\theta'}$ denote the base benefit that a type θ agent gets from a relationship with an agent of type θ'. This comes from borrowing and lending activities, as we discuss in more detail below.

The realized utility from a relationship also involves an idiosyncratic noise term ε_{ij} that i gets from being friends with j. This could be personality compatibility, or some other benefits. Thus, an agent i gets a value $v_{\theta_i\theta_j} + \varepsilon_{ij}$ from a connection with j, where ε_{ij} is distributed according to an atomless distribution F.

was some entry by other MFIs: what is reassuring is that the numbers for the non-microfinance villages are small, even for H types.

A useful expression is

$$\mathrm{E}^{+}[v] = \mathrm{E}\left[v + \varepsilon_{ij} \,|\varepsilon_{ij} > -v\right] = v + \frac{\int_{-v}^{\infty} \varepsilon_{ij} dF}{\int_{-v}^{\infty} dF},$$

which denotes the expectation of $v + \varepsilon_{ij}$ conditional the value of $v + \varepsilon_{ij}$ being positive. This is the expected utility that an agent gets from a relationship with base value v, conditional upon having being willing to form the friendship.

An agent of type θ's expected utility if they expect to have $d_{\theta\theta'}$ friends of type θ' is then

$$\sum_{\theta' \in \Theta} d_{\theta\theta'} \mathrm{E}^{+}[v_{\theta\theta'}].$$

4.2. **Efforts and Link Formation.** Each agent chooses an effort $e_i \in [0,1]$, which represents the amount of time they spend socializing and maintaining or forming links. In the case of the villagers, this could be time spent in the town square or tea shop, where they meet with other villagers.[10] As will become evident, our model is meant to capture both link formation and link maintenance; and our main focus is on the indirect spillover effects and interdependencies.

Two agents i and j who have chosen efforts e_i and e_j have probability proportional to $e_i e_j$ of meeting. The model therefore rules out "directed search" since the probability of meeting is independent of the agent's type, conditional on their effort. Time goes in periods $t \in \{0, 1, 2, \ldots\}$.

Let $g^t \in \{0,1\}^{n \times n}$ be the adjacency matrix representing network at time t. If $g_{ij}^{t-1} = 1$. We assume that already connected agents keep their friendship if they meet with each other during time t - so that keeping the relationship requires seeing each other over time. Therefore agents i and j keep their friendship with probability $e_i e_j$ and lose it with probability $1 - e_i e_j$.

If $g_{ij}^{t-1} = 0$, then agents i, j form a friendship with probability

$$e_i e_j \left(1 - F(-v_{\theta_i \theta_j})\right) \left(1 - F(-v_{\theta_j \theta_i})\right).$$

This is the probability that they meet *and* they both find the friendship of positive value–a friendship requires mutual consent.

Thus, the effort of agents does two things: maintains old relationships by continuing an interaction, but also allows them to meet new people. Link formation also requires mutual consent conditional upon meeting.

4.3. **Steady-State Equilibrium.** A *steady-state equilibrium* is a list of efforts $(e_\theta)_{\theta \in \Theta}$, and a corresponding set of expected degree levels $(d_{\theta\theta'})_{\theta\theta' \in \Theta^2}$ such that e_θ maximizes each

[10]See Currarini, Jackson, and Pin (2009, 2010); Cabrales, Calvó-Armengol, and Zenou (2011); Canen, Jackson, and Trebbi (2017) for other models where socialization takes effort and there is random meeting.

agent's expected utility, and the expected degree levels are in steady state as generated by the efforts.[11]

We prove in the appendix that in all equilibria all agents of the same type choose the same action, and that the equilibrium, provided that costs of effort are not trivial.

The requirement that degrees are in steady state and generated by the efforts can be formally represented as follows. Let $n_{\theta\theta'}$ denote the number of agents of type θ' with whom an agent of type θ could potentially form friendships. If $\theta' \neq \theta$ then this will generally be the number of agents of type θ',[12] while if it is of type θ then it will less by one to account for the agent herself.

Out of those agents only an expected fraction of $(1 - F(-v_{\theta\theta'}))\,(1 - F(-v_{\theta'\theta}))$ will ever be friends with an agent of type θ, given the mutual consent requirement. Thus, let

$$m_{\theta\theta'} = n_{\theta\theta'}\,(1 - F(-v_{\theta\theta'}))\,(1 - F(-v_{\theta'\theta}))\,.$$

This is the effective size of the pool of agents of type θ' with which an agent of type θ will be friends over time.

Degree at the end or beginning of a period is then the maintained relationships plus the new ones formed:

$$d_{\theta\theta'} = e_\theta e_{\theta'} d_{\theta\theta'} + (m_{\theta\theta'} - d_{\theta\theta'})\,e_{\theta'} e_\theta,$$

which simplifies to

$$d_{\theta\theta'} = m_{\theta\theta'} e_\theta e_{\theta'} = n_{\theta\theta'}\,(1 - F(-v_{\theta\theta'}))\,(1 - F(-v_{\theta'\theta}))\,e_\theta e_{\theta'}.$$

Thus, in steady state, the degree is then proportional to the number of available agents of the other type, weighted by the probability that there is a mutual compatibility, and by the socializing efforts.

The expected utility of an agent involves the benefits from relationships, the costs of socialization, $-\frac{1}{2}c_\theta e_\theta^2$, as well as a base benefit from socializing, $u_\theta e_\theta$. An agent may get some base value from going to the town square or getting tea, etc., independently of who else is there.

Overall this leads to a utility of

$$V_\theta(e_\theta) = \underbrace{u_\theta e_\theta - \frac{1}{2}c_\theta e_\theta^2}_{\text{base benefit and costs of effort}} + \underbrace{\sum_{\theta'\in\Theta} \mathrm{E}^+[v_{\theta\theta'}] d_{\theta\theta'} e_{\theta'} e_\theta}_{\text{expected maintenance of existing friendships by effort}}$$

[11]We solve the model in terms of steady-state and expected values, but it will be clear from the analysis that one can also do this in terms of realized values. The equilibrium will still be unique, the complementarities still apply in the same manner, and the equilibria have the same comparative statics. The complication is that strategies then need to be specified as a function of more than just type, as the realized noise terms then matter. As this adds no insight, we work with the more transparent version of equilibrium.

[12]It could also incorporate some other taboos or restrictions, for instance if some types simply are not permitted to form relationships, which would be captured by the vs.

$$+\underbrace{\sum_{\theta' \in \Theta} \mathrm{E}^{+}[v_{\theta\theta'}]\left(m_{\theta\theta'} - d_{\theta\theta'}\right) e_{\theta'} e_{\theta}}_{\text{expected new friendships from effort}}$$

Using the expressions for $m_{\theta\theta'}$ and $d_{\theta\theta'}$, we can write

$$V_{\theta}\left(e_{\theta}\right) = u_{\theta} e_{\theta} - \frac{1}{2} c_{\theta} e_{\theta}^{2} + \sum_{\theta' \in \Theta} \mathrm{E}^{+}[v_{\theta\theta'}] n_{\theta\theta'} \left(1 - F(-v_{\theta\theta'})\right)\left(1 - F(-v_{\theta'\theta})\right) e_{\theta'} e_{\theta}.$$

If we take $u_{\theta} > 0, c_{\theta} > 0$ for all θ and $\mathrm{E}^{+}[v_{\theta\theta'}] \geq 0$ for all θ, θ', then an equilibrium requires that:[13]

$$e_{\theta} = \min\left\{1,\ \frac{1}{c_{\theta}}\left(u_{\theta} + \sum_{\theta' \in \Theta} \mathrm{E}^{+}[v_{\theta\theta'}] n_{\theta\theta'} \left(1 - F(-v_{\theta\theta'})\right)\left(1 - F(-v_{\theta'\theta})\right) e_{\theta'}\right)\right\}.$$

4.4. **Equilibrium Existence and Some Comparative Statics.** This is a game of strategic complements, and therefore, as is well-known,[14] equilibria exist and form a complete lattice. If $u_{\theta} = 0$ for all θ, then there exists a corner equilibrium in which all agents exert 0 effort. To examine the more interesting case, we presume that $u_{\theta} > 0$ for all agents, so that agents gain some utility from socializing regardless of the connections they form from it. In this case, for high enough costs there exists a unique equilibrium. We also note a fact about the model that helps explain why the model will generate results consistent with our empirical findings. The following result is proven in Appendix A.

PROPOSITION **1.** *Let $u_{\theta} > 0, c_{\theta} > 0$ for all θ. For sufficiently large $c_{\theta} > 0$'s, there is a unique equilibrium and it is stable and interior ($0 < e_{\theta} < 1$ for all θ), and agents of the same type take the same efforts. In addition, if $\mathrm{E}^{+}[v_{\theta\theta'}] > 0, n_{\theta\theta'} > 0$ for each θ, θ',[15] and $v_{\theta\theta'}$ is reduced for some $\theta\theta'$ (holding all other parameters constant), then $e_{\theta''}$ goes down for all θ'', and $d_{\theta''\theta'''}$ goes down for all $\theta''\theta'''$.*

The characterization of equilibrium is as follows. Let u be the $|\Theta|$-dimensional vector with entries $\frac{1}{c_{\theta}} u_{\theta}$ and E be the $|\Theta| \times |\Theta|$ matrix with θ, θ' entries

$$\frac{1}{c_{\theta}} \sum_{\theta' \in \Theta} \mathrm{E}^{+}[v_{\theta\theta'}] n_{\theta\theta'} \left(1 - F(-v_{\theta\theta'})\right)\left(1 - F(-v_{\theta'\theta})\right).$$

[13] These come from the first order conditions, capped by the bound on efforts. Second order conditions are $-c_{\theta}$ and so are negative. Thus, these conditions are also sufficient.

[14] For instance, see Van Zandt and Vives (2007).

[15] All that is needed for this result is that this is true for a cycle of θ and θ's that include all types. Note also that $\mathrm{E}^{+}[v_{\theta\theta'}] > 0$ does not require that all people form links, just that there is a non-zero probability that any two types could find a high enough noise term to form a friendship.

Then the unique equilibrium is given by

$$e = (I - E)^{-1} u,$$

which we show is well-defined for large enough costs in the appendix.

A major implication of the proposition is that a decrease in the returns from any type of relationship decreases *all* efforts and degrees. The decrease in value $v_{\theta\theta'}$ for some $\theta\theta'$ directly affects their efforts, decreasing those. Then, given the strict strategic complementarities, there is then a decrease in other efforts; and the feedback can lead to a substantial drop in all efforts.

As we will see, open sets of parameters can lead to the largest drop to be among the types who were not directly affected (the LL's).

4.5. **Externalities in Network Formation.** Even though our model does not include direct externalities in payoffs between links, the network formation process still exhibits significant external effects since some agents' decisions to form more or fewer links (their effort levels) affect others' potential payoffs and their network formation decisions. Most models of strategic network formation are based on explicit externalities in payoffs – so that the marginal payoff to a given link depends on its indirect connection to other links - so that a person cares about indirect relationships: friends-of-friends are valued either positively or negatively, etc.[16] In those models inefficiencies arise because the broader societal externalities across links are not taken into account by the two people involved in forming a particular link. Here, our analysis implies that even in the complete absence of any externalities based on link configurations, externalities in the network formation process itself can also result in broad strategic interactions and in inefficiencies. We add further direct externalities to the picture in the appendix, but they are not needed to get the main insights here.

This makes a point beyond the current setting: network formation can be inefficient or distorted not simply because of externalities in the payoffs across links, but also because meeting people takes efforts which also result in strategic complementarities and substantial externalities. Here, decreased socialization efforts by the High types makes socialization less appealing for the Low types, which then decreases their efforts as well, consistent with what we see both anecdotally and in terms of the empirical measurements above.

4.6. **Specializing to Microfinance and Structural Estimation.**

4.6.1. *Two-Types.* We now specialize the model to the case of $\Theta = \{H, L\}$. Let λ be the share of H types in the population. In this case, a steady-state is a solution to the equations:

$$c_H e_H^\star = u_H + \mathrm{E}^+[v_{HH}]\,(\lambda n - 1)\,(1 - F(-v_{HH}))^2\, e_H^\star \tag{4.1}$$

[16] For references see Jackson (2003, 2008).

$$+\,\mathrm{E}^{+}[v_{HL}]\,(1-\lambda)\,n\,(1-F(-v_{HL}))\,(1-F(-v_{LH}))\,e^{\star}_{L},$$

$$c_L e^{\star}_{L} = u_L + \mathrm{E}^{+}[v_{LL}]\,((1-\lambda)\,n-1)\,(1-F(-v_{LL}))^{2}\,e^{\star}_{L} \tag{4.2}$$
$$+\,\mathrm{E}^{+}[v_{LH}]\lambda n\,(1-F(-v_{LH}))\,(1-F(-v_{HL}))\,e^{\star}_{H},$$

$$d_{HL} = ((1-\lambda)\,n)\,e^{\star}_{H}e^{\star}_{L}\,(1-F(-v_{HL}))\,(1-F(-v_{LH}))\,, \tag{4.3}$$

$$d_{LH} = d_{HL}\frac{\lambda}{1-\lambda}, \tag{4.4}$$

$$d_{HH} = (\lambda n-1)\,(e^{\star}_{H})^{2}\,(1-F(-v_{HH}))^{2}\,, \tag{4.5}$$

$$d_{LL} = ((1-\lambda)n-1)\,(e^{\star}_{L})^{2}\,(1-F(-v_{LL}))^{2}\,. \tag{4.6}$$

We can then compute $e^{\star}_{H}$ and $e^{\star}_{L}$, $d_{HH}, d_{HL}, d_{LH}, d_{LL}$ as the solution to the above six equations, as a function of the parameters.

4.6.2. *Application to Microfinance: An Example.* Before showing the solutions to the model as the parameters change, it is useful to get some idea of how we might expect microfinance to change the underlying payoffs in the model. To do this, we discuss a natural rationalization for the payoffs from borrowing and lending.

In particular, we can rationalize the values $v_{\theta\theta'}$ from "financial" connections considering payoffs from borrowing and lending. Doing this can give us some idea of how $v_{HH}, v_{HL}, v_{LH}, v_{LL}$ will change in response to H's getting microcredit.

A simple setting has lending giving some net return of r, which would be the effective expected interest rate from informal loans less the opportunity cost of that money. Borrowing leads to a net return from getting a loan after repaying the loan of b which is to be thought of as the difference between the interest rate charged by a network "friend" and the market rate. Generally, we expect $b > 0$ and $b > r$,[17] as otherwise such relationships make little sense. Whether r is positive or negative is not obvious since there are clearly social expectations to help out friends in need (which may make r negative), and may depend on context.

A household can be in one of three states of the world: they have money to lend, they need to borrow, or neither. An H household has a probability α_H of having money to lend and a probability $\beta_H \leq 1-\alpha_H$ of needing to borrow, and with the remaining probability $1-\alpha_H-\beta_H$ neither occurs. There are similar probabilities α_L and β_L for the L types.

[17]The limited evidence we have on peer to peer lending suggests that markups tend to be small, potentially even negative. b by contrast ought to be substantial and positive.

The base payoff to an agent of type $\theta \in \{H, L\}$ of being matched to agent of type $\theta' \in \{H, L\}$ is then

$$v_{\theta\theta'} = \alpha_\theta \beta_{\theta'} r + \beta_\theta \alpha_{\theta'} b.$$

The introduction of microfinance changes these parameters. For instance, if repayments are weekly and the household has made borrowed from the MFI to make a sizable purchase, it may have to cut back on lending smaller sums to others in the village and may even start borrowing small amounts to repay the loans when cash is short, leading to a decline in α_H and perhaps an increase in β_H. In addition if there are complementarities between formal and informal loans because receiving a MF loan allows the household to overcome a non-convexity, β_H could go up. In contrast, if re-lending of formal credit to network partners is common, a type H may have a probability $\alpha'_H \geq \alpha_H$ of being able to lend once he gets access to microfinance. His probability of needing to borrow may also go down to $\beta'_H \leq \beta_H$, if microfinance loans are a substitutes for network credit. In any case, we maintain that the L's needs for borrowing and lending are unaltered by the introduction of microfinance.

Let

$$\Delta\beta_H = \beta'_H - \beta_H \quad \text{and} \quad \Delta\alpha_H = \alpha'_H - \alpha_H$$

be the changes in the probabilities that the H-types have borrowing and lending needs after microfinance. We assume that the corresponding numbers of the L type are unchanged.

Let $\Delta_{\theta\theta'}$ denote the resulting change in $v_{\theta\theta'}$. To get a feeling for how this depends on $\Delta\alpha_H$ and $\Delta\beta_H$, note that for small values of $\Delta\alpha_H$ and $\Delta\beta_H$, we get approximations

$$\begin{aligned} \Delta_{HH} &= (\alpha_H \Delta\beta_H + \beta_H \Delta\alpha_H)(r + b) & \Delta_{LL} &= 0 \\ \Delta_{HL} &= \alpha_L \Delta\beta_H b + \beta_L \Delta\alpha_H r & \Delta_{LH} &= \alpha_L \Delta\beta_H r + \beta_L \Delta\alpha_H b. \end{aligned}$$

Given that $\Delta\beta_H$ and $\Delta\alpha_H$ can each go up or down, it is very hard to say anything very general about the signs of these expressions. However one obvious special case is when both α_H and β_H go down. In this case, as long as both b and r are positive, it is easy to see that all of v_{HL}, v_{LH} and v_{HH} must go down.

However this is not the only possibility: consider the special case in which $\alpha_L = \alpha_H$, $\beta_L = \beta_H$ and $\alpha_H \Delta\beta_H + \beta_H \Delta\alpha_H = 0$. In this case $\Delta_{HH} = 0$. Notice that $b - r$ is likely to be positive. Meanwhile as discussed above we expect $b - r$ to be positive. Then Δ_{HL} should be positive whereas Δ_{LH} should be negative as long as $\Delta\beta_H > 0$ and $\Delta\alpha_H < 0$.[18]

4.7. **Extensions of the model.** We discuss several extensions of the model that were omitted from the above example for clarity.

The model is solved in steady-state. Adding a population of unlinked (say "new born") agents to the population of the unmatched is straightforward, as is having agents exit.

[18]See the calculations in Appendix F.

Our basic model also has no place for triads, which we previously saw to play a key role. This can be added directly, simply by having triples meet if they are all present in the town square. The extension is straightforward and thus omitted (see Chandrasekhar and Jackson (2018) for more detail on such an extension). In such a model, analogous to the pairs case above, it would be direct that LL and LLL decline more than their counterparts (HH and LH; HHH, LHH, and LLH, respectively).

Note that it is plausible that when one aspect of a relationship gets less importance there is some risk that the entire relationship breaks up since there are fixed costs of maintaining a relationship as well as other reasons for complementarities. By adding other types of links that are maintained and formed at the same time as these financial links, we can get similar effects on other links as well. As we saw above in Section 3.4, in Table 4 when we look at advice-based links, the effects are more or less of the same magnitude in proportional terms and in the same direction as the financial links.

4.8. **Alternative Explanations.** In this section we try to address two issues. First, can we account for the facts without going to a model with undirected search while maintaining our assumptions about changes in payoffs? Second, are there alternative assumptions about changes in payoffs that can help us account for the facts in combination with a simpler model of network formation?

4.8.1. *Alternative models of network formation.* In Online Appendix C, we discuss four other models of network formation, variations of which are already in the literature. We show why one needs a model that goes beyond those models to generate the patterns in our data.

A model that does not involve some randomness in meetings - some form of undirected search - would have to involve strange externalities in payoffs to get LL and LLL relationships to drop in response to a decrease in Hs' willingness to link to Ls. This suggests that there needs to be some efforts made that result in meetings - so that the decreased willingness of Hs to link with Ls results in lower social efforts by Ls. The reason that this requires both social efforts and conditional choices to link conditional upon meeting, is that the Hs themselves are not lowering linking as much as Ls - so it is the decreased prospects of linking - not just the efforts - that drive the results.

In summary, explaining the data requires a model that explains why so many LL pairs and LLL triples decrease, both old and new, even though they do not involve any Hs, and should have increased if the Ls could just costlessly meet with each other.

The additional dimensions of link maintenance and multiplexing then directly address other aspects of the data.

4.8.2. *Alternative models of match value.* We have so far assumed that match value depends only types and does not depend the pattern of matching. It is possible, for example, that matches are substitutes so that when a lot LH links break, the value of LL links may go up. This obviously goes in the wrong direction. It is also possible, though less plausible, that links are complements: perhaps when an L can no longer borrow from the Hs, she gives up the entire project and therefore also stops borrowing from the Ls. However in this case the LL links break because some LH links have disappeared and therefore the effect on LL links should be smaller than the effect on LH links in proportional terms.

Another alternative view is that the reason LL links drop is that Ls recognize that even if they don't participate in microfinance, it is available to them. This is probably true for some of them, but because we use microfinance eligibility to determine who is an L, it is less true for them than for the H (who also don't all borrow). An H is therefore more likely to break their link with an L on these grounds than another L.

The second alternative is based on the idea that the very fact that Hs tend to socialize with Hs in microfinance meetings would provide a force unique to participants, hence Hs, to form new links and hence HH links should decline less than other types of links which rules out this story. We examine this in Online Appendix E. We show that our main results hold even if we condition on all pairs where neither member joined microfinance. This, of course, comes from the fact that the vast majority (86%) of existing links have no MF participant on either end.

A third alternative, ex-ante, is simply that the mechanism may be a version of our undirected search model where the H types simply do not have time to meet with the Ls anymore. Notice our general form of the model allows for this, so there is no conflict. This is through heterogeneity on the c_H parameter: allowing for the cost of socializing to be differentially higher by type under treatment, which we do find.

A last possibility is that the entry of microfinance leads to rapid economic growth in the village, so that both H and L type don't need to maintain informal relationships any more. This is not only inconsistent with the extensive literature on microfinance, which finds little impact of microfinance entry on average village or neighborhood level outcomes such as consumption, investment or business profit (Angelucci, Karlan, and Zinman, 2015; Attanasio, Augsburg, De Haas, Fitzsimons, and Harmgart, 2015; Augsburg, De Haas, Harmgart, and Meghir, 2015; Banerjee, Karlan, and Zinman, 2015b; Banerjee, Duflo, Glennerster, and Kinnan, 2015a; Crépon, Devoto, Duflo, and Parienté, 2015; Tarozzi, Desai, and Johnson, 2015) (see Meager (2015) for a meta-analysis), but is also inconsistent with the household finances and borrowing that we see going on in the villages.[19] Moreover, per se, this explanation would also not account for larger effects for the L households.

5. Conclusion

By studying a setting in which microcredit was introduced to a subset of 75 villages, we established that not only did the village social networks change in response, but further, those who experienced the greatest change in their local networks were the most ex-ante unlikely to join microcredit. The data show that, further, these agents faced large losses in links even when those links were not to those who were likely to take-up microcredit. This suggests a new sort of general equilibrium effect in network formation.

To explain the data we introduce a model in which agents put in effort in order to socialize, whom they meet has an undirected component, and agents engage in mutual consent to build links. Such a model embodies a global externality. If exposure to microcredit, for instance, reduces the desire to lend money because these individuals now have to make regular repayments, then the value of a link between a microcredit taker and an individual who does not join declines. Consequently, even those unlikely to join microcredit may reduce their own effort, knowing that microcredit takers would find the link less helpful and they would also find the link with microcredit takers less useful. As such, in equilibrium, those who are unlikely to be involved with microcredit can suffer the greatest losses in links when the village is exposed to microfinance.

The fact that our model provides patterns consistent with the data, of course, does not imply that it is the right or only thing that is behind the empirical observations. It will take further research to develop a full causal understanding of our empirical observations. But the facts, in particular the evidence of negative spillovers on the non-beneficiaries, are robust and have wide-ranging and important implications. The previous literature has shown that there may be important positive spillovers from microfinance on participant households, especially in terms of strengthened network connections. But this could come at a significant cost of weakening connections in the rest of the community.

Regardless of the explanation for the changes, the more general point is that social networks can involve spillovers, externalities, and complex relationships so that changing one

[19]All of these studies, other than Augsburg et al. (2015), estimate the treatment effect at the community (urban neighborhood or village) level; these are therefore directly comparable to the effect on the village networks in our study. Moreover only Crépon et al. (2015) allows for the estimation of spillovers and does not find any support for the view that there are some people who get large benefits, while other lose out. One exception to this evidence of a lack of impact is Breza and Kinnan (2018) who find negative effects of shutting down microcredit in Andhra Pradesh, India. Their interpretation of this result is that there are spillovers, possibly from a demand shortage generated by a large-scale demand crunch. However their unit of comparison is a district, which has many hundreds of villages, and therefore there is a much bigger scope for spillovers. The villages in our paper are relatively far from each other and there is much less chance of cross-village spillovers.

part of the network can have quite extensive and unanticipated consequences. As a result, interventions into a community can change the social structure and interactions in ways that no one intended, with potentially large costs for some non-participants. Being mindful of these possibilities is important in designing more effective policies.

References

Altonji, J. G., T. E. Elder, and C. R. Taber (2005): "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of political economy*, 113, 151–184. 3.2

Ambrus, A., M. Mobius, and A. Szeidl (2014): "Consumption risk-sharing in social networks," *American Economic Review*, 104, 149–82. 1

Angelucci, M., D. Karlan, and J. Zinman (2015): "Microcredit impacts: Evidence from a randomized microcredit program placement experiment by Compartamos Banco," *American Economic Journal: Applied Economics*, 7, 151–82. 4.8.2

Arrow, K. J. (2000): "Observations on social capital," *Social capital: A multifaceted perspective (World Bank Publisher)*, 6, 3–5. 1

Attanasio, O., B. Augsburg, R. De Haas, E. Fitzsimons, and H. Harmgart (2015): "The impacts of microfinance: Evidence from joint-liability lending in Mongolia," *American Economic Journal: Applied Economics*, 7, 90–122. 4.8.2

Augsburg, B., R. De Haas, H. Harmgart, and C. Meghir (2015): "The impacts of microcredit: Evidence from Bosnia and Herzegovina," *American Economic Journal: Applied Economics*, 7, 183–203. 4.8.2, 19

Bala, V. and S. Goyal (2000): "A noncooperative model of network formation," *Econometrica*, 68, 1181–1229. 1, C

Banerjee, A., E. Breza, E. Duflo, and C. Kinnan (2014a): "Does Microfinance Foster Business Growth? The Importance of Enterpreneurial Heterogeneity," Working paper. 1

Banerjee, A., A. Chandrasekhar, E. Duflo, and M. O. Jackson (2013): "Diffusion of Microfinance," *Science*, 341, DOI: 10.1126/science.1236498, July 26 2013. 1, 2, 6, 2.4, B.3, B.4

——— (2014b): "Gossip and Identifying Central Individuals in a Social Network," *working paper*. 1

Banerjee, A., E. Duflo, R. Glennerster, and C. Kinnan (2015a): "The miracle of microfinance? Evidence from a randomized evaluation," *American Economic Journal: Applied Economics*, 7, 22–53. 1, 2.1.2, 4.8.2

Banerjee, A., D. Karlan, and J. Zinman (2015b): "Six randomized evaluations of microcredit: Introduction and further steps," *American Economic Journal: Applied Economics*, 7, 1–21. 4.8.2

Beaman, L. and J. Magruder (2012): "Who gets the job referral? Evidence from a social networks experiment," *American Economic Review*, 102, 3574–93. 1

Belloni, A. and V. Chernozhukov (2009): "Least squares after model selection in high-dimensional sparse models," *MIT Department of Economics Working Paper*. 3.1

Belloni, A., V. Chernozhukov, and C. Hansen (2014a): "High-dimensional methods and inference on structural and treatment effects," *The Journal of Economic Perspectives*, 28, 29–50. 1, 3.1

——— (2014b): "Inference on treatment effects after selection among high-dimensional controls," *The Review of Economic Studies*, 81, 608–650. 1, 3.1

Binzel, C., E. Field, and R. Pande (2013): "Does the Arrival of a Formal Financial Institution Alter Informal Sharing Arrangements? Experimental Evidence from Village India," Working paper. 1

Bloch, F., G. Genicot, and D. Ray (2008): "Informal insurance in social networks," *Journal of Economic Theory*, 143, 36–58. 1, C

Blumenstock, J. and X. Tan (2016): "Social Networks and Migration: Theory and Evidence from Rwanda," . 1

Blumenstock, J. E., N. Eagle, and M. Fafchamps (2016): "Airtime transfers and mobile communications: Evidence in the aftermath of natural disasters," *Journal of Development Economics*, 120, 157–181. 1

Boucher, V. (2015): "Structural homophily," *International Economic Review*, 56, 235–264. 1, C

Breza, E. (2016): "Field experiments, social networks, and development," *The Oxford handbook of the economics of networks*. 1

Breza, E. and C. Kinnan (2018): "Measuring the equilibrium impacts of credit: Evidence from the Indian microfinance crisis," Tech. rep., National Bureau of Economic Research. 19

Cabrales, A., A. Calvó-Armengol, and Y. Zenou (2011): "Social interactions and spillovers," *Games and Economic Behavior*, 72, 339–360. 1, 10

Canen, N., M. O. Jackson, and F. Trebbi (2017): "Endogenous Networks and Legislative Activity," *SSRN http://ssrn.com/abstract=2823338*. 1, 10

Chandrasekhar, A. and M. O. Jackson (2018): "A Network Formation Model Based on Subgraphs," *SSRN Working Paper: http://ssrn.com/abstract=2660381*. C.5

Christakis, N., J. Fowler, G. Imbens, and K. Kalyanaraman (2010a): "An Empirical Model for Strategic Network Formation," *NBER Working Paper*. 1

Christakis, N. A., J. H. Fowler, G. W. Imbens, and K. Kalyanaraman (2010b): "An empirical model for strategic network formation," Tech. rep., National Bureau of Economic Research. C

Comola, M. and S. Prina (2014): "Do Interventions Change the Network? A Dynamic Peer Effect Model Accounting for Network Changes," Working paper. 1

Crépon, B., F. Devoto, E. Duflo, and W. Parienté (2015): "Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in Morocco," *American Economic Journal: Applied Economics*, 7, 123–50. 4.8.2, 19

Currarini, S., M. O. Jackson, and P. Pin (2009): "An economic model of friendship: Homophily, minorities, and segregation," *Econometrica*, 77, 1003–1045. 1, 10

——— (2010): "Identifying the roles of race-based choice and chance in high school friendship network formation," *Proceedings of the National Academy of Sciences*, 107, 4857–4861. 1, 10

Currarini, S. and M. Morelli (2000): "Network formation with sequential demands," *Review of Economic Design*, 5, 229 – 250. 1, C

Dutta, B. and S. Mutuswami (1997): "Stable networks," *Journal of economic theory*, 76, 322–344. 1, C

Fafchamps, M. and S. Lund (2003): "Risk-sharing networks in rural Philippines," *Journal of development Economics*, 71, 261–287. 1

Feigenberg, B., E. Field, and R. Pande (2013): "The economic returns to social interaction: Experimental evidence from microfinance," *The Review of Economic Studies*, rdt016. 1, C.3.4, E

Field, E., R. Pande, J. Papp, and Y. J. Park (2012): "Repayment flexibility can reduce financial stress: a randomized control trial with microfinance clients in India," *PloS one*, 7, e45679. 1

Herings, P. J.-J., A. Mauleon, and V. Vannetelbosch (2009): "Farsightedly stable networks," *Games and Economic Behavior*, 67, 526–541. 1, C

Hess, S. H., D. Jaimovich, and M. Schündeln (2018): "Development Projects and Economic Networks: Lessons From Rural Gambia," . 1

Jackson, M. O. (2003): "The stability and efficiency of economic and social networks," *Advances in Economic Design, Heidelberg: Springer-Verlag, edited by Koray, S. and Sertel, M.* 16

——— (2008): *Social and economic networks*, Princeton: Princeton University Press. 5, 16

Jackson, M. O., T. R. Rodriguez-Barraquer, and X. Tan (2012): "Social Capital and Social Quilts: Network Patterns of Favor Exchange," *American Economic Review*, 102, 1857–1897. 1, 3.3, C, C.3

Jackson, M. O. and A. Van den Nouweland (2005): "Strongly stable networks," *Games and Economic Behavior*, 51, 420–444. 1, C

Jackson, M. O. and A. Watts (2002): "The evolution of social and economic networks," *Journal of Economic Theory*, 106(2), 265–295. 1, C

Jackson, M. O. and A. Wolinsky (1996): "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71, 44–74. 1, C

Karlan, D., M. Mobius, T. Rosenblat, and A. Szeidl (2009): "Trust and Social Collateral," *The Quarterly Journal of Economics*, 24, 1307–1361. 1

König, M. D., C. J. Tessone, and Y. Zenou (2014): "Nestedness in networks: A theoretical model and some applications," *Theoretical Economics*, 9, 695–752. 1

Meager, R. (2015): "Understanding the impact of microcredit expansions: A bayesian hierarchical analysis of 7 randomised experiments," *arXiv preprint arXiv:1506.06669*. 4.8.2

Mele, A. (2017): "A structural model of Dense Network Formation," *Econometrica*, 85, 825–850. 1, C

Munshi, K. and M. Rosenzweig (2016): "Networks and misallocation: Insurance, migration, and the rural-urban wage gap," *American Economic Review*, 106, 46–98. 1

Putnam, R. (2000): *Bowling Alone: The Collapse and Revival of American Community*, New York: Simon and Schuster. 1

Tarozzi, A., J. Desai, and K. Johnson (2015): "The impacts of microcredit: Evidence from Ethiopia," *American Economic Journal: Applied Economics*, 7, 54–89. 4.8.2

Udry, C. (1994): "Risk and insurance in a rural credit market: An empirical investigation in northern Nigeria," *The Review of Economic Studies*, 61, 495–526. 1

Van Zandt, T. and X. Vives (2007): "Monotone equilibria in Bayesian games of strategic complementarities," *Journal of Economic Theory*, 134, 339–360. 14, A

Watts, A. (2001): "A dynamic model of network formation," *Games and Economic Behavior*, 34, 331–341. 1

Tables

Table 1. Sample Statistics

	Microfinance villages	Non-Microfinance villages
Panel A: Wave 1 Data		
Average Degree (Mean)	14.827	13.335
Average Degree (Std. Dev.)	2.5577	2.4427
Average Clustering (Mean)	0.2799	0.3121
Average Clustering (Std. Dev.)	0.0723	0.0897
Average Closeness (Mean)	0.3655	0.3783
Average Closeness (Std. Dev.)	0.0367	0.0464
Number of Households (Mean)	223.21	165.81
Number of Households (Std. Dev.)	56.17	48.94
Panel B: Wave 2 Data		
Average Degree (Mean)	17.3168	17.4601
Average Degree (Std. Dev.)	3.3268	4.3358
Average Clustering (Mean)	0.2872	0.3244
Average Clustering (Std. Dev.)	0.0477	0.0623
Average Closeness (Mean)	0.4483	0.4838
Average Closeness (Std. Dev.)	0.0375	0.0460
Number of Households (Mean)	246.5349	175.8438
Number of Households (Std. Dev.)	67.0797	53.4911

Notes: This table presents summary statistics from the 75 villages in our sample, 43 of them are villages that received microfinance.

Table 2. Characteristics of H versus L

Panel A: Demographics and Amenities variables

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	GMOBC	Latrine	Num. Rooms	Num. Beds	Thatched Roof	RCC Roof
H	-0.160	-0.0375	-0.0850	-0.0394	0.00384	-0.0154
	(0.0285)	(0.0157)	(0.0380)	(0.0342)	(0.00411)	(0.00763)
	[0.000]	[0.019]	[0.028]	[0.253]	[0.353]	[0.047]
Observations	14,904	14,904	14,904	14,904	14,904	14,904
Depvar mean	0.629	0.266	2.358	0.840	0.0235	0.117

Panel B: Network variables

	(1)	(2)	(3)	(4)
VARIABLES	Degree	Links to L	Links to H	Eig. Cent.
H	0.935	-0.911	1.846	0.00963
	(0.229)	(0.231)	(0.197)	(0.00176)
	[0.000]	[0.000]	[0.000]	[0.000]
Observations	14,904	14,904	14,904	14,904
Depvar mean	8.972	5.456	3.516	0.0524

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets.

Table 3. Difference in Differences

VARIABLES	(1) Density	(2) Density	(3) Density	(4) Clustering	(5) Clustering	(6) Clustering	(7) Closeness	(8) Closeness	(9) Closeness
Microfinance x Post	-0.0119	-0.0128	-0.0128	0.00357	0.00968	0.00968	-0.0225	-0.0153	-0.0155
	(0.00678)	(0.00690)	(0.00716)	(0.0146)	(0.0147)	(0.0153)	(0.00970)	(0.0117)	(0.0122)
	[0.0836]	[0.0669]	[0.0769]	[0.807]	[0.513]	[0.528]	[0.0234]	[0.193]	[0.208]
Microfinance	-0.0205	0.00477	0.00204	-0.0408	-0.0179	-0.00638	-0.0129	0.00947	0.00963
	(0.00842)	(0.00555)	(0.00227)	(0.0159)	(0.0148)	(0.00551)	(0.00993)	(0.0101)	(0.0106)
	[0.0175]	[0.393]	[0.373]	[0.0123]	[0.230]	[0.250]	[0.199]	[0.353]	[0.366]
Post	-0.0117	-0.0145	-0.0145	-0.00913	0.00852	0.00852	0.105	-0.0472	-0.00499
	(0.00576)	(0.0107)	(0.0111)	(0.0100)	(0.0249)	(0.0258)	(0.00762)	(0.0522)	(0.0778)
	[0.0454]	[0.182]	[0.198]	[0.366]	[0.733]	[0.742]	[0]	[0.369]	[0.949]
Observations	150	150	150	150	150	150	150	150	150
Double-Post LASSO		✓	✓		✓	✓		✓	✓
Village FE			✓			✓			✓
Depvar Mean	0.0983	0.0983	0.0983	0.307	0.307	0.307	0.418	0.418	0.418

Notes: Standard errors (clustered at the village level) are reported in parentheses. *p*-values are reported in brackets. Controls consist of the share of upper caste households, number of households in the village, share of households in self-help groups, share Hindu, share with a latrine in the house, share that own the household, share that have electricity and share that are leaders.

Table 4. Link Evolution for Info and Financial Links

VARIABLES	(1) Info Linked Post-MF	(2) Info Linked Post-MF	(3) Financial Linked Post-MF	(4) Financial Linked Post-MF
Microfinance × LL	-0.052	-0.001	-0.046	-0.003
	(0.024)	(0.005)	(0.027)	(0.005)
	[0.033]	[0.793]	[0.092]	[0.563]
Microfinance × LH	-0.051	-0.003	-0.025	-0.005
	(0.019)	(0.003)	(0.022)	(0.003)
	[0.009]	[0.390]	[0.260]	[0.115]
Microfinance	-0.002	-0.011	-0.022	-0.008
	(0.021)	(0.005)	(0.025)	(0.005)
	[0.941]	[0.040]	[0.396]	[0.089]
LL	0.017	-0.009	0.018	-0.006
	(0.019)	(0.005)	(0.022)	(0.004)
	[0.371]	[0.070]	[0.430]	[0.133]
LH	0.005	-0.007	-0.012	-0.004
	(0.016)	(0.003)	(0.019)	(0.003)
	[0.770]	[0.029]	[0.526]	[0.104]
Observations	37,044	866,893	27,072	876,865
Linked Pre-MF	Yes	No	Yes	Yes
Depvar Mean	0.326	0.038	0.333	0.034
HH, Non-MF Mean	0.346	0.053	0.364	0.047
MF + MF × LL = 0 p-val	0.007	0.020	0.004	0.044
MF + MF × LH = 0 p-val	0.001	0.001	0.008	0.001
MF × LL = MF × LH p-val	0.923	0.589	0.279	0.483

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. All columns include a full set of controls. Centrality controls are a vector of flexible controls (a polynomial) for centrality of both nodes. Household characteristics are caste and a number of wealth proxies including number of rooms, number of beds, electrification, latrine presence, and roofing material. Household predictor variables consist of all variables that are used in the random forest classification. In every case we include interactions of all of these network, demographic, and classification variables with microfinance.

Table 5. Link Evolution

VARIABLES	(1) Linked Post-MF	(2) Linked Post-MF	(3) Linked Post-MF	(4) Linked Post-MF	(5) Linked Post-MF	(6) Linked Post-MF
Microfinance × LL	-0.064	-0.060	-0.067	-0.004	-0.006	-0.005
	(0.022)	(0.022)	(0.022)	(0.008)	(0.007)	(0.006)
	[0.006]	[0.009]	[0.003]	[0.605]	[0.437]	[0.428]
Microfinance × LH	-0.069	-0.067	-0.072	-0.005	-0.006	-0.006
	(0.017)	(0.017)	(0.017)	(0.005)	(0.005)	(0.005)
	[0.000]	[0.000]	[0.000]	[0.349]	[0.223]	[0.226]
Microfinance	0.002	-0.016	-0.010	-0.017	-0.012	0.000
	(0.021)	(0.023)	(0.110)	(0.009)	(0.008)	(0.026)
	[0.928]	[0.469]	[0.927]	[0.051]	[0.126]	[0.988]
LL	0.021	0.012	0.019	-0.014	-0.010	-0.005
	(0.016)	(0.017)	(0.016)	(0.007)	(0.006)	(0.006)
	[0.205]	[0.470]	[0.256]	[0.050]	[0.127]	[0.421]
LH	0.014	0.011	0.016	-0.011	-0.008	-0.006
	(0.013)	(0.013)	(0.013)	(0.005)	(0.004)	(0.004)
	[0.255]	[0.386]	[0.219]	[0.025]	[0.064]	[0.163]
Observations	57,376	57,376	57,376	846,561	846,561	846,561
Linked Pre-MF	Yes	Yes	Yes	No	No	No
Centrality control		✓	✓		✓	✓
Depvar Mean	0.441	0.441	0.441	0.064	0.064	0.064
HH, Non-MF Mean	0.462	0.462	0.462	0.089	0.089	0.089
MF + MF × LL = 0 p-val	0.001			0.012		
MF + MF × LH = 0 p-val	0.000			0.001		
MF × LL = MF × LH p-val	0.731	0.673	0.759	0.810	0.870	0.920
Household type predictor control			✓			✓
Household characteristics control			✓			✓

Notes: Standard errors (clustered at the village level) are reported in parentheses. *p*-values are reported in brackets. Centrality controls are a vector of flexible controls (a polynomial) for centrality of both nodes. Household characteristics are caste and a number of wealth proxies including number of rooms, number of beds, electrification, latrine presence, and roofing material. Household predictor variables consist of all variables that are used in the random forest classification. In every case we include interactions of all of these network, demographic, and classification variables with microfinance.

Table 6. Evolution of LL links as a function of support from H links

VARIABLES	(1) Linked Post-MF
Num High Friends in Common / Num Friends in Common x MF	0.0491
	(0.0333)
	[0.145]
Num High Friends in Common / Num Friends in Common	-0.0164
	(0.0253)
	[0.519]
Microfinance	-0.0514
	(0.0271)
	[0.0621]
Num Friends in Common x MF	-0.00802
	(0.00509)
	[0.119]
Num Friends in Common	0.0131
	(0.00414)
	[0.00219]
Observations	19,220
Linked Pre-MF	Yes
LL links only	Yes
Depvar Mean	0.459

Notes: Regression includes fixed effects for number of friends in common and interaction of these dummies with MF. Number of High Friends in Common has a mean of 0.49 with a standard deviation of 1.03. Number of High Friends in Common / Number of Friends in Common has a mean of 0.61 with a standard deviation of 0.42. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets.

Table 7. Triples Evolution

VARIABLES	(1) Full triangle linked Post-MF	(2) Full triangle linked Post-MF	(3) Full triangle linked Post-MF	(4) Any link in triangle survived Post-MF	(5) Any link in triangle survived Post-MF	(6) Any link in triangle survived Post-MF
Microfinance × LLL	-0.103	-0.102	-0.106	-0.103	-0.108	-0.105
	(0.0330)	(0.033)	(0.032)	(0.030)	(0.030)	(0.028)
	[0.00254]	[0.003]	[0.001]	[0.001]	[0.001]	[0.000]
Microfinance × LLH	-0.0846	-0.087	-0.094	-0.061	-0.065	-0.067
	(0.0276)	(0.029)	(0.027)	(0.025)	(0.026)	(0.025)
	[0.00305]	[0.003]	[0.001]	[0.016]	[0.015]	[0.009]
Microfinance × LHH	-0.0569	-0.059	-0.063	-0.043	-0.044	-0.046
	(0.0191)	(0.019)	(0.018)	(0.019)	(0.019)	(0.018)
	[0.00393]	[0.003]	[0.001]	[0.026]	[0.021]	[0.014]
Microfinance	0.0294	0.009	0.032	0.011	0.001	0.090
	(0.0261)	(0.026)	(0.148)	(0.021)	(0.021)	(0.128)
	[0.265]	[0.716]	[0.828]	[0.615]	[0.981]	[0.481]
LLL	0.0260	0.018	0.026	0.031	0.029	0.034
	(0.0258)	(0.027)	(0.025)	(0.020)	(0.021)	(0.020)
	[0.317]	[0.513]	[0.302]	[0.125]	[0.169]	[0.088]
LLH	0.0106	0.009	0.018	0.002	0.003	0.010
	(0.0203)	(0.020)	(0.018)	(0.018)	(0.018)	(0.017)
	[0.604]	[0.643]	[0.313]	[0.893]	[0.855]	[0.543]
LLL	0.00599	0.008	0.014	0.004	0.006	0.010
	(0.0160)	(0.015)	(0.014)	(0.014)	(0.014)	(0.013)
	[0.709]	[0.578]	[0.318]	[0.781]	[0.667]	[0.431]
Observations	53,233	53,233	53,233	53,233	53,233	53,233
Linked Pre-MF	Yes	Yes	Yes	Yes	Yes	Yes
Centrality control		✓	✓		✓	✓
Depvar Mean	0.197	0.197	0.197	0.808	0.808	0.808
HHH, Non-MF Mean	0.216	0.216	0.216	0.834	0.834	0.834
MF + MF × LLL = 0 p-val	0.013			0.000		
MF + MF × LLH = 0 p-val	0.018			0.003		
MF + MF × LHH = 0 p-val	0.261			0.101		
MF × LLL = MF × LLH p-val	0.360	0.468	0.540	0.013	0.007	0.009
MF × LLL = MF × LHH p-val	0.104	0.127	0.122	0.020	0.010	0.010
MF × LLH = MF × LHH p-val	0.114	0.119	0.082	0.270	0.209	0.200
Household type predictor control			✓			✓
Household characteristics control			✓			✓

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. Centrality controls are a vector of flexible controls (a polynomial) for centrality all nodes. Household characteristics are caste and a number of wealth proxies including number of rooms, number of beds, electrification, latrine presence, and roofing material. Household predictor variables consist of all variables that are used in the random forest classification. In every case we include interactions of all of these network, demographic, and classification variables with microfinance.

Table 8. Borrowing patterns

VARIABLES	(1) MFI	(2) Friend	(3) SHG	(4) Family	(5) Moneylender	(6) Full Network
MF × Post × H	1,420.789	626.636	411.933	-71.899	129.978	976.332
	(219.156)	(480.269)	(523.082)	(806.320)	(1,256.401)	(1,175.546)
	[0.000]	[0.196]	[0.433]	[0.929]	[0.918]	[0.409]
Microfinance × Post	320.280	-1,046.038	-1,016.528	109.471	-855.003	-1,943.946
	(232.040)	(513.993)	(471.430)	(740.379)	(1,150.991)	(1,166.455)
	[0.172]	[0.045]	[0.034]	[0.883]	[0.460]	[0.100]
Post × H	104.786	-1,069.177	-283.731	-382.132	-997.410	-1,769.285
	(278.031)	(407.767)	(298.380)	(707.445)	(1,102.186)	(965.261)
	[0.707]	[0.011]	[0.345]	[0.591]	[0.368]	[0.071]
Microfinance	260.573	-105.892	8.255	190.591	-40.219	96.014
	(197.514)	(37.933)	(241.509)	(463.169)	(402.089)	(514.685)
	[0.191]	[0.007]	[0.973]	[0.682]	[0.921]	[0.853]
High	-41.558	123.278	271.808	59.082	159.143	472.771
	(262.973)	(38.215)	(130.453)	(564.628)	(246.315)	(587.881)
	[0.875]	[0.002]	[0.041]	[0.917]	[0.520]	[0.424]
Post	218.554	2,795.997	2,084.137	1,434.949	3,380.911	6,365.265
	(175.949)	(449.740)	(382.122)	(555.236)	(1,001.433)	(1,018.074)
	[0.218]	[0.000]	[0.000]	[0.012]	[0.001]	[0.000]
Observations	27,981	27,113	27,981	27,981	27,981	27,113
Depvar Mean	738	1104	1916	1809	2876	4927

Notes: This table presents the effect of microfinance access on the loan amounts borrowed from microfinance institutions, friends, family, banks and moneylenders. All columns control for surveyed in wave 1 fixed effects. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets.

Table 9. Borrowing patterns: H Respondents

VARIABLES	(1) MFI	(2) Friend	(3) SHG	(4) Family	(5) Moneylender	(6) Full Network
MF × No. of Hs × Post	133.033	-28.868	123.798	-582.572	-242.789	-488.035
	(191.001)	(65.215)	(229.319)	(602.388)	(361.192)	(676.775)
	[0.488]	[0.659]	[0.591]	[0.337]	[0.504]	[0.473]
Microfinance × Post	3,708.154	-1,297.353	-49.368	587.360	2,596.432	-628.933
	(1,647.354)	(1,049.217)	(590.078)	(2,461.298)	(1,346.604)	(2,967.033)
	[0.027]	[0.220]	[0.934]	[0.812]	[0.058]	[0.833]
No. of Hs × Post	98.437	-53.927	-146.951	556.989	272.774	350.736
	(78.440)	(55.132)	(173.447)	(375.197)	(246.089)	(411.292)
	[0.213]	[0.331]	[0.400]	[0.142]	[0.271]	[0.397]
MF × No. of Hs	-124.943	5.143	132.866	72.609	196.673	211.851
	(177.287)	(21.623)	(177.805)	(434.277)	(260.920)	(454.866)
	[0.483]	[0.813]	[0.457]	[0.868]	[0.453]	[0.643]
Microfinance	-1,592.624	-41.771	-39.990	-1,360.441	53.111	-1,564.379
	(1,574.730)	(59.574)	(287.417)	(1,958.299)	(456.758)	(2,123.553)
	[0.315]	[0.485]	[0.890]	[0.489]	[0.908]	[0.464]
No. of Hs	-38.203	17.920	-38.355	-482.349	-124.387	-500.842
	(41.261)	(18.866)	(156.666)	(334.017)	(214.914)	(356.254)
	[0.358]	[0.345]	[0.807]	[0.153]	[0.564]	[0.164]
Post	116.230	3,868.362	2,683.170	2,499.233	769.021	9,367.234
	(139.538)	(923.906)	(447.037)	(798.056)	(1,131.388)	(1,396.839)
	[0.408]	[0.000]	[0.000]	[0.002]	[0.499]	[0.000]
Degree × Post	-34.395	-178.767	-16.678	-412.152	52.033	-626.235
	(47.965)	(65.540)	(111.923)	(250.327)	(156.579)	(275.900)
	[0.476]	[0.008]	[0.882]	[0.104]	[0.741]	[0.026]
Degree	47.970	14.599	139.209	428.998	196.159	589.871
	(41.763)	(10.011)	(97.802)	(235.069)	(153.415)	(248.361)
	[0.254]	[0.149]	[0.159]	[0.072]	[0.205]	[0.020]
MF × Degree	224.935	-9.865	-76.414	113.433	-61.100	33.738
	(259.033)	(9.819)	(104.718)	(427.507)	(156.876)	(440.949)
	[0.388]	[0.318]	[0.468]	[0.791]	[0.698]	[0.939]
MF × Degree × Post	-239.588	102.703	-104.236	220.835	-249.478	212.586
	(259.127)	(69.544)	(133.063)	(510.511)	(204.374)	(542.615)
	[0.358]	[0.144]	[0.436]	[0.667]	[0.226]	[0.696]
Observations	10,354	10,033	10,354	10,354	10,354	10,033
Depvar Mean	1013	944.1	2182	1800	2840	5027

Notes: This table presents the effect of microfinance access, no. of H neighbours and their interactions on the loan amounts borrowed from microfinance institutions, friends, family, banks and moneylenders. All columns control for surveyed in wave 1 fixed effects. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets.

TABLE 10. Borrowing patterns: L Respondents

VARIABLES	(1) MFI	(2) Friend	(3) SHG	(4) Family	(5) Moneylender	(6) Full Network
MF × No. of Hs × Post	-353.476	111.771	54.494	-343.383	833.626	-171.527
	(147.379)	(111.260)	(173.335)	(392.103)	(718.622)	(449.429)
	[0.019]	[0.318]	[0.754]	[0.384]	[0.250]	[0.704]
Microfinance × Post	329.527	-2,710.266	-449.116	-183.203	-724.114	-3,331.248
	(209.283)	(1,091.349)	(617.565)	(1,002.670)	(1,277.177)	(1,773.786)
	[0.120]	[0.015]	[0.469]	[0.856]	[0.572]	[0.064]
No. of Hs × Post	115.244	-60.975	-6.440	-441.698	26.892	-511.861
	(124.383)	(101.443)	(144.049)	(258.775)	(542.616)	(353.578)
	[0.357]	[0.550]	[0.964]	[0.092]	[0.961]	[0.152]
MF × No. of Hs	222.650	-44.243	87.780	161.409	-270.460	203.811
	(135.074)	(23.751)	(146.163)	(270.452)	(489.541)	(310.086)
	[0.104]	[0.066]	[0.550]	[0.552]	[0.582]	[0.513]
Microfinance	152.889	35.291	-208.213	-267.788	-107.078	-469.690
	(144.766)	(47.984)	(209.995)	(636.178)	(774.163)	(736.708)
	[0.294]	[0.464]	[0.325]	[0.675]	[0.890]	[0.526]
No. of Hs	-139.465	43.646	-26.134	288.818	-198.403	306.826
	(115.774)	(21.680)	(103.979)	(209.203)	(318.732)	(209.164)
	[0.232]	[0.048]	[0.802]	[0.172]	[0.536]	[0.147]
Post	424.048	5,725.117	2,392.551	1,786.294	4,098.845	10,111.798
	(142.278)	(1,000.090)	(455.338)	(870.131)	(832.852)	(1,593.008)
	[0.004]	[0.000]	[0.000]	[0.044]	[0.000]	[0.000]
Degree × Post	-89.724	-300.463	-27.107	199.200	-131.045	-142.099
	(100.806)	(72.555)	(105.741)	(148.992)	(356.087)	(207.299)
	[0.376]	[0.000]	[0.798]	[0.185]	[0.714]	[0.495]
Degree	89.425	13.037	112.368	-68.810	321.879	58.285
	(94.550)	(10.972)	(66.172)	(146.201)	(243.635)	(158.682)
	[0.347]	[0.239]	[0.094]	[0.639]	[0.191]	[0.714]
MF × Degree	-126.566	8.271	-41.081	-134.348	152.640	-165.004
	(103.176)	(12.923)	(90.629)	(167.947)	(431.562)	(201.324)
	[0.224]	[0.524]	[0.652]	[0.426]	[0.725]	[0.415]
MF × Degree × Post	233.243	138.041	-99.269	352.025	-545.820	389.299
	(113.836)	(86.535)	(135.307)	(270.164)	(543.288)	(297.665)
	[0.044]	[0.115]	[0.465]	[0.197]	[0.318]	[0.195]
Observations	17,627	17,080	17,627	17,627	17,627	17,080
Depvar Mean	576.7	1198	1759	1815	2898	4867

Notes: This table presents the effect of microfinance access, no. of H neighbours and their interactions on the loan amounts borrowed from microfinance institutions, friends, family, banks and moneylenders. All columns control for surveyed in wave 1 fixed effects. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets.

Appendix A. Proof of Proposition 1

We show there is a unique equilibrium and characterize it, here letting each agent's utility be fully dependent upon their label i.

From our discussion above, it follows directly that a best response must satisfy[20]

$$e_i = \min\left\{1,\ \frac{1}{c_i}\left(u_i + \sum_{j\neq i} \mathrm{E}^{+}[v_{ij}]\left(1 - F(-v_{ij})\right)\left(1 - F(-v_{ij})\right) e_j\right)\right\}.$$

Given the bound that $e_j \leq 1$, and the fact that $u_i > 0$, it follows that for sufficiently large c_i,

$$e_i = \frac{1}{c_i}\left(u_i + \sum_{j\neq i} \mathrm{E}^{+}[v_{ij}]\left(1 - F(-v_{ij})\right)\left(1 - F(-v_{ij})\right) e_j\right),$$

and is strictly between 0 and 1.

Thus, taking c_i to be sufficiently large for each i, we let u be the n-dimensional vector with entries $\frac{1}{c_i}u_i$ and E be the $n\times n$ matrix with ij entries $\frac{1}{c_i}\sum_{j\neq i} \mathrm{E}^{+}[v_{\theta\theta'}]\left(1 - F(-v_{ij})\right)\left(1 - F(-v_{ij})\right)$. Then the characterization of equilibria can be written as

$$e = u + Ee,$$

which has a unique solution of $e = (I - E)^{-1}u$, given that E has nonnegative values that are less than 1 and so $(I - E)$ is invertible.

Note that two agents of the same type take the same effort by the symmetry of the expected utility in type and uniqueness of equilibrium overall.

Rewriting u be the $|\Theta|$-dimensional vector with entries $\frac{1}{c_\theta}u_\theta$ and E be the $|\Theta|\times|\Theta|$ matrix with θ, θ' entries $\frac{1}{c_\theta}\sum_{\theta'\in\Theta} \mathrm{E}^{+}[v_{\theta\theta'}]n_{\theta\theta'}\left(1 - F(-v_{\theta\theta'})\right)\left(1 - F(-v_{\theta'\theta})\right)$, the unique equilibrium is given by

$$e = (I - E)^{-1}u.$$

The result on the comparative statics follows from Proposition 16 in Van Zandt and Vives (2007), noting the strict monotonicity of the best responses in the payoffs and actions of others, and the interiority of the equilibrium.

[20]This drops the $n_{\theta\theta'}$ terms, but one can include an indicator n_{ij} and nothing in the argument below changes.

Appendix B. Random Forest Model Description

We use a random forest algorithm to classify our respondents into two types: those that have a high probability of taking up microfinance loans (H) and those that have a low probability (L), when offered.

B.1. Algorithm Inputs.

Input Data:

- N = Set of respondents from all villages,
- N_{mf} = Set of respondents from microfinance villages,
- Y_i = Loan take-up binary outcome for each $i \in N_{mf}$,
- X_i = Set of predictor variables for each $i \in N_{mf}$.

Algorithm Parameters:

- T = Set of trees to grow,
- p = Total number of predictors,
- m = Number of predictors selected at each split,
- C = Misclassification cost matrix to penalize False Positives and False Negatives (explained in subsection B.3),
- t = Fraction of sample to be used as training dataset,
- d = Number of splits/ Number of nodes/ Tree depth.

B.2. Basic Algorithm.

Step 1: Randomly select (with replacement) training data S and testing data S' from N_{mf}. The size of S will be $t \cdot n(N_{mf})$ and the size of S' will be $(1-t) \cdot n(N_{mf})$.

Step 2: For each tree $t \in T$,
- Randomly select (without replacement) a sample of size $n(S)$ from S.
- At each node n of the tree t, randomly select (with replacement) a set of predictors of size m from p.
- At each node until d, construct a split based on the "fitctree" rule in MATLAB which uses Gini's Diversity Index (gdi) to determine the split.
- For every tree t, each $i \in N_{mf}$ will be assigned a classification $\hat{Y}_{it} \in \{0, 1\}$.

Step 3: After classifying each $i \in N_{mf}$, for each tree t, the final classification can be computed as follows,

$$\hat{Y}_i = 1\left\{\frac{1}{n(S)} \sum_{t=1}^{n(T)} \hat{Y}_{it} > 0.5\right\}$$

and therefore $\theta_i = \hat{Y}_i \cdot H + (1 - \hat{Y}_i) \cdot L$.

B.3. Our Parameter Choices.

- T: We use 500 trees.

- p: We use 5 predictors and the choice of predictors is explained in subsection B.4.
- m: We select 3 out of 5 predictors at every split. Conventionally, a third of the total number of predictions are used but we choose to use two-thirds because our selected set of predictors p are already trimmed to those that, based on the rules of the bank and Banerjee et al. (2013), are more likely to be approached for loan takeup.
- C: We specify a ratio of False Positives to False Negatives as 1:12. The alogrithm oversamples unique observations from the class that has a large penalty. In our data from microfinance villages, the take-up rate of loans is 18%. Therefore, we chose to penalize missclassifying an L as an H more heavily.[21]
- t: We use 0.75 of our sample to train the data.
- d: We use the default number of splits i.e., $n(S) - 1$.

B.4. **Selection of predictors.** To select our set of predictors, we use the eligbility rules of the microfinance firm as well as network and social distance to leaders in the village, which should predict how likely it is that a household heard of microfinance. The predictors we use are as follows:

- dummy for being a BSS leader, who are the people that the MFI would approach when entering a village,
- dummy for if the household has a female of eligible age for a microfinance loan, which is a requirement for the household to be able to participate,
- the average closeness (mean of inverse of network distance) to leaders, which is relevant because as in Banerjee et al. (2013) those who are closer to leaders should be more likely to hear of microfinance,
- the average closeness (mean of inverse distance) to same-caste leaders, because interactions within-caste are more likely and therefore should influence the likelihood of being informed,
- the share of same-caste leaders in the village, as above.

Table 2 shows the characteristics on which the H respondents differ from the L respondents. We see that H households are much more likely to be SCST, have smaller houses in terms of room count, much less likely to have a latrine in the household, and much less likely to have an RCC roof, all of which suggests that they tend to be poorer. Finally, we see that H households and L households have comparable degree (H types have 0.94 more friends on a base of 8.97), but the composition exhibits considerable homophily: H types have a lower number of links to L types and a higher number of links to H types. But H households are more eigenvector central in the network.

[21]Our results are robust to changing these.

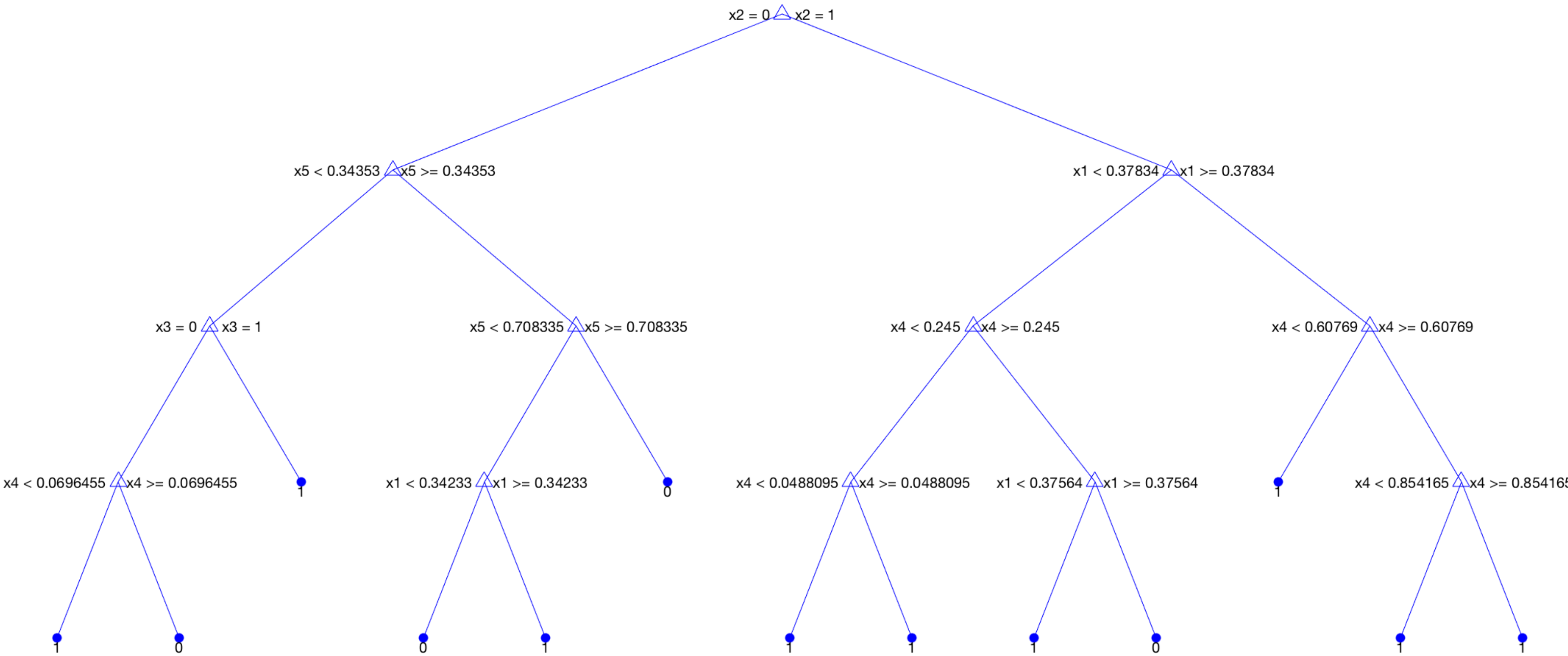


Figure B.1. This presents an example of a decision tree. For the sake of simplicity, we limit the maximum number of splits to 12. The actual procedure has a considerably more complex tree. Here x_1 is the average closeness to leaders, x_2 is whether the household is eligible by having a female of eligible age, x_3 is whether the household is a leader, x_4 is the share of same-caste leaders in the village, and x_5 is the closeness to same-caste leaders.

Appendix C. Alternative Models: Existing Models in the Literature

In this section we describe several alternative models, and emphasize why they fail to generate the patterns in our data; and we also describe an extension of our model to include direct payoff externalities across links. The goal is to provide the reader with a perspective as to why our model is new and why insights from existing models are insufficient. Of course, this is not an exhaustive list of the large number of models in the literature, but are representative of the types of models that would be natural candidates for this application.

In what follows, we take the setup of Section 4.6 to work through these models. We study four specific alternatives.

The first two involve exogenous random matching and mutual consent. These are analogous to the type of models studied by **?**Jackson and Watts (2002); Christakis, Fowler, Imbens, and Kalyanaraman (2010b); Mele (2017), albeit presented in a much more simplified manner for clarity of argument.

First, Section C.1, presents the case when links are historically given but may break as a result of a shock, such as the introduction of microfinance. New links are however slow to form and in the short run the dominant effect of shock is that links break (in the longer run new links presumably form). This is as in Jackson et al. (2012). The second model takes on the opposite extreme case where links get renewed every period from scratch. So in section C.2), we imagine an exogenous set of unlinked individuals who form new links, with random matching opportunities and mutual consent for link formation.

The third model, presented in Section C.3, returns to the case where links are easy to break but slow to form, but focuses on triads rather than pairs. This introduces the idea of support–the that the presence of one link may help sustain other links involving some of the same set of people (Jackson et al., 2012).

Despite their very different perspectives, these three models all point to similar conclusions: that the number of HL links should go down in microfinance villages, while the number of LL links should stay the same or, if it does decline, should decline less than mixed link types. Further LLL triads should decline less than LLH or LHH.

The fourth model, presented in Section C.4, returns to the setup where networks essentially re-form every period, but now introduces "directed search". With directed search, agents are free to choose which other types of agents they want to link with. In such a model, we find that while HL links should decline in microfinance villages, LL links should certainly increase. This fits with the main strand of the network formation literature (e.g., Jackson and Wolinsky (1996); Dutta and Mutuswami (1997); Bala and Goyal (2000); Currarini and Morelli (2000); Jackson and Van den Nouweland (2005); Bloch, Genicot, and Ray (2008); Herings, Mauleon, and Vannetelbosch (2009); Jackson, Rodriguez-Barraquer, and Tan (2012); Boucher (2015)...).

Taken together, these four perspectives, which take either exogenous or directed search, with mutual consent and possibly support, cannot generate patterns consistent with the data.

C.1. **The impact on pre-existing links.** The first model takes the view that villagers are in a pre-existing network, and while links are easy to break, forming new links can be very slow and thus not on the same time-scale. We start from a setting where we take these network connections as given before the arrival of microcredit. Where microcredit arrives, people have the choice of continuing or breaking off those relationships and breaking is unilateral (consistent with mutual consent models). In control villages we assume that nothing changes.

Let us write that the payoff to node i of type θ_i of being linked to j of type θ_j is given by

$$\alpha_{\theta_i}\beta_{\theta_j}r + \beta_{\theta_i}\alpha_{\theta_j}b - \epsilon_{ij},$$

where G is the CDF of ϵ, a mean-zero random variable, so as before the expected value is

$$v_{\theta\theta'} = \alpha_\theta\beta_{\theta'}r + \beta_\theta\alpha_{\theta'}b.$$

What is the effect on the number of relationships of each type: HH, LH, and LL? Clearly the number of HH relations goes down and the number of LL relationships should be unchanged. The number of HL relationships however depends on both the willingness of the H to partner with an L, which has gone down and the willingness of an L to partner with an H, which might have gone up. The number of LH pairs in MF villages given by

$$G(v_{HL} + \Delta v_{HL}) \cdot G(v_{LH} + \Delta v_{LH})$$

compared to

$$G(v_{HL}) \cdot G(v_{LH})$$

in non-MF villages. For relatively small changes in the value of the relationships the difference in the number of HL pairs can be written as

$$\begin{aligned} G'(v_{HL})\Delta v_{HL} + G'(v_{LH})\Delta v_{LH} &= G'(v_{HL})[\Delta v_{HL} + \Delta v_{LH}] \\ &= (\alpha_H\Delta\beta_H + \beta_H\Delta\alpha_H)(r+b) < 0. \end{aligned}$$

The last inequality follows from the fact that if relending is small relative to the change in appetite for borrowing (as is the case in the empirical literature), then $\Delta_{HH} < 0$ which is the same condition as above.

Therefore the number of HL relations must also fall. Only the number LL relationships do not go down when MF arrives.

Claim **1.** *Starting with a given set of links, the introduction of microfinance should*

(1) *reduce HH links,*

(2) *reduce LH links,*
(3) *leave LL links unchanged,*
(4) *and the total number of links should decline and be less than in non-microfinance villages.*

C.2. **Introducing link formation.** We now turn to a model at the other extreme: there is no persistence in links whatsoever, so the network is essentially re-formed every period. Thus we can consider the formation of new links from an unmatched population.

As before the pairs are formed if both parties want the link, which happens with probability $G(v_{\theta\widetilde{\theta}}) \cdot G(v_{\widetilde{\theta}\theta})$ for a $\theta\widetilde{\theta}$ link. From above, the fraction of new HH and LH links should go down in microfinance villages but that of new LL links should remain the same.

Claim **2.** *If new links are formed by randomly matching, the introduction of microfinance should*

(1) *reduce new HH links,*
(2) *reduce new LH links,*
(3) *leave new LL links unchanged,*
(4) *and the total number of new links should be less than in non-microfinance villages.*

C.3. **A model with supported links.** Our third model again takes the perspective that links are easy to break but slow to form, but in this case we focus on the value of a link being supported in the sense of Jackson et al. (2012). This model builds a natural connection between what happens to the Hs (who are directly affected by microcredit) and what happens to Ls. An LL link can break because it is no longer supported by an H. However for reasons that will become clear it cannot explain the patterns we observe in the data.

C.3.1. *Payoffs.* We start with a set of HH, LH, and LL links. However some of these links also support each other in the sense that some are part of HHH, LHH, LLH, or LLL triangles. We assume that no one has more than two links, to keep the problem manageable. We assume that the payoff to i from the links between i (a type θ) and j (a type $\widetilde{\theta}$) that is supported by k (a type θ') is given by

$$W_{ijk}(\theta, \widetilde{\theta}|\theta') = v_{\theta\widetilde{\theta}} + \max\{\varepsilon_{ij}, \varepsilon_{ik}\}$$

where $v_{\theta\widetilde{\theta}}$ is defined as in Section 4 and ε_{ij} and ε_{ik} are drawn, as before, i.i.d. from a distribution G.

This formulation makes sense in a world where there is no crowd-out in borrowing or lending – when an agent is in the borrowing state he gets twice the benefit b if he can borrow from two sources and when he is in the lending state he gets twice the benefit r if he can lend to two people. The modeling of the relationship specific utility term captures the intuition that when three people are hanging out together, the effect of the relationship to each of them depends on its best parts from their point of view.

When the relation is not supported, i.e., there is either just one pair or there is a potential triad but not all 3 pairs are connected, the payoff from it is, as before

$$W_{ij}(\theta,\widetilde{\theta}|\emptyset) = v_{\theta\widetilde{\theta}} + \varepsilon_{ij}$$

where the ε_{ij} is drawn, as before, i.i.d. from a distribution G.

C.3.2. *Analysis of the model.* The decision to be made is simple: whether to stay linked. However starting from a trilateral relationship there are potentially multiple equilibria: i might leave because she expects k to leave and vice versa. To reduce the number of cases, assume that the equilibrium selection rule is always to choose the triad equilibrium if it existed in the pre-period and is still an equilibrium. In other words, each participant of triad only checks whether they want to stay in the relationship if the other two members of the triad were to stay. If the triad is no longer an equilibrium then each pair in the erstwhile triad independently decides whether or not to stay together as a pair (and clearly at least one will not) and the equilibrium is unique.

Clearly some of the Hs who are in a triad and have access to microfinance will want to break at least one link since both v_{HH} and v_{HL} decline. Once this is taken as given, the value of each remaining relationship goes down at least weakly and in some fraction of cases those relationships will also break up because they were sustained by the higher ε associated with the triad. The only triads that will unaffected are the LLL triads. All other types of triads will break up more in MF villages than in non-MF villages. It is also easy to see that LHH triads are more likely to break up than LLH triads with microfinance, simply because the LH links are the vulnerable points.

This model can explain why lots of pre-existing LL links break up in MF villages. The argument would be that most of these links were part of a triad with an $H$and the H has less incentive to continue in the triad. It does however suggest that less LL links should break up than LH links, since under this theory LL links only break up because an LH links that sustained that LL relationship broke up.

Claim **3.** *In the model with supported links, when microfinance is introduced,*

(1) *LL decline but LH should decline by more,*
(2) *$LHHs$ are more likely to decline than $LLHs$ which are more likely to decline than $LLLs$.*

C.3.3. *Simulation.* To make this transparent, we present a simulation exercise. We look at networks of size $n = 300$. We set the payoff parameters $r = 0.1$, $b = 1$, and $\alpha_H = \alpha_L = \beta_H = \beta_L = 1/3$. We set $\alpha'_H = 1.45\alpha_H$ and we will vary the needing to borrow probability under microfinance, $\beta'_H \in \{0.25, 0.3, \ldots, 0.65\}$ for the simulations. Under these parameters we have v_{HH}, v_{HL}, v_{LH}, and v_{LL} satisfying the assumptions maintained throughout this paper, described in Section 4. We let $G(\varepsilon) = \mathcal{N}(0, 1/100)$ and let half the population be H and the other half be L.

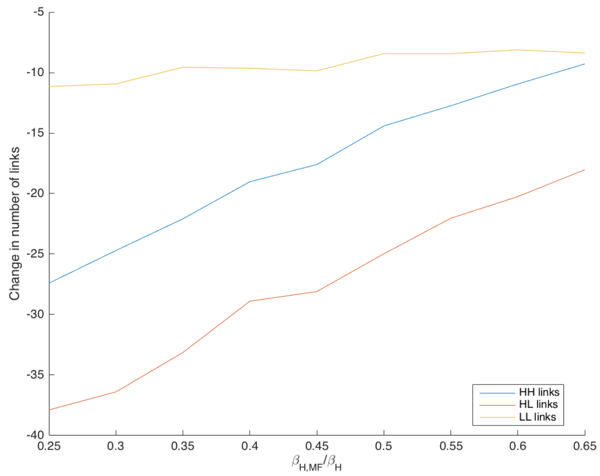
We repeat 100 simulations of the following procedure. We seed the graph by connecting collections of mutually exclusive sets of three nodes at random. We then draw ε_{ij} and compute an equilibrium network under no-MF payoffs and an equilibrium network under MF payoffs, holding fixed the seed and the shocks as above. Specifically, any triangle that exists initially and for which it is still an equilibrium under the shocks and payoff parameters to maintain are maintained. If not, then constituent links are checked. A resulting equilibrium graph holding fixed seeds and shocks can be computed for each simulation draw under both non-MF and MF payoffs.

Figure C.1 presents the results. We plot the change in the number of links (and the change in the number of triangles) comparing MF networks to non-MF networks. We see that MF networks uniformly lead to a decline in every link and triangle type. Furthermore, the gap between the models declines the closer β'_H is to β_H. Nonetheless, what is striking is that LL links drop much less than its counterparts HH and LH, as do LLL triangles compared to HHH, LLH and LHH.

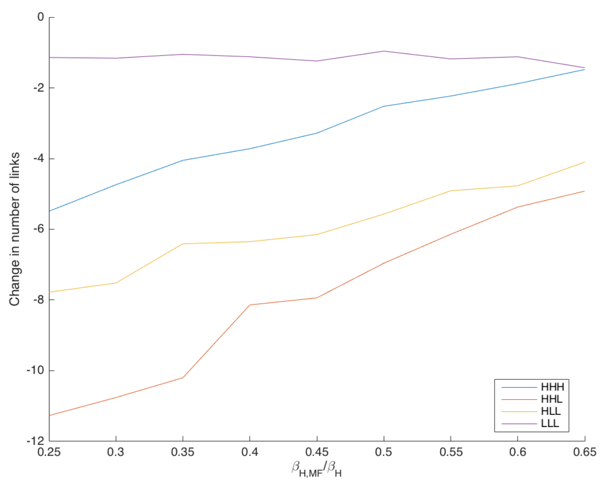
C.3.4. *Summary so far.* The simple models discussed so far with or without the idea of support all point to the same conclusion: that the number of LH and HH links should go down faster in MF villages than the number of LL links. There is however one additional factor, ignored so far, which might makes the effect on the number of HH links in MF villages potentially ambiguous. This is the fact that microfinance itself promotes connections between group members, who will tend to be Hs (Feigenberg et al., 2013).

In addition, when we think of triads rather than pairs, our final model predicts that LLL triangles should be less likely to break apart in microfinance villages than LHH or LLH triangles. Because our data shows that LL links break the fastest (or at least faster than HH) and LLL triangles break the fastest (significantly more than mixed triads or HHH triads), none of these models can explain the data.

C.4. **A model of directed search.** Let us take the set up of the model where networks essentially re-form every period, but now introduce directed search. Instead of matching randomly, we now assume that each agent can select the population within which they will match. Once they observe who they are matched to, which happens randomly within



(A) Evolution of HH, LH, and LL



(B) Evolution of HHH, LHH, LLH, and LLL

FIGURE C.1. Supported Links Model

the group, they get to decide whether they will actually form a link. Link formation is unilateral. There are three possible populations: HH (i.e., just Hs), LL (i.e., just Ls), and LH (i.e., mixed, with the fractions endogenously determined). Within the HH and LL groups everyone will get matched (assuming even numbers). Within the HL group the outcomes depends on the fraction of the two types, but we assume that the maximum possible number of matches are formed.

In this model there are spillovers from the decisions of the Hs on the decisions of the Ls. If Hs decide to stop matching with the Ls, then Ls might be forced to change their matching habits. However for reasons that will become clear, this model does not deliver the desired patterns.

In non-MF villages we have assumed that the payoffs for Hs and Ls are identical and therefore there are many possible equilibria. However in all the equilibria the shares of H and L types in the LH group must be the same.

In MF villages, observe that

$$\begin{aligned}\Delta v_{HL} - \Delta v_{HH} &= \alpha_H \Delta\beta_H b + \beta_H \Delta\alpha_H r - (\alpha_H \Delta\beta_H + \beta_H \Delta\alpha_H)(r+b) \\ &= -\alpha_H \beta_H \left[\frac{\Delta\beta_H}{\beta_H} r + \frac{\Delta\alpha_H}{\alpha_H} b\right].\end{aligned}$$

This leaves us with two possibilities. Either $\frac{\Delta\beta_H}{\beta_H} r + \frac{\Delta\alpha_H}{\alpha_H} b > 0$ or not. Assume the expression is positive. Since we started from a situation where $v_{HL} = v_{HH} = v_{LL}$, the condition implies that in MF villages $v_{HH} > v_{HL}$. Therefore all Hs will chose the HH option. Paradoxically the same condition also tells us that $\Delta v_{LH} > 0$, so in MF villages $v_{LL} < v_{LH}$. An L will prefer to be matched with an H. However the probability of being matched with an H is zero for an L, since all Hs will choose the HH option. Therefore all Ls will choose the LL option.

Or second, $\frac{\Delta\beta_H}{\beta_H} r + \frac{\Delta\alpha_H}{\alpha_H} b < 0$. In this case Hs will want to match with Ls but not the other way around. Therefore once again we will see full homophily. The fraction of both the HH and LL populations will go up and that of HL will go down in both cases. However in both cases the value of HH links has gone down ($\Delta v_{HH} < 0$), while that of LL links is unchanged. Therefore the fraction of HH links actually formed may go up or down. The fraction of LL links should go up. However in the population as a whole, the LH population turns into HHs and LLs in MF villages. Randomly formed LL pairs out of this population have the same probability of turning into an actual link as randomly formed LH pairs, but randomly formed HH pairs have lower chance of turning into an actual link. The total number of realized links should therefore be lower in MF villages.

This example is extreme but it captures a robust intuition. If microfinance makes Ls want to pair with Hs rather than with Ls it also makes Hs want to pair with Hs, and vice versa, which is why there are no LH pairs in MF villages.

Claim **4.** *If new links are formed by directed matching, the introduction of microfinance should*

(1) *either reduce or increase new HH links,*
(2) *reduce new LH links,*
(3) *increase new LL links,*
(4) *and the total number of new links should be less than in non-microfinance villages.*

We can see from the result that directed search will be inconsistent with the data namely because the effect on LL should be to increase, not decrease their presence, whereas the number of LH links will go down. Therefore, endogenous matching must generate spillovers in a another way.

C.5. **Adding More General Dependencies to Our Model.** We now describe an extension of our model to include dependencies in link presence. We describe a variation on the subgraph formation model of Chandrasekhar and Jackson (2018).

Let G be some set of potential subgraphs on n nodes. For instance, instead of just a list of all possible links, it could also include triangles, or various other cliques, stars, and so forth.

We abuse notation and let $i \in g$ for some $g \in G$ denote that i is one of the nodes that has links in g. Let $v_i(g)$ denote the utility of i if g forms. The total utility that i obtains is the sum over all subgraphs that i is part of - so rather than just a network, the resulting object is a multigraph.

We let m_g denote a relative frequency adjustment for the type of subgraph in question, as some may be more or less likely to form as a function of the efforts.

The probability that some g forms if it is not present is then

$$m_g \times_{i \in g} e_i(1 - F(v_i(g)))$$

which is the product of the socialization efforts and the probability that each i involved in g finds it valuable to form g.

The probability that a subgraph is maintained if it is already present is[22]

$$\times_{i \in g} e_i.$$

Letting $E^+[v_i(g)]$ denote the expected utility that i gets from subgraph g conditional on finding it worthwhile to form, and $\mathcal{G}^t$ denote the set of subgraphs present at the beginning of time t, then the expected utility that i gets from effort e_i is

$$\begin{aligned} V_i\left(e_i\right) = u_\theta e_\theta - \frac{1}{2} c_\theta e_\theta^2 + \sum_{g \in \mathcal{G}^t} \mathrm{E}^+[v_i(g)] \times_{j \in g} e_j \\ + \sum_{g \notin \mathcal{G}^t} \mathrm{E}^+[v_i(g)] m_g \times_{j \in g} e_j(1 - F(v_j(g))). \end{aligned}$$

The model then functions just as the model described in the text, simply with these augmented preferences over richer collections of subgraphs.

[22]One could adjust the relative impact of effort for maintaining a subgraph to be some other function than simply the product, depending on the context.

Online Appendix: Not for Publication

Appendix D. Results without Kin and Only Kin

Table D.1. Link Evolution without Kin

VARIABLES	(1) Linked Post-MF	(2) Linked Post-MF	(3) Linked Post-MF	(4) Linked Post-MF
Microfinance × LL	-0.051	-0.048	-0.004	-0.006
	(0.024)	(0.024)	(0.008)	(0.007)
	[0.038]	[0.050]	[0.620]	[0.428]
Microfinance × LH	-0.057	-0.055	-0.004	-0.006
	(0.017)	(0.017)	(0.005)	(0.005)
	[0.001]	[0.002]	[0.392]	[0.236]
Microfinance	-0.011	-0.030	-0.016	-0.011
	(0.023)	(0.023)	(0.008)	(0.008)
	[0.631]	[0.200]	[0.054]	[0.140]
LL	0.004	-0.004	-0.014	-0.009
	(0.019)	(0.019)	(0.007)	(0.006)
	[0.840]	[0.834]	[0.048]	[0.138]
LH	0.002	-0.001	-0.011	-0.008
	(0.013)	(0.013)	(0.005)	(0.004)
	[0.857]	[0.962]	[0.017]	[0.052]
Observations	57,376	57,376	846,561	846,561
Linked Pre-MF	Yes	Yes	No	No
Centrality control		✓		✓
Depvar Mean	0.416	0.416	0.061	0.061
HH, Non-MF Mean	0.450	0.450	0.085	0.085
MF + MF × LL = 0 p-val	0.002	0.001	0.010	0.016
MF + MF × LH = 0 p-val	0.000	0.000	0.002	0.005
MF × LL = MF × LH p-val	0.692	0.611	0.866	0.949

Notes: Standard errors clustered at village level in parentheses.

Table D.2. Link Evolution with Only Kin

VARIABLES	(1) Linked Post-MF	(2) Linked Post-MF
Microfinance × LL	0.012	0.015
	(0.029)	(0.029)
	[0.681]	[0.604]
Microfinance × LH	-0.010	-0.006
	(0.024)	(0.025)
	[0.692]	[0.807]
Microfinance	-0.023	-0.044
	(0.029)	(0.028)
	[0.417]	[0.121]
LL	0.010	-0.000
	(0.020)	(0.020)
	[0.612]	[0.998]
LH	-0.000	-0.005
	(0.017)	(0.017)
	[0.981]	[0.767]
Observations	13,853	13,853
Linked Pre-MF	Yes	Yes
Centrality control		✓
Depvar Mean	0.570	0.570
HH, Non-MF Mean	0.581	0.581
MF + MF × LL = 0 p-val	0.632	0.193
MF + MF × LH = 0 p-val	0.179	0.038
MF × LL = MF × LH p-val	0.246	0.261

Notes: Standard errors clustered at village level in parentheses.

Table D.3. Triples Evolution without Kin

VARIABLES	(1) Full triangle linked Post-MF	(2) Full triangle linked Post-MF	(3) Any link in triangle survived Post-MF	(4) Any link in triangle survived Post-MF
Microfinance × LLL	-0.071	-0.071	-0.090	-0.095
	(0.033)	(0.032)	(0.032)	(0.031)
	[0.032]	[0.029]	[0.006]	[0.003]
Microfinance × LLH	-0.050	-0.055	-0.053	-0.058
	(0.027)	(0.028)	(0.027)	(0.029)
	[0.071]	[0.059]	[0.056]	[0.048]
Microfinance × LHH	-0.035	-0.039	-0.032	-0.034
	(0.017)	(0.017)	(0.018)	(0.018)
	[0.043]	[0.028]	[0.077]	[0.065]
Microfinance	0.001	-0.015	-0.004	-0.014
	(0.027)	(0.028)	(0.023)	(0.023)
	[0.979]	[0.589]	[0.864]	[0.537]
LLL	0.005	-0.002	0.016	0.013
	(0.027)	(0.027)	(0.022)	(0.023)
	[0.858]	[0.944]	[0.488]	[0.554]
LLH	-0.012	-0.011	-0.013	-0.011
	(0.021)	(0.021)	(0.019)	(0.019)
	[0.566]	[0.602]	[0.504]	[0.563]
LHH	-0.001	0.003	-0.005	-0.003
	(0.014)	(0.014)	(0.012)	(0.012)
	[0.937]	[0.844]	[0.668]	[0.831]
Observations	53,233	53,233	53,233	53,233
Linked Pre-MF	Yes	Yes	Yes	Yes
Centrality control		✓		✓
Depvar Mean	0.174	0.174	0.786	0.786
HHH, Non-MF Mean	0.844	0.844	0.844	0.844
MF + MF × LLL = 0 p-val	0.012	0.002	0.000	0.000
MF + MF × LLH = 0 p-val	0.020	0.003	0.003	0.000
MF + MF × LHH = 0 p-val	0.127	0.021	0.080	0.016
MF × LLL = MF × LLH p-val	0.273	0.345	0.023	0.012
MF × LLL = MF × LHH p-val	0.168	0.187	0.021	0.009
MF × LLH = MF × LHH p-val	0.362	0.356	0.202	0.159

Notes: Standard errors clustered at village level in parentheses.

Table D.4. Triples Evolution with Only Kin

VARIABLES	(1) Full triangle linked Post-MF	(2) Full triangle linked Post-MF	(3) Any link in triangle survived Post-MF	(4) Any link in triangle survived Post-MF
Microfinance × LLL	0.038	0.030	0.025	0.021
	(0.080)	(0.081)	(0.033)	(0.034)
	[0.635]	[0.709]	[0.449]	[0.530]
Microfinance × LLH	-0.053	-0.058	0.037	0.038
	(0.069)	(0.069)	(0.028)	(0.028)
	[0.441]	[0.405]	[0.190]	[0.170]
Microfinance × LHH	-0.015	-0.012	0.029	0.033
	(0.082)	(0.080)	(0.026)	(0.025)
	[0.857]	[0.879]	[0.269]	[0.198]
Microfinance	0.011	-0.004	-0.024	-0.035
	(0.067)	(0.065)	(0.024)	(0.025)
	[0.871]	[0.947]	[0.336]	[0.161]
LLL	-0.078	-0.079	-0.006	-0.003
	(0.055)	(0.057)	(0.025)	(0.027)
	[0.164]	[0.168]	[0.805]	[0.926]
LLH	-0.026	-0.030	-0.026	-0.025
	(0.045)	(0.046)	(0.022)	(0.022)
	[0.568]	[0.514]	[0.228]	[0.249]
LHH	-0.061	-0.064	-0.015	-0.014
	(0.053)	(0.052)	(0.018)	(0.018)
	[0.256]	[0.223]	[0.404]	[0.429]
Observations	3,618	3,618	3,618	3,618
Linked Pre-MF	Yes	Yes	Yes	Yes
Centrality control		✓		✓
Depvar Mean	0.433	0.433	0.929	0.929
HHH, Non-MF Mean	0.844	0.844	0.844	0.844
MF + MF × LLL = 0 p-val	0.309	0.609	0.951	0.559
MF + MF × LLH = 0 p-val	0.282	0.101	0.470	0.865
MF + MF × LHH = 0 p-val	0.938	0.730	0.822	0.936
MF × LLL = MF × LLH p-val	0.043	0.054	0.650	0.543
MF × LLL = MF × LHH p-val	0.376	0.484	0.895	0.695
MF × LLH = MF × LHH p-val	0.448	0.367	0.718	0.819

Notes: Standard errors clustered at village level in parentheses.

Appendix E. Building Social Capital among MF Takers Cannot Explain Results

In Table E.1 we repeat our main regression of whether a link exists in Wave 2 as a function of microfinance exposure and interactions with household-type. In columns 2 and 4 we include indicators for whether at least one of the households involved joined microfinance, so the main effects are for households not involved in microfinance whatsoever. (This is clearly not to be causally interpreted, but merely illustrative.)

From this we see that our results are essentially unchanged. That is, for household pairs of type HH, HL, or LL, when a link exists in Wave 1, the greatest relative declines in the probability of the link surviving in MF villages versus non-MF villages come from LL and HL, rather than HH. The differential effects of having the households (typically Hs) joining microfinance could not have been driving our main result (the interactions are insignificant and have small point estimates). A similar phenomenon holds in column 4. Thus, we find that for the vast majority of pairs, which are not at all involved in microfinance, in microfinance villages they experience a relative decline in probability of being linked in the second period and the decline is larger for LL than for HH pairs.

Then in Table E.2 we regress whether a link exists in Wave 2 further interacting by whether one or both of the households involved joined microfinance. We can see again that our main results (for those who have no parties joining microfinance) are unchanged, demonstrating that our results are not driven by new links among microfinance members. However, it is interesting to note that HH pairs that both enroll in microfinance, that are previously unlinked, are considerably more likely (1.5pp relative to a mean of 6.4%) to form a new link, consistent with Feigenberg et al. (2013). Of course, the main effect of having microfinance for this pair is a 2.2pp decline in the probability of forming the new link to begin with, so this means that on net there is no effect: that the new relationships forged by meeting others in microfinance centers serve only to offset the greater decline in social capital overall.

Taken together, we see that (a) even looking at parties that never joined microfinance, LL types experience greater social capital losses than HH, and (b) while HHs involved in microfinance are able to stave off some of the loss in linking rates in MF villages, because microfinance takers wind up forming some links to each other, they are not nearly numerous enough to drive our main results (noting that 86% of pairs households in microfinance villages involve households that did not take-up).

TABLE E.1. Link Evolution

VARIABLES	(1) Linked Post-MF	(2) Linked Post-MF	(3) Linked Post-MF	(4) Linked Post-MF
Microfinance × LL	-0.0499	-0.0414	-0.00453	0.00123
	(0.0237)	(0.0240)	(0.00789)	(0.00816)
	[0.0387]	[0.0895]	[0.568]	[0.881]
Microfinance × LH	-0.0565	-0.0477	-0.00494	-0.000606
	(0.0166)	(0.0171)	(0.00535)	(0.00557)
	[0.00106]	[0.00686]	[0.358]	[0.914]
Microfinance	-0.00776	-0.0174	-0.0168	-0.0229
	(0.0226)	(0.0223)	(0.00871)	(0.00887)
	[0.733]	[0.437]	[0.0575]	[0.0118]
LH	0.00219	0.00219	-0.0114	-0.0114
	(0.0122)	(0.0122)	(0.00462)	(0.00462)
	[0.858]	[0.858]	[0.0160]	[0.0160]
LL	0.00600	0.00600	-0.0138	-0.0138
	(0.0177)	(0.0177)	(0.00690)	(0.00690)
	[0.736]	[0.736]	[0.0496]	[0.0496]
MF × LL × At least 1 in MF		0.00667		-0.00619
		(0.0253)		(0.00526)
		[0.793]		[0.243]
MF × LH × At least 1 in MF		-0.0165		-0.00627
		(0.0126)		(0.00310)
		[0.195]		[0.0465]
MF × At least 1 in MF		0.0197		0.0133
		(0.0164)		(0.00401)
		[0.232]		[0.00137]
Observations	57,376	57,376	846,562	846,562
Linked Pre-MF	Yes	Yes	No	No
Depvar Mean	0.441	0.441	0.0636	0.0636
HH, Non-MF Mean	0.472	0.472	0.0891	0.0891
MF + MF × LL = 0 p-val	0.00300	0.00200	0.00900	0.00800
MF + MF × LH = 0 p-val	0	0	0.00200	0.00100
MF × LL = MF × LH p-val	0.642	0.654	0.917	0.645

Notes: Standard errors clustered at village level in parentheses.

TABLE E.2. Link Evolution

VARIABLES	(1) Linked Post-MF	(2) Linked Post-MF
Microfinance × LL	-0.0414	0.00123
	(0.0240)	(0.00816)
	[0.0895]	[0.881]
Microfinance × LH	-0.0477	-0.000606
	(0.0171)	(0.00557)
	[0.00686]	[0.914]
Microfinance	-0.0174	-0.0229
	(0.0223)	(0.00887)
	[0.437]	[0.0118]
LL	0.00600	-0.0138
	(0.0177)	(0.00690)
	[0.736]	[0.0496]
LH	0.00219	-0.0114
	(0.0122)	(0.00462)
	[0.858]	[0.0160]
One takes MF	-0.0604	-0.00370
	(0.0344)	(0.0119)
	[0.0836]	[0.757]
Both take MF	0.0267	0.0152
	(0.0181)	(0.00412)
	[0.143]	[0.000437]
MF × LH × One in MF	0.0180	0.00572
	(0.0283)	(0.00878)
	[0.526]	[0.517]
MF × LL × One in MF	0.0689	0.0125
	(0.0488)	(0.0115)
	[0.162]	[0.282]
MF × LH × Both in MF	-0.0151	-0.00717
	(0.0138)	(0.00340)
	[0.275]	[0.0384]
MF × LL × Both in MF	0.0108	-0.00916
	(0.0312)	(0.00556)
	[0.730]	[0.103]
Observations	57,376	846,562
Linked Pre-MF	Yes	No
Depvar Mean	0.441	0.0636
MF + MF × LL = 0 p-val	0.00200	0.00800
MF + MF × LH = 0 p-val	0	0.00100
MF × LL = MF × LH p-val	0.654	0.645

Notes: Standard errors clustered at village level in parentheses.

Appendix F. Misc. Calculation

Consider the special case where $\alpha_L = \alpha_H$, $\beta_L = \beta_H$ and $\alpha_H \Delta\beta_H + \beta_H \Delta\alpha_H = 0$. In this case $\Delta_{HH} = 0$.

Now suppose first that $\Delta\beta_H > 0$ and therefore $\Delta\alpha_H < 0$. In this case

$$0 < \mathit{\Delta}_{HL} = \alpha_H \Delta\beta_H b + \beta_H \Delta\alpha_H r \Leftrightarrow r\frac{\beta_H|\Delta\alpha_H|}{\alpha_H \Delta\beta_H} = r < b$$

and

$$0 < \mathit{\Delta}_{LH} = \alpha_H \Delta\beta_H r + \beta_H \Delta\alpha_H b \Leftrightarrow r\frac{\alpha_H \Delta\beta_H}{\beta_H|\Delta\alpha_H|} = r > b.$$

In the case where $\Delta\beta_H < 0$ and therefore $\Delta\alpha_H > 0$, these inequalities get reversed and we get

$$0 < \Delta_{HL} \Leftrightarrow r > b$$

and

$$0 < \Delta_{LH} \Leftrightarrow r < b.$$

In other words, in this special case, Δ_{HL} and Δ_{LH} move in opposite directions and which one goes up depends on which of r and b is bigger and whether or not $\Delta\beta_H > 0$.

Since $b > r$, in this special example we would expect Δ_{HL} to be positive and Δ_{LH} to be negative as long as $\Delta\beta_H > 0$ and the reverse otherwise. In other words, it is entirely possible for v_{HL} to go up, v_{LH} to go down and v_{HH} to be unchanged but it requires α_H to go down and β_H to go up.