Peer Pressure and Loan Repayment: Evidence from a Natural Experiment

Emily Breza
MIT

October 2010

Abstract

An important feature of microfinance around the world is the peer structure of borrower repayment. Most academic analyses of peer lending to date have focused on the joint liability contract structures common to these loans. However, now that many MFIs are moving away from joint liability, it is more important than ever to understand the non-contractual social incentives for loan repayment. In this paper, I estimate the effect of peer repayment on an individual's own repayment decision. To circumvent the standard problems when trying to identify peer effects, I use a unique data set from a natural experiment in India, where 100% of borrowers temporarily defaulted on their loans. I argue that the timing of the defaults was plausibly "random" and use variation in each borrower's dynamic repayment incentive (which I find to be quite strong) to instrument for eventual loan repayment. I find that if a borrower's peers shift from full default to full repayment, she is 10% more likely to repay her own loan. I also find preliminary evidence that these peer effects are decreasing in social distance. Finally, it is important to note that repayment peer effects can create both positive and negative incentives for borrowers.

JEL: C81, G21, O12, O16

PRELIMINARY AND INCOMPLETE

1 Introduction

Group lending has always been a central feature of Grameen Bank-style microfinance. The Grameen Bank website even claims "there is more to the bank than just the balance sheet; it ties lending to a process of social engineering."[1] The peer lending context has been exported and replicated across the globe to diverse cultures and settings and has remained surprisingly successful at providing strong incentives for loan repayment. However, we know surprisingly little about how the peer

[1] See *http://www.grameen-info.org*

dimension of these loans actually affects a borrower's repayment decisions or if there are aspects of the peer structure that carry over to other areas of the borrower's financial life.

Many group lending schemes have historically been characterized by group level joint liability. In these contract structures, there is a direct role for peer decisions to affect repayment rates. For example, if one member defaults on her loan, then the remaining members must bear the cost of that defaulted loan if they intend to continue to recieve loans from the organization. This extra cost may result in other repayers choosing to default and walk away from the lending relationship. Alternately, the non-defaulting borrowers could use a local enforcement technology to coax the defaulter into repaying her loan. In both of these scenarios, the actions of the peer group have direct consequences for an individual's own repayment decisions. Several theoretical models examine various mechanisms through which joint liability operates including screening, monitoring and enforcement. Candidate pathways include moral hazard and project selection, Stiglitz (1990), moral hazard and project effort, Banerjee, Besley and Guinnane (1994), adverse selection of borrowers, Ghatak (1999), and village sanctions and limited contract enforcement, Besley and Coate (1995). These models have different predictions for borrower repayment, but all conclude that peer behaviors should affect individual decisions. Ahlin and Townsend (2002) use data from Thailand to test the theoretical predictions and find that higher degrees of joint liability coincide with lower repayment as do higher levels of cooperation within borrower groups. Their results highlight the potential for perverse social effects on repayment. Using quasi-random group formation data, Karlan (2007) finds that stronger social connections imply higher repayment rates in joint liability groups in Peru and that default is detrimental to social ties.

While the joint liability literature gives a rich theoretical framework for thinking about peer effects in microfinance, the social forces seem to be deeper than the details of the contract structure. Many MFIs have maintained the group repayment format (weekly public repayment meetings, oaths etc.) but have begun to eliminate the joint liability feature of each loan contract. A small but growing set of empirical investigations attempts to link social capital and microfinance in the absence of joint liability. Gine and Karlan (2007, 2008) randomize between individual and joint liability contracts to already formed joint liability groups while maintaining the group repayment meetings. Over 1 and 3 year horizons, moral hazard does not appear to increase when clients are assigned to the individual liability treatment, default rates do not increase, and the individual liability policy attracts more new clients. However, the remaining question is to what effect, aside from screening, does the peer format of the individual liability loans contribute to repayment incentives. Another study, Feigenberg et al (2010), examines social effects in the absence of joint liability. The authors find that individuals randomly assigned to weekly vs. monthly repayment schedules form stronger ties with their fellow group members, visit fellow group members more frequently, and exhibit more trust among themselves. The weekly groups also display better repayment records throughout their second loan cycle. The authors argue that the stronger social safety nets caused by the more frequent meetings allow borrowers to better smooth shocks and thus to pay their loan installments more reliably. While their results are suggestive, we still don't understand how the social ties affect repayment or the durability of the social effects under aggregate shocks.

I seek to better understand the effects of peer repayment decisions on each individual's own repayment behavior using a novel idenification strategy. In general, it is hard to measure peer effects due to the problem of unobserved, correlated shocks (see Manski (1993)). To circumvent these problems, I use data from a natural experiment in India. In March 2006, the District Collector of Krishna District in the state of Andhra Pradesh announced that all borrowers should stop repaying their microloans. Within two days of the announcement, all borrowers had ceased making loan payments. Soon after the defaults, the local microlenders began to reestablish collections in the affected villages and also suspended the joint liability feature of the loans. Some individuals resumed payment within a few months of the crisis, and as of November 2009, 40-50% of individuals had fully repaid their liabilities. I exploit the random timing of the shock to provide identification of repayment peer effects. Since microlenders use the promise of new loans to encourage repayment, borrowers who are closest to receiving a new loan have the biggest potential benefits from repayment. Thus, each individual's location in the 50-week loan cycle at the time of the defaults can be used as an instrument for repayment. Since loan cycles are all 50 weeks long, individuals in weeks 45-50 will have much stronger incentives than individuals in weeks 0-5. I therefore can also treat the identification problem as a good candidate for the fuzzy regression discontinuity approach. Since I have data on the full universe of loans, including village of residence and group membership, I can also shed light on which level of peer interaction (i.e social distance) displays the largest effects on repayment behavior.

I find that overall, borrowers are very sensitive to their own dynamic repayment incentives and that each completed payment before the defaults corresponds to a 1% increase in eventual repayment. Furthermore, borrowers appear to become more satiated and less incentivised to repay with each additional loan; borrowers are 5% less likely to repay with each consecutive loan cycle, conditional on loan size variables. I do find evidence that peer repayment behavior influences loan repayment in addition to these individual effects. Borrowers are 10% more likely to repay their loans if their entire center[2] repays. The estimates at the village-level peer group are quite a bit smaller, and provide suggestive evidence that the peer repayment effect is a largely local phenomenon. I plan to further disentangle the mechanisms driving these results using data that are still being collected in the field.

While a mass default episode might seem like a special case in which to look for peer effects, estimates from the Krishna defaults may be helpful in understanding the effects of stresses to the microfinance industry. It is exactly during crises when peer effects could matter most and when joint liability becomes impossible to enforce.[3] In many microlending contexts including India, loan repayment rates amazingly hover close to 100%. However, in the three years following the Krishna defaults, the MFI was only able to coax approximately half of the borrowers into paying their debts.

[2] Borrowers are assigned to groups of 6-10 individuals. Every 3-6 groups are then combined to form a center. All members of each center meet together at the same time and place each week to make their loan payments.

[3] Spandana fully suspended joint liability in the wake of the crisis.

This is especially puzzling since after the crisis subsided, little about the operations of the MFI, the loan products offered or the needs of the borrowers had changed. Peer coordination equilibria might help us to rationalize these outcomes. Furthermore, popular financial news sources such as the Economist and the Wall Street Journal have expressed concern about the vulnerability of microfinance (see *Froth at the Bottom of the Pyramid: the Debate Over a "Bubble" in Microlending,* Economist. Aug. 2009 or *A Global Surge in Tiny Loans Spurs Credit Bubble in a Slum*, WSJ. Aug. 2009). This study sheds light on the peer component of microfinance crises.

The body of the paper procedes as follows. Section 2 describes the setting of the natural experiment and the data set used. Section 3 provides a graphical analysis of the key exogenous variables and outlines the intuition behind the identification strategy. Section 4 describes the empirical model in more depth. Section 5 details the results, and section 6 concludes.

2 Empirical Setting

2.1 Spandana Group Loans

Before describing the default crisis in detail, it first might be useful to summarize some basic information about the MFI from which all of the data for this analysis come. Spandana Sphoorty Financial Limited is one of the largest MFIs in India and was one of the primary microlenders operating in the Krishna District at the time of the defaults. The standard loan product offered by Spandana before the defaults was a weekly repayment, joint liability loan, which was only available to female borrowers and is still the main product offered by the MFI across 10 states in India. Before the first loan disbursement, each borrower is assigned to a borrowing group of approximately 10 women. Every 2-5 borrowing groups within the same village are then combined to form a center. All borrowers belonging to a center meet at the same time each week to make their installment payments to the credit officer, who travels from the branch office to the borrowers' village. All group members cosign for the loans of the other members under the joint liability contract structure. Groups within a center also cosign for the other centers. However, it is unclear if the cross-group joint liability is ever enforced. After the successful completion of a 50-week loan cycle, new, generally larger loans are disbursed to each borrower. Borrowers within a group may have different loan sizes, depending on the MFI's assessment of each borrower's ability to repay. Since I have access to administrative data, my notion of peer hierarchies is derived from Spandana's repayment structure. I use group, center and village, in increasing order of social distance, to define peer groups for the rest of the analysis.

2.2 The AP Crisis

The setting for my investigation into peer effects and borrower repayment is a natural experiment in the Krishna District of the state of Andhra Pradesh, India. In March 2006, the District Collector, Navin Mittal, closed over 50 branches of two large MFIs, Share and Spandana. This move resulted in the cessation of all weekly repayments across the district, a potential loss of close to Rs 200cr (~$40mm) of outstanding loans. The district government alleged that MFIs were setting interest rates too high, using unethical means to encourage loan repayment, and stealing clients from state banks and SHGs (self help groups). Furthermore, several farmer suicides were blamed on the stress from having to repay microloans. The local media (whose chit fund companies stood to benefit from the departure of MFIs from the region) began a campaign of bad press and personal attacks highlighting the evils of the microfinance industry. There is some weak evidence from the local newspapers that Mittal scheduled his announcement around International Women's Day, which occurs each year on March 8. Sa Dhan and an alliance of MFIs put pressure on the government to rescind the District Collector's statement. The worst of the crisis finally came to an end in early 2007, and the Krishna government permitted borrowers and MFIs to resume their lending relationships freely.

The District Collector's announcement spurred a diverse set of reactions across the region. Some borrowers started repaying their loans within a few months of the announcement and picked up where they left off in the repayment schedule. In other parts of the district, angry villagers threatened Spandana and Share field staff and forced branch closures and the halt of all collection attempts. Some defaulters still claim that the Disrict Collector's statement remains in effect even though Mittal eventually issed a retraction and was transferred to a different region. As of November 2009, approximately 45% of the outstanding portofolio on March 9, 2006 had been repaid. This number is too large to reveal strong preferences for loan default, but too small to be explained by borrowers simply being able to resume payment where they left off.

The media's treatment of the crisis highlights the controversy associated with microfinance in the Krishna District and acknowledge the importance of peer interactions. Between March and June 2006, there were frequent negative articles about microfinance in the Vijayawada edition of The Hindu, the English language newspaper. A common view seems to be that "the microfinance companies sanction loans to SHGs liberally without insisting on security but charge exorbitant interest and collect the instalments using peer pressure of the group." A stronger complaint came in an article entitled "Microfinance victims petition rights panel," which claimed that borrowers "were caught in a perennial debt trap by the MFIs through machinations. The breach of human rights by the firms drove at least 10 persons to suicide in the district." (I could not find any articles substantiating the link between MFIs and suicides in Krishan District.) It is clear that the media understood that the peer enforcement channel mattered in the process of loan collections. One strongly-worded article notes, "micro-finance institutions have hit upon a new and unscrupulous method of recovering outstanding loans – pitting members of self-help groups against another."

Due to the political climate, Spandana was forced to take a measured response to the crisis. While the defaulted loans had been issued under group joint liability, Spandana was forced to abandon the policy to encourage loan repayment. They made the decision to reward repayers with future credit, regardless of the time spent in default and could satisfy all demand for new credit. After one year of only marginally successful collection efforts, Spandana also started offering refinancing plans, where small new loans were disbursed to encourange borrowers to begin making regular loan payments and to eventually repay all outstanding debt.

The Krishna District Spandana defaults represent an ideal natural experiment for studying the determinants of microloan repayment. First, the defaults were instigated by a federal beaurocrat from a different part of India. The defaults did not spread across district borders, indicating that true political upheaval did not drive the crisis. Since loan repayment rates remained at close to 100% in neighboring districts, it is safe to assume that in the absence of Mital's announcement, Krishna loan repayment rates would also have been almost 100%. Moreover, according to the MIX (Microfinance Information Exchange), Spandana had less than 0.01% of its portfolio overdue more than 90 days in both 2004 and 2005. In 2008, after the crisis had subsided, the reported portfolio at risk > 90 days was again very low at 0.02%. Second, Spandana is one of the largest, most efficient MFIs in the world. The MFI was able to withstand the liquidity shock from the suspension of loan payments on almost 200,000 loans and fulfill its promise of future credit to all repayers. ICICI Bank owned many of the defaulted loans, further insulating Spandana from liquidity effects. Spandana also was able to retain and pay its field staff even when all collections had ceased. Usually default crises are accompanied by other problems endemic to the MFI. Finally, as a result of the crisis, SKS and other MFIs decided not to expand their client base into Krishna district. Spandana and Share have also agreed not to add new borrowers in the district. This improves the repayment incentives for defaulted borrowers since alternate sources of MFI credit are not available.

2.3 Data

Spandana employees graciously shared all of their available electronic records with me. The data used in the analysis represent a close to complete set of loans outstanding during the AP crisis and report on loans seviced by all 23 branches[4] that are currently operating in the district. Spandana borrowers make loan payments weekly for 50 weeks and then receive a new loan disbursement, which is generally larger each cycle. The modal first cycle loan is Rs 8,000 (~160), with the loan amount generally increasing to Rs 10,000 (~200) upon successful repayment. All of the borrowers in the dataset are women, which is standard practice for Indian MFIs that follow the Grameen Bank model. The data set includes information on group name, center number and village or slum name as well as details about the specific loans including loan size, date of disbursement, loan cycle and repayment information. Table 1 gives an overview of the dataset.

Note that the data set contains information on 194,312 loans and 162,835 borrowers. The average loan size is Rs 6,462 (~$130) with loans ranging in from Rs 1,000 to Rs 30,000. The portfolio at risk on March 9, 2006 reflected in the data totals >$11mm. The average loan for which a payment was missed was disbursed 200 days before 3/9/2006. Since loan cycles are 50 weeks,

[4]Though Spandana has consolidated some of its branches since the collector's statement, these 23 branches service all affected loans.

or 350 days long. This means that the average loan was 57% repaid at the onset of the default. Loans within 1 standard deviation of date of disbursement were handed out ±104 days around the mean.

Additionally, 20% of borrowers had multiple loans outstanding at the time of default. Borrowers are given the chance to take small, interim loans after making many on-time installment payments. The interim loans also require fixed instalment payments for 50 weeks and add to the client's total liability. Once the main loan has been fully repaid, clients are permitted to take a new main loan, even when the interim loans are still outstanding. Thus, it is common for borrowers to have two Spandana loans simultaneously outstanding. These interim loans also explains the high incidence of loans smaller than Rs 6,000.

The data set does not have unique identifiers, even though many borrowers have both main and interim loans outstanding. I use fuzzy matching on the borrower name, group name, center number, and village name to identify multiple borrowing. Since village name, center numbers and group names also have multiple spellings, I have probably underestimated the incidence of multiple borrowing. I plan to improve the matching in future versions, which should decrease the number of borrowers affected by the collector's statement.

The administrative records also include week-specific payment and delinquency information that allows me to determine when a borrower resumed paying her loan and when she completed making payments on the delinquent loan. Of the 194,312 loans in my dataset, approximately 60% are still in arrears as of November 2009. The average loan outstanding conditional on being positive is Rs 3,346, which is greater than the average amount due at the time default. This is consistent with the result that borrowers with a higher debt burden are less likely to repay.

For the empirical analysis, I drop all villages with fewer than 50 borrowers. This is for two reasons. First, in the peer effects regressions, villages with only one group or center would be automatically dropped since there is no extra-group variation available. Second, it is possible that villages with only a few borrowers have miscoded names. Additionally, I drop any villages without documented cycle numbers, since I would like to be able control for weeks in the lending relationship with Spandana. Lastly, I focus only on each borrower's main loan, since these are the largest loans and furthermore, the main loans are what matter for group formation. I also drop any loans smaller than Rs. 3000, since these are most likely mis-coded interim loans. Making these cuts reduces the universe of loans to approximately 125,000 loans and borrowers. The average loan size and default size are larger after the cuts. The number of villages falls to 668 as well.

Finally, I am currently collecting additional loan data from archived hand-written ledger books stored in Spandana branch offices. These data will include additional information on borrower savings balances (which also affect repayment incentives) and will also help to correct coding errors in the electronic records, further allowing me to match borrowers across loans and distinguish general loans from interim loans.

3 Graphical Analysis

3.1 Repayment Incentives Across the Loan Cycle

The key task in this analysis is to find plausibly exogenous variation in the repayment behavior of each borrower's peer group. Note that each borrower's incentive to repay her loan changes over the 50-week cycle. Since loan installments are all the same size and are made weekly, the cost of paying off the remainder of the loan is decreasing as the loan cycle progresses (i.e. borrowers approach the maturity date of their loans). Additionally, MFIs almost universally use dynamic incentives to encourage repayment. Lenders use the promise of new, often larger amounts of credit to motivate borrowers to repay their loans. Thus, as the weeks in the loan cycle progress, the borrower is closer to receiving the next loan disbursement. So with discounting, the costs of repaying the remaining loan burden are decreasing and the benefits of paying off the loan in full are increasing. Therefore, repayment incentives are strongest in week 49 and weakest in week 0 of each loan cycle. Note that since the Krishna defaults all occured within a very small time window and since the timing of the initial loan disbursements was staggered, Mittal's announcement induced variation in the repayment incentives of borrowers across the district. As a result, week in the loan cycle is a good candidate for quasi-exogenous variation in repayment. I can also aggregate the individual incentives to the peer group level.

This argument is presented visually in Figure 1. Suppose that an MFI opens for business at time T=0. Borrower 1 expresses interest in borrowing from the MFI and receives a loan at T=0. The loan cycle is 50 weeks long and if everything goes as planned, the loan will be paid off at T=50, and a second loan will be disbursed. Also suppose that the MFI is constrained in how quickly it can expand its lending practices. Borrower 2 also expresses interest in taking a loan, and she receives her first loan at T=10, to be paid off at week T=60. Now suppose that the collector makes a statement instructing both borrowers to cease repayment at T=55. At the time of the defaults, borrower 1 is in week 5 of her second loan, while borrower 2 is in week 45 of her first loan. Thus, borrower 2 is only 5 weeks of payments from finishing the loan and recieving a second loan. On the other hand, borrower 1 has 45 installment payments to make before the third loan is disbursed. The difference in dynamic incentives implies that if borrowers 1 and 2 are otherwise identical, the probability that borrower 2 repays will be higher than the probability that borrower 1 repays.

If the collector's announcement did not contain any real information, why should the repayment probabilities be any less than 100%? If the collector's statement was really a time-out, then why do we only see 45%-50% repayment rates? Gaining a better understanding of this puzzle is one of the motivations for this project. The peer channel, and potential coordination equilibria comprise one possible explanation.

Figure 1 also brings to light some of the assumptions required for identification of repayment incentives using weeks in the loan cycle. First, it must be the case that conditional on observables, the individuals that fall to one side or the other of the 50-week point are not systematically different.

For example, it might be the case that local leaders are the first to adopt microfinance in any given village or neighborhood. Then the difference in repayment outcomes between borrowers 1 and 2 might also pick up varying tendencies to repay as a function of leader status. However, since we know when each borrower started taking loans from the MFI, we can control for smooth functions of this timing variable.[5] A similar argument might be made for loan size, since the loan size increases at each new disbursement. Again, we can include controls for functions of loan size and use the variation in timing for identification, partialling out these other effects.

Another concern would be that the district collector timed his statement to coincide with the loan cycles of key constituents. The assumption required for identification is that the timing of the announcement was not related to any of the cyclical timings of borrowers. It would also be problematic if the announcement coincided with some change in Spandana's expansion strategy 45-55 weeks prior. In section 4.2, I examine this further.

Figure 2 shows the distribution of the number of loans by week in the loan cycle on March 9, 2006. The largest concentration of loan disbursements occured between 40-50 weeks before the defaults. If the District Collector had intended to maximize the windfall gain of the borrowers of the Krishna district, it would have been optimal to wait one or two months before making his announcement, so that more borrowers could default on fresh loans. There is a second, smaller spike in loan concentration around week 80. It is also comforting that this increase does not coincide with any multiple of 50. The figure also shows that relatively few of the borrowers were on their third loans at the time of the defaults. This pattern could either be the result of slow initial growth in loan disbursements when Spandana entered the district or the result of borrower attrition between loan cycles, as I can only observe a borrower if she held a loan on March 9, 2006.

The relationship between week in the loan cycle and repayment is plotted in Figure 3. For the bulk of the analysis, repayment is an indicator for the individual having repaid the entire loan by November 20, 2009.[6] Each point in the scatter plot represents the average repayment as of November 2009 of all loans disbursed in the same week for borrowers in the same cycle. The points between 0 and 50 correspond to borrowers in their first loan cycle, while the points in weeks 50-100 correspond to second loans etc. The smooth line in the figure is a median cubic spline of the relationship between weeks with the MFI and repayment. Note that there is a large discontinuous drop in the repayment probability at each multiple of 50. Within each loan cycle, the relationship between weeks and repayment appears to be linear with the same slope in each cycle. However, the overall likelihood of repaying the defaulted loan seems to decrease with the cycle. I will investigate these relationships more in the regressions later in the paper. The relationship between the dynamic incentive and loan repayment is quite clear in this figure. The discountinuous jumps at the start of each new loan cycles also suggests a regression discontinuity approach when trying to estimate

[5] However, since borrowers were in cycles 1,2, and 3 at the time of the announcement, it is not always the case that the leaders had the better incentives. I plan to investigate this further in the future.

[6] November, 2009, which is three years after the resolution of the collector's statement, is an appropriate at which to separate long term repayers from long term defaulters. The Spandana staff predicts that it might be possible to collect at maximum 10% of the remaining debt outstanding.

peer effects.

3.2 Preliminary Peer Effect Evidence

Figures 4-9 give some suggestive evidence that there are repayment peer effects, at least at some definitions of the peer group. Figures 4-6 focus on the village level peer group, while figures 7-9 focus on the center level peer group. Figure 4 plots the estimated density of the average number of weeks borrowing from Spandana, averaged at the village level. Note that this distribution is single-peaked and centered around 40. Figure 5 gives the density by average week in the loan cycle and has mean close to 30 and a symmetric shape. Figure 6 plots the distribution of average village repayment, the key outcome variable, across all 668 villages in the sample. The shape of the repayment distribution is single-peaked and roughly symmetric, with the majority of villages experiencing repayment levels between 20% and 80%. Very few villages have close to universal default or universal repayment. The repayment distribution looks roughly consistent with the weeks in loan cycle distribution. These facts don't necessarily point to strong village peer group correlations.

Figures 7-9 give us a very different picture of potential within-center peer correlations. The plot of the densities of center repayment suggests strong correlations of repayment behavior within each center. While the average week in the loan cycle within each center is spread across all 50 possible points in Figure 8, the density of repayment has a clear, double-peaked shape with a large fraction of centers exhibiting repayments of close to either 0 or 1. The bi-modal shape of center repayment is suggestive of a strong peer effect that works in both perverse and virtuous ways. There is a strong attraction toward full default or full repayment that is not observed at the average village level densities. These densities give preliminary evidence of repayment peer effects. The striking difference in the repayment profile over villages versus centers also suggests that social mechanisms might be stronger at closer social distances. However, other sources of correlation within peer groups might also be responsible for these patterns, so we need to use the quasi-random variation induced by the timing of the shock to say something more definitive.

4 Empirical Strategy

The graphical analysis shows a relationship between individual repayment and peer repayment, and shows how the discontinuity in repayment incentives allows for the estimation of the peer effect. Before a further discussion of the results, it is important to understand the statistical inference problem at hand. The equation of interest is

$$repay_i = \beta^0 + \beta^1 repay_{p(i)} + \beta^2 X_{i,p(i)} + \varepsilon_{i,p(i)} \tag{1}$$

where i indexes the individual and $p(i)$ indexes the peer group net of the individual. The variable, $repay_i$ is a measure of individual i's loan repayment, $repay_{p(i)}$ is a measure of repayment by i's peers, and $X_{i,p(i)}$ is a set of additional individual and peer-level controls.

The biggest problem confronting most peer effects or social learning empirical identification strategies is that the peer effect cannot be separated from correlated shocks or other omitted group-level characteristics using an OLS framework. In a simple OLS regression, shocks to the entire peer group could be misinterpreted as peer effects. Manski (1993) shows that without an instrument or strong restrictions on the form of the peer effect, β^1 can't be identified, and that the OLS estimate $\hat{\beta}^1$ would pick up any correlation between $repay_{p(i)}$ and the error term.

4.1 Consistent Strategies in the Literature

Peer effects questions frequently arise in the labor and development economics literatures, and researchers have developed several approaches to consistently identify the setting-specific equivalents of equation 1. One set of approaches involves separating the peer group from the common shock group. In their paper on learning about new technologies in Ghana, Conley and Udry (2010) identify information peers and use geographical neighbors to control for common growing conditions. Similarly, in their study of social views towards contraception, Munshi and Myaux (2006) separate each individual's own religous group from other religious groups in the same village. The authors use the fact that the relevant social norms for any individual come exclusively from her own religious group. Others have tried an instrumental variables approach to solve the identification problem. For example Duflo and Saez (2002) instrument peer behavior with average group characteristics when analyzing social effects on 401K contributions. The identification of education externalities also faces the same problems. Acemoglu and Angrist (2000) use separate instruments for own education and for the education of an individual's peers. In order to estimate education externalities across age cohorts in Indonesia, Duflo (2004) finds an instrument for peer education that is orthogonal to the individual's own educational attainment.

4.2 Identification Discussion

My strategy for identification involves using week in the loan cycle at the time of default conditional on other observables as an instrument for individual and peer repayment incentives. I make two important identifying assumptions: first, in absence of Mittal's announcement, 100% of borrowers would have repaid; and second, the timing of Mittal's announcement was exogenous. These assumptions both seem plausible for the reasons discussed in the previous section. In all but a few border villages, the neighboring district maintained near-perfect repayment during the Krishna District crisis.

Table 2 compares the average week in the loan cycle at the time of default with village characteristics from the Census of India. Ideally, village characteristics will not be correlated with the distribution of weeks remaining in the loan cycle for the borrowers who live there. The Indian Census has information on population, caste break-down, education facilities, medical facilities, access to taps or wells, communication facilities, banking facilities and types of roads. The table only includes information for rural villages and may be incomplete due to different spellings of village names between the Spandana data and the Census.[7] However, for the 310 matched villages, the average weeks variable does not seem to be related to many of the covariates available in the Census that capture demographic information, land area, access to finance, access to education, and access to health care. The first column looks at the relationship between various village covariates and the average village week for the full sample. Each regression coefficient comes from a separate univariate regression, and standard errors are in parentheses below. Most of the variables are not correlated with weeks in the cycle. However, places with lower week averages tend to have higher populations, might be marginally farther from the nearest town and have fewer primary schools per capita. It would be problematic if Mittal chose to make the announcement when his cronies had most to gain, so I also include the fraction of the village between weeks 0 and 5 and the fraction between 45 and 50 in the loan cycle as regressors. Columns 2-3 show the coefficients from regressions of village characteristics on both the fraction of the village in weeks 0-5 and the fraction in weeks 45-50. Again, most of the coefficients are not significantly different from 0. There is a positive relationship between distance to town and both of these variables. However, the difference is not significant. Finally, column 3 restricts the sample to only those villages with a majority of borrowers in weeks 0-5 or a majority of borrowers in weeks 45-50. This reduces the sample size, however, to only 115 villages. Villages with high concentrations of borrowers in weeks 0-5 are relatively bigger and have relatively more health centers per capita. However, none of the education, transport, finance or irrigation variables are significantly different. It is a bit disconcerting that population is correlated with the instrument, but I can control for size of village peer group in the regression specifications.

In the estimation procedure, the implicit baseline first stage for each individual, is

$$repay_j = \alpha + \gamma^1 weeks_j + \gamma^2 f_1 \left(loanamount_j\right) + \gamma^3 f_2 \left(MFI\_weeks_j\right) + \delta^1 X_{p(j)} + e_j \tag{2}$$

where, $week_j$ is the number of weeks elapsed in the loan cycle at the time of default. The variable $loanamount_j$ provides a control for each individual's loan size, which is endogenous to other peer factors. Spandana tries to assess a borrower's debt capacity, so wealthier households are generally allocated larger loan sizes. This could interact with the peer network in the community, so I add a third degree polynomial of the loan size. Similarly, the order in which villagers signed up for their initial Spandana loans might also be correlated with peer structures. Since the gap in the loan cycle is correlated with this ordering variable, I also control functions of MFI_weeks_j, the number of weeks at the time of default since the disbursement of the borrower's first loan. Identification comes from the fact that loan cycles only last 50 weeks. Those individuals who are early in their

[7] I hope to have higher match rates in future versions of this paper.

second loan cycles should have similar characteristics (i.e. continuous, not discrete differences) to individuals late in the first loan cycle. $X_{p(j)}$ is a vector of peer group controls, explained below.

The real equation of interest is

$$repay_{i,p(i)} = \beta^0 + \beta^1 \sum_{j \in p(i)} \frac{repay_j}{\|p(i)\|} + \beta^2 weeks_i + \beta^3 X_{i,p(i)} + \varepsilon_{i,p(i)} \tag{3}$$

Where $\sum_{j \in p(i)} \frac{repay_j}{\|p(i)\|}$ is average peer repayment in person i's peer group $p(i)$. If the peer effect is linear, then the endogenous peer repayment term can be instrumented using the average weeks in the loan cycle conditional on average village loan sizes and starting dates with Spandana. See A for a derivation. The key requirement for identification is that $weeks_i \perp \sum_{j \in p(i)} \frac{weeks_j}{\|p(i)\|} | X_{i,p(i)}$. Because of this requirement, all peer groups are calculated excluding the borrowing group. So, $p(i)$ is either village ex group or center ex group in the various specifications. I argue that all of the peer group information lies in the starting date with Spandana. Thus, the orthogonality requirement is plausible conditional on functions of start date.[8]

So the peer-group level first stage is

$$\sum_{j \in p(i)} \frac{repay_j}{\|p(i)\|} = \delta_0 + \delta_1 \sum_{j \in p(i)} \frac{weeks_j}{\|p(i)\|} + \delta_2 X_{i,p(i)} + \eta_{p(i)} \tag{4}$$

where $X_{i,p(i)}$ are the appropriate individual and peer level controls.

As an alternate specification, I also estimate the peer effects regression using the following aggregate first stage:

$$\sum_{j \in p(i)} \frac{repay_j}{\|p(i)\|} = \gamma_0 + \gamma_1 \sum_{j \in p(i)} \frac{(0 \leq weeks_j < 5)}{\|p(i)\|} + \gamma_2 \sum_{j \in p(i)} \frac{(45 \leq weeks_j < 50)}{\|p(i)\|} + \gamma_3 X_{i,p(i)} + \psi_{p(i)} \tag{5}$$

In this specification, the instruments are the fraction of the peer group in weeks 0-5 of the loan cycle and the fraction of the peer group in weeks 45-50 of the loan cycle. I also use dummy variables for a peer group exceeding specific concentrations of individuals in one of these categories. In the vector of controls $X_{i,p(i)}$, I include controls for the highest and lowest values of the variables weeks with Spandana and loan size within the peer group.

The figures in section 3 suggest a possible RD interpretation of the week in loan cycle variation. Hahn, Todd and Van Der Klaauw (2001) and Van Der Klaauw (2002) establish a strong connection between IV and fuzzy regression discontinuity designs. One option, which is used by Angrist and Lavy (1999), is to run the same IV specification, but with the sample restricted to only the data points close to the discontinuities. In their paper on the application of regression discontinuity, Imbens and Lemieux (2008) reiterate the equivalence between local linear regression on either side of the discontinuity and Two Stage Least Squares using a dummy variable for data points to right

[8]This requirement can be somewhat relaxed by using a regression discontinuity strategy on average peer week, holding difference between an individual's week and average peer week constant.

of the threshhold as the instrument. This procedure also involves restricting the sample to a small window around the discontinuity. The new first stage is

$$R_{p(i)} = \delta_0 + \delta_1 W_{p(i),T} + \delta_2 X_{i,p(i)} + \eta_{p(i)} \tag{6}$$

where

$$\begin{aligned} R_{p(i)} &= \sum_{j \in p(i)} \frac{repay_j}{\|p(i)\|} \\ W_{p(i),T} &= 1\left(\sum_{j \in p(i)} \frac{(45 \leq weeks_j < 50)}{\|p(i)\|} > T \right) \end{aligned}$$

and T is the threshhold. The regressions are restricted to peer groups for which either

$$1\left(\sum_{j \in p(i)} \frac{(45 \leq weeks_j < 50)}{\|p(i)\|} > T \right) = 1$$

or

$$1\left(\sum_{j \in p(i)} \frac{(0 \leq weeks_j < 5)}{\|p(i)\|} > T \right) = 1$$

These regressions are also performed for both definitions of the peer group and only use information close to the discontinuity.

5 Results

5.1 OLS Results

Table 3 details the OLS estimates of equation 1. Column 1 shows a simple bi-variate regression of individual repayment on village repayment excluding the borrowing group's repayment. For the remainder of the analysis, all peer variables exclude the borrowing group. The reason for this is that most borrowers within a group receive their loans each cycle at the same meeting, so there is little to no variation in weeks in loan cycle at the group level. Therefore, I focus on the village peer effect and the center peer effect excluding the group. Column 1 shows a relationship of 100% between village repayment and individual repayment. This means that if the entire village switches from full default to full repayment, the individual borrower makes the same switch. Column 2 separates the village peer effect from the center peer effect. The village peer effect is on the order of 0.4, while the center peer effect is on the order of 0.6 with no other controls. Columns 3 and 4 repeat the same regressions, but add branch fixed effects. The peer repayment coefficients are slightly smaller, but still quite large. Column 5 adds individual controls to the village and center regressions. These basic controls include loan size, weeks with the MFI and weeks in the loan cycle at the time of the defaults. Finally column 6 includes peer group level controls for average loan size and weeks with the MFI. In the full specification, the OLS estimate for the peer effect is 0.772, which is quite precise. Note that two-thirds of the "peer effect" comes from the center. This fact is consistent with the shapes of Figures 6 and 9. Again, we need to be cautious in our interpretation of these estimates. These types of estimates tend to be greatly overstated in the case of unobserved correlated covariates or shocks.

5.2 Individual First Stage Results

We rarely have quasi-random variation in loan repayment incentives, so the individual first stage regressions, which describe the relationship between default and individual characteristics are of interest in their own right. The relationship between the week in the loan cycle and the fraction of repayers detailed in figure 3 is strongly increasing. Figure 10 shows a collapsed version of this relationship. More installments already paid by 3/9/2006 translates into a greater probability of full loan repayment following the default. The repayment probability ranges from 20% for those with the full loan to repay to 60% for those with very little time left on the loan. The slope of this line implies an approximately 1% increase in repayment probability per additional week in the loan cycle. It is curious that at 49 weeks, the repayment fraction is significantly lower than 100%. The borrower does have the option of repaying, taking a new, larger loan and subsequently defaulting. For some reason this behavior does not seem to happen, which might point to adverse peer effects.

Table 4 captures this relationship in a regression format. In all specifications, the effect of one extra week in the loan cycle at the time of the defaults corresponds to a 1% greater likelihood of repaying. In other words, individuals in week 50 are 50% more likely to repay their loans than borrowers in week 0 of their loan cycles. The standard errors of all of the estimates are extremely small. Note that the 1% coefficient is not very sensitive to the inclusion of branch fixed effects or village level peer group controls.

It is rare to have the opportunity to look at determinants of default/repayment in the microfinance setting, so Table 4 gives some important information about the effectiveness of the dynamic incentives in microfinance. Where it is significant, the weeks with the MFI variable is negative, even controlling for functions of loan size. Recall that this negative pattern appears in Figure 3. The column 5 estimates suggest that a borrower in loan cycle 2 is more than 4%-5% less likely to repay than a borrower in cycle 1 at the same week. This evidence suggests that the dynamic incentives lose their power as borrowers complete more cycles. Borrowers could become satiated with credit after having had previous opportunities to use microfinance loans for investment or durable purchases. Alternatively, the increases in loan sizes slow as borrowers mature in their relationships with the MFI and clients may no longer have strong enough future incentives. Perhaps, this deceleration is too much to keep the borrowers interested in repaying their loans. Rational theories of dynamic incentives such as Bulow and Rogoff (1979) predict that if the rate of increase of loan size is lower than the interest rate, rational borrowers should always eventually default on their loans.

Loan size does not appear to be correlated with repayment conditional on the weeks with the MFI for each borrower. We might expect a negative relationship between loan size and repayment if loan size was randomly assigned. However, Spandana lends according to it's beliefs about the borrower's ability to repay, so loan size could proxy for borrower wealth or other characteristics. A small coefficient might even suggest that Spandana chooses loan size optimally for its borrowers, at least in the first cycle.

5.3 Aggregate First Stage Results

Not surprisingly, the aggregate first stage regressions of peer repayment on average peer week in the loan cycle look quite similar to the individual first stage regressions. Tables 5 and 6 show variations on equations 4 and 5 respectively. The first column of each table shows the first stage village-ex-group peer repayment regressions. Columns 2 and 3 show the separate first stages for the village-ex-center and center-ex-group repayment rates. Finally, column 4 gives the center-ex-group stand-alone peer first stage with village fixed effects. Table 5 reinforces the 1% per week improvement in repayment as the weeks completed in the loan cycle increase at the time of default. Interestingly, there is no effect of village average weeks on center repayment.

Table 6 uses the fraction of peer members in the extremes of the loan cycle as instruments for peer repayment. We see that if the entire peer group goes from 0% in weeks 0-5 to 100% in weeks 0-5, the repayment likelihood decreases by 20-25%. The coefficient on the fraction in weeks 45-50 variable implies that if the entire group switches from 0% to 100% in weeks 45-50, the repayment likelihood increases by 20-25%. The sum of the coefficients on the 0-5 and 45-50 variables are indistinguishable from zero. This implies a symmetric effect of weeks in the loan cycle on repayment incentives. This is consistent with the linear shape of the individual incentives observed in figures 3 and 10. Again, notice that the village peer variables are not significant in the center regressions. This is a sort of peer effect reduced form regression, and provides early evidence that the village peer effect is not very strong. Note as in the individual first stage regression, the coefficients of interest in both tables are estimated quite precisely.

5.4 Reduced Form Results

Tables 7 and 8 display results for the reduced for regressions of individual repayment on the peer group weeks in the loan cycle variables. As in the first stage tables, table 7 presents the results using the average weeks instrument, while table 8 shows results for the extreme weeks instrument.[9] Column 1 presents the reduced form using the village-ex-group peers. The weeks variable is not significant in column 1. Failure to identify the village effect may be due to power, since there are only 600 villages in the data set, and the village average does not contain very much variation since

[9]The extreme weeks reduced form regressions are weak tests for non-linear or asymmetric peer effects.

individuals within each group were staggered. Column 2 includes village-ex-center and center-ex-group weeks variables. The coefficient on the center level peer group instrument is much larger than the coefficient on the village peer group variable and is of the same magnitude as the coefficient in column 1. The center-level weeks variable is significant. However, since the estimate on the village variable is so imprecise, we cannot conclude that the village and center peer effects are different, at least in the reduced form. Columns 3 and 4 show results for the reduced form only using the center-level variable. Column 4 uses village fixed effects. The coefficients are both significant and are of a similar magnitude. These reduced form regressions give some evidence that the peer effect is stronger at the more local level.

Again table 8 shows a similar pattern. Only the center level variables are statistically different from zero. We also see that only the weeks 45-50 variables are significant at any standard level. However, in all of the specifications, it cannot be ruled out that the center level coefficients sum to zero, so there is no conclusive evidence here of asymmetric effects. Finally, column 5 of table 8 shows the center level regression with the average-weeks and extreme weeks instrument. On their own, none of these variables is significant. Again, there is no strong evidence of asymmetries here.

5.5 IV Results

The results of the Two Stage Least Squares estimation procedure on the full sample are detailed in table 9. Parameter estimates of equation 3 are shown using the average weeks instrument in columns 1-3 and the fraction extreme weeks instruments in columns 4-6. The average weeks specifications again give village repayment peer effects (ex group) estimates at around 11.2%. This translates into a 1% increase in the probability of repaying the loan for every additional 10% repayment by the peer group. However, as the reduced form regression suggested, this estimate is not statistically significant. Column 2 breaks this effect into a center and a village-ex-center effect. The center effect is positive and significant at 0.0936. The remaining village effect is much smaller in magnitude, but is estimated with very large standard errors. Looking only at the center peer effect in column 3, the estimate is 11%. I include the linear component of the individual and village controls as well as the weeks in cycle variable. The identification strategy requires that the own repayment incentive through weeks in the cycle should not be affected by the inclusion of the peer weeks in the cycle variables controlling for functions of weeks with the MFI and loan size. The coefficients on weeks in cycle are all precisely identified at approximately 0.01, the same value as in the first stage regressions, where the peer repayment incentives were suppressed. Again, repayment is decreasing as individuals develop longer relationships with the MFI and loan size does not seem to matter. The village level controls for average length of relationship and average loan size appear to be negative.

Columns 4-6, which use an alternate function of the peer group week variables as an instrument for peer repayment, show very similar patterns. Again, none of the village peer effect coefficients are significant, but both of the center coefficients are significant at the 1% level. The magnitude of the center peer effect is larger at 13.7% in the alternate specifications.

As discussed in section 4.2, there is a close connection between instrumental variables and fuzzy regression discontinuity approaches. Table 10 shows IV results using the alternate instruments, while restricting the sample to those peer groups with a majority of borrowers in either the 0-5 or the 45-50 week buckets. The regressions include the full set of individual and peer controls. The specifications that include the village peer effect restict the entire village peer group, while the center-only regressions restict the sample at the center level. (i.e. in the center regressions, a subsample of every village is potentially used in the analysis.) This sample restriction is similar to the approach used by Angrist and Lavy (1999). Columns 1-3 show the village and center peer effects regressions. Note that in the specifications that include village repayment, the sample size drops by a factor of 10. The center-level regressions maintain closer to 20% of the original sample. Again, no village effect is evident, but in column 2, we can reject the hypothesis that the village and center peer effects are the same size. In this specification, the center peer effect is close to 25%. Column 3 only looks at the center effect. The coefficient on peer repayment is 0.128, which is very similar to the estimates in table 9. Column 4 of table 10 shows the center level regression limited to peer centers with more that 75% of borrowers in the extreme buckets. The results are almost identical to those of column 3.

Finally, table 11 presents the regression specification with the fuzzy RD interpretation as suggested by Imbens and Lemieux (2008). Columns 1-3 all express different specifications of the IV procedure which uses equation 6 for the first stage. The threshhold values increase as the columns move to the right. Again the coefficients on peer repay fall between 0.09 and 0.13. The coefficient is no longer significant when we restrict to centers that have more than 85% of borrowers grouped into one of the extreme buckets. Many of the centers all disbursed loans on the same week, but the variation I use comes from centers that had staggered disbursement schedules across groups. This might explain why the standard errors grow so fast as the threshhold is increased in table 11. If the centers that have staggered group disbursements are different from centers that don't have staggered disbursements, then the coefficients might be biased. However, the most plausible direction of this bias would be toward zero, since we might expect more synchronized centers to be characterized by closer relationships among its members. The results also look very similar when I drop all of the covariates except for own week in loan cycle.

5.6 Non-Linear Peer Effects

A linear peer effect does predict hump-shaped adoption/repayment patterns. However, is a non-linear peer effect partially driving the stark shape of repayment observed in figure 9? Non-linear peer effects might help to explain why there is so much aggregation towards the poles of full repayment and full default. I am interested in estimating

$$repay_{i,p(i)} = \beta^0 + g\left(\sum_{j \in p(i)} \frac{repay_j}{\|p(i)\|}\right) + \beta^2 weeks_i + \beta^3 X_{i,p(i)} + \varepsilon_{i,p(i)} \tag{7}$$

where the function $g(.)$ is not necessarily linear. Using the nonparametric IV, control function approach of Newey, Powell and Vella (1999), I plot the non-linear center peer effect in figure 11. The estimates use series regression with fith order polynomials. The shape of the estimated peer repayment relationship is not linear. It is characterized by steep regions around 0 and 1 and a much shallower slope between 0.2 and 0.8. However, the error bars are quite large. It is practically only possible to conclude that $g(1) > g(0)$. However, an S-shaped peer effect would be consistent with a coordination equilibrium-type story. Peer groups may coordinate to either all pay or all default. It is also interesting how symmetric this relationship appears to be, with an inflection point around 50% peer repayment. A relationship with this shape would also imply both virtuous and perverse peer effects, where the forces pulling toward default are similar to those pulling toward repay. While the resulting function is by no means conclusive, it is suggestive to oberve an S-shaped peer effect.

6 Conclusion

Several conclusions emerge from the analysis. First, the microfinance system that boasts near perfect repayment rates in good times can be very fragile in response to crises. Spandana, which was in the best possible position to make collections, had only moderate success over a three year time horizon. Second, the dynamic incentives and the ability for an MFI to continuously offer new loan disbursements are extremely important in encouraging repayment. However, the promise of future credit is not enough to encourage most borrowers early in their loan cycles to full repay their loans. It also appears that the power of these dynamic incentives decreases across the loan cycles.

I find strong evidence that repayment peer effects do exist. The decision of a peer group to move from 0% to 100% repayment corresponds to an individual becoming 10% more likely to repay her own loan. There is also some preliminary evidence that the center peer effect is stronger than the village peer effect. This is consistent with a story of local relationships being more influential to an individual's decision-making. I find no evidence that the peer effect is asymmetric, and it appears that both perverse and virtuous incentives arise from the peer effects.

The results are generally consistent with peer groups playing coordination equilibria. However, more work is necessary to fully unpack the mechanisms underlying these outcomes. Possible mechanisms could be that repayment sends a signal about credit-worthiness in informal financial transactions, information about the likelihood of recieving new loans diffuses through peer groups, groups coordinate on default to prevent collections agents from bothering a community or that groups believe that the MFI will cease to lend to a community unless a threshhold group repays.

A The Reflection Problem

Suppose that all peer groups are of size n and that the peer effect operates through the average repayment in the peer group. Then we are interesting in identifying the structural parameter, α_2. Note that the problem is symmetric for all individuals in the same peer group, so

$$\begin{aligned}
repay_1 &= \alpha_0 + \alpha_1 date_1 + \alpha_2 \sum_{j\neq 1} \frac{repay_j}{n-1} + \varepsilon_1 \\
repay_2 &= \alpha_0 + \alpha_1 date_2 + \alpha_2 \sum_{j\neq 2} \frac{repay_j}{n-1} + \varepsilon_2 \\
&\quad \ldots \\
repay_n &= \alpha_0 + \alpha_1 date_n + \alpha_2 \sum_{j\neq n} \frac{repay_j}{n-1} + \varepsilon_n
\end{aligned}$$

So, first sum equations 2-n

$$\sum_{j\neq 1} \frac{repay_j}{n-1} = \frac{n}{n-1}\alpha_0 + \sum_{j\neq 1} \frac{date_j}{n-1} + \alpha_2 \frac{1}{n-1} \sum_{i=2}^{n} \sum_{j\neq i} \frac{repay_j}{n-1} + \sum_{j\neq 1} \frac{\varepsilon_j}{n-1}$$

where

$$\begin{aligned}
&\alpha_2 \frac{1}{(n-1)^2} \sum_{i=2}^{n} \sum_{j\neq i} repay_j \\
&= \alpha_2 \left[\frac{repay_1}{(n-1)} + \frac{(n-2)}{(n-1)} \sum_{j\neq 1} \frac{repay_j}{(n-1)} \right]
\end{aligned}$$

So

$$\sum_{j\neq 1} \frac{repay_j}{n-1} = \frac{1}{1-\alpha_2 \frac{(n-2)}{(n-1)}} \left(\frac{n}{n-1}\alpha_0 + \alpha_1 \sum_{j\neq 1} \frac{date_j}{n-1} + \alpha_2 \frac{repay_1}{(n-1)} + \sum_{j\neq 1} \frac{\varepsilon_j}{n-1} \right)$$

Plugging this back into the first equation, we get

$$\begin{aligned}
repay_1 &= \alpha_0 + \alpha_1 date_1 + \frac{\alpha_2}{1-\alpha_2 \frac{(n-2)}{(n-1)}} \left(\frac{n}{n-1}\alpha_0 + \alpha_1 \sum_{j\neq 1} \frac{date_j}{n-1} + \alpha_2 \frac{repay_1}{(n-1)} + \sum_{j\neq 1} \frac{\varepsilon_j}{n-1} \right) + \varepsilon_1 \\
&= \tilde{\alpha}_0 + \frac{1}{\left(1 - \frac{\alpha_1 \alpha_2}{(n-1)-\alpha_2(n-2)}\right)} \left(\alpha_1 date_1 + \frac{\alpha_1 \alpha_2}{1-\alpha_2 \frac{(n-2)}{(n-1)}} \sum_{j\neq 1} \frac{date_j}{n-1} \right) + \tilde{\varepsilon}
\end{aligned}$$

Now, let's go back and look at the average peer repayment equation, since this is in essence, the first stage of my regressions

$$\begin{aligned}
\sum_{j\neq 1}\frac{repay_j}{n-1} &= \frac{1}{1-\alpha_2\frac{(n-2)}{(n-1)}}\left(\frac{n}{n-1}\alpha_0+\alpha_1\sum_{j\neq 1}\frac{date_j}{n-1}+\alpha_2\frac{repay_1}{(n-1)}+\sum_{j\neq 1}\frac{\varepsilon_j}{n-1}\right)\\
&= \phi+\frac{1}{1-\alpha_2\frac{(n-2)}{(n-1)}}\left(\alpha_1+\frac{\alpha_2^2}{n-1}\right)\sum_{j\neq 1}\frac{date_j}{n-1}
\end{aligned}$$

where ϕ includes all of the other terms. So the coefficient on average date in this regression (excluding $date_1$ which is orthogonal to the other date variable) is

$$\frac{1}{1-\alpha_2\frac{(n-2)}{(n-1)}}\left(\alpha_1+\frac{\alpha_2^2}{n-1}\right)$$

So the ratio of the reduced form coefficient over the first stage coefficient is what IV gives us, so

$$\begin{aligned}
\frac{\frac{\alpha_1\alpha_2}{1-\alpha_2\frac{(n-2)}{(n-1)}}}{\frac{1}{1-\alpha_2\frac{(n-2)}{(n-1)}}\left(\alpha_1+\frac{\alpha_2^2}{n-1}\right)} &= \frac{\alpha_1\alpha_2}{\alpha_1+\frac{\alpha_2^2}{n-1}}\\
&= \frac{\alpha_2}{1+\frac{\alpha_2^2}{\alpha_1(n-1)}}\\
&\approx \alpha_2
\end{aligned}$$

If anything, the small sample bias makes this estimate too low. IV gives a consistent estimate of the peer effect.

References

[1] Acemoglu, Daron and Josh Angrist (1990): "How Large are the Social Returns to Education? Evidence from Compulsory Attendance Laws", *NBER Macroeconomics Annual.* 15.

[2] Ahlin, Christian and Robert Townsend (2002): "Using Repayment Data to Test Across Models of Joint Liability Lending", *University of Chicago.*

[3] Angrist, Joshua and Victor Lavy (1999): "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement", *Quarterly Journal of Economics.* 114(2): 533-575.

[4] Banerjee, Abhijit, Timothy Besley and Timoty Guinnane (1994): "Thy Neighbor's Keeper: The Design of a Credit Cooperative with Theory and a Test", *Quarterly Journal of Economics.* May: 491-515.

[5] Besley, Timothy and Stephen Coate (1995): "Group lending, repayment incentives and social collateral", *Jounal of Development Economics.* 46: 1-18.

[6] Bond, Philip and Ashok Rai (2009): "Borrower Runs", *Jounal of Development Economics.* 88(2): 185-191

[7] Brown, Martin and Christian Zehnder (2007): "Credit Reporting, Relationship Banking and Loan Repayment", *Journal of Money, Credit and Banking.* 39(8): 1883-1918.

[8] Camerer, Colin (2003): "Behavioural Studies of Strategic Thinking in Games", *Trends in Cognitive Science.* 7(5): 225-231.

[9] Conley, Timothy and Chris Udry (2010): "Learning about a New Technology: Pineapple in Ghana", *American Economic Review.* Forthcoming.

[10] Duflo Esther (2004): "The medium run effects of educational expansion: evidence from a large school construction program in Indonesia", *Journal of Development Economics.* 74 (1): 163-197.

[11] Duflo, Esther and Emmanuel Saez (2002): "Participation and Investment Decisions in a Retirement Savings Plan: the Influence of Colleagues' Choices", *Journal of Public Economics.* 85 (1): 121-148.

[12] Duflo, Esther and Emmanuel Saez (2003): "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment", *Quarterly Journal of Economics.* 118 (3): 815-842.

[13] Feigenberg, Benjamin, Erica Field and Rohini Pande (2010): "Building Social Capital through Microfinance", *Working Paper.*

[14] Foster, Andrew and Mark Rosenzweig (1995): "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture", *Journal of Political Economy.* 103 (6): 1176-1209.

[15] Ghatak, Maitreesh (2000): "Screening by the company you keep: Joint liability lending and the peer selection effect", *The Economic Journal.* 110 (July): 601-631.

[16] Gine, Xavier and Dean Karlan (2007): "Group versus Individual Liability: a Field Experiment in the Philippines", *Working Paper 111 Center for Global Development.*

[17] Gine, Xavier and Dean Karlan (2008): "Peer Monitoring and Enforcement: Long Term Evidence from Microcredit Lending Groups with and without Group Liability", *Working Paper.*

[18] Hahn, Jinyong, Petra Todd and Wilbur Van Der Klauuw (2001): "Identification and estimation of treatment effects iwth a regression discontinuity design", *Econometrica.* 69:201-209.

[19] Imbens, Guido and Thomas Lemieux (2008): "Regression dicontinuity designs: A guide to practice", *Journal of Econometrics.* 142: 615-635.

[20] Karlan, Dean (2005): "Using Experimental Economics to Measure Social Capital and Predict Financial Decisions", *American Economic Review.* 95 (5): 1688-1699.

[21] Karlan, Dean (2007): "Social Connections and Group Banking," *The Economic Journal.* 117: F52-F84.

[22] Kinnan, Cynthia and Robert Townsend (2010): "Kinship Networks, Financial Access and Consumption Smoothing," *Working Paper.*

[23] Manski, Charles (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies.* 60(3): 531-542.

[24] Microfinance Information Exchange (MIX). Available at www.mixmarket.org.

[25] Newey, Whitney and James Powell (2003): "Instrumental Variable Estimation of Nonparametric Models", *Econometrica.* 71(5): 1565-1578.

[26] Newey, Whitney, James Powell and Francis Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models", *Econometrica.* 67(3): 565-603.

[27] Karlan, Dean, Markus Mobius, Tanya Rosenblat and Adam Szeidl (2009): "Trust and Social Collateral", *Quarterly Journal of Economics.* 124 (3): 1307-1361.

[28] Van Der Klaauw, Wilbur (2002): "Estimating the effect of financial aid offers on college enrollment: a regression-discontinuity approach", *International Economic Review.* 43: 1249-1287.

B Tables and Figures

	Full Sample		Analysis Sample	
Descriptive Statistics	Mean	Std. Dev	Mean	Std. Dev
As of 3/9/2006				
Number of Loans	194,312		124,488	
Number of Borrowers	162,835		124,488	
Fraction of Borrowers with Multiple Loans	19.14%			
Loan Size (Rs)	6,462	2,549	7,515	1,947
Loan Outstanding (Rs)	2,975	2,894	3,492	3,099
Date of Disbursement	8/20/2005	104 days	8/20/2005	111 days
Number of Villages/Slums	1,266		668	
Number of Centers	7,347		5,538	
Number of Groups	18,499		13,876	
Number of Borrowers in 1st Loan Cycle			68,696	
Number of Borrowers in 2nd Loan Cycle			49,052	
Number of Borrowers in 3nd Loan Cycle			6,740	
As of 11/20/2009				
Number of Loans Still in Arrears	106,386		70,448	
Loan Outstanding Given Not Repaid (Rs)	3,346	3,059	4,502	3,312

Table 1: Summary Statistics of the Data Universe and the Subset Used in the Analysis

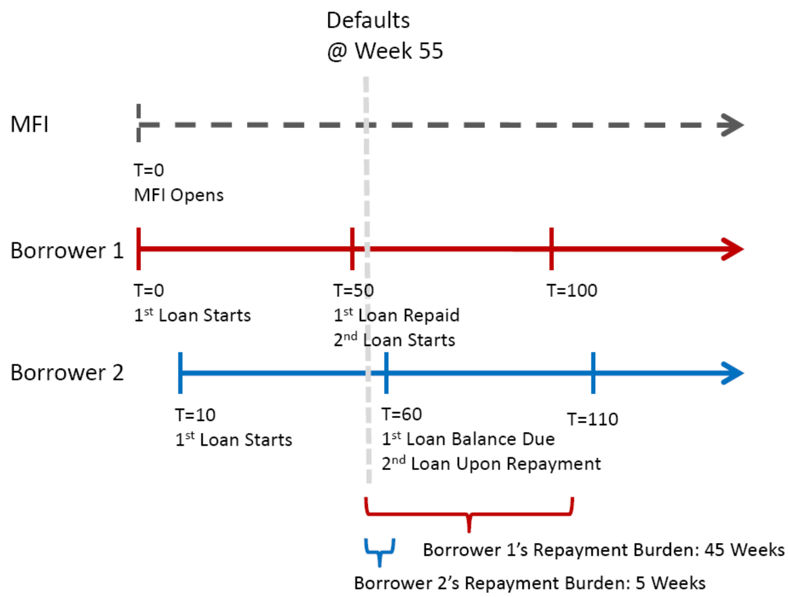

Figure 1: Differential Repayment Incentives Across the Loan Cycle

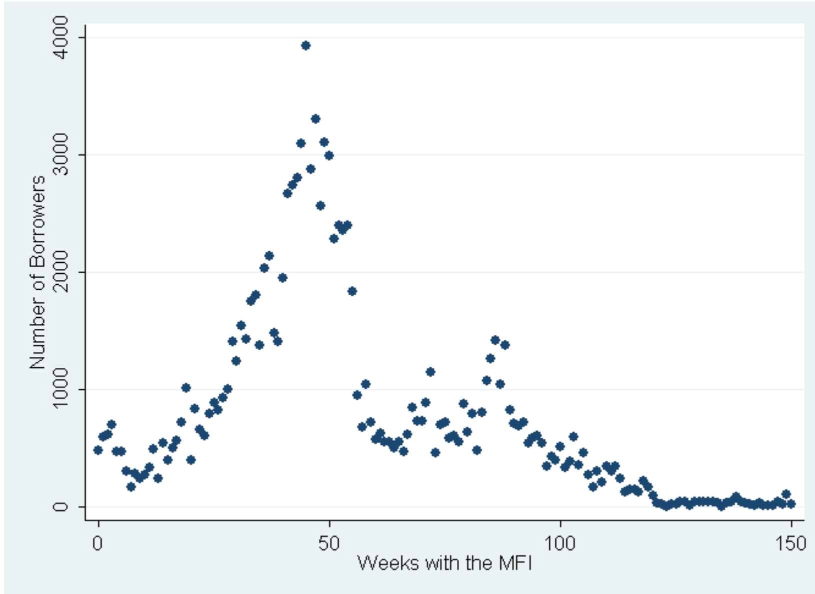

Figure 2: Number of Borrowers by Week with the MFI

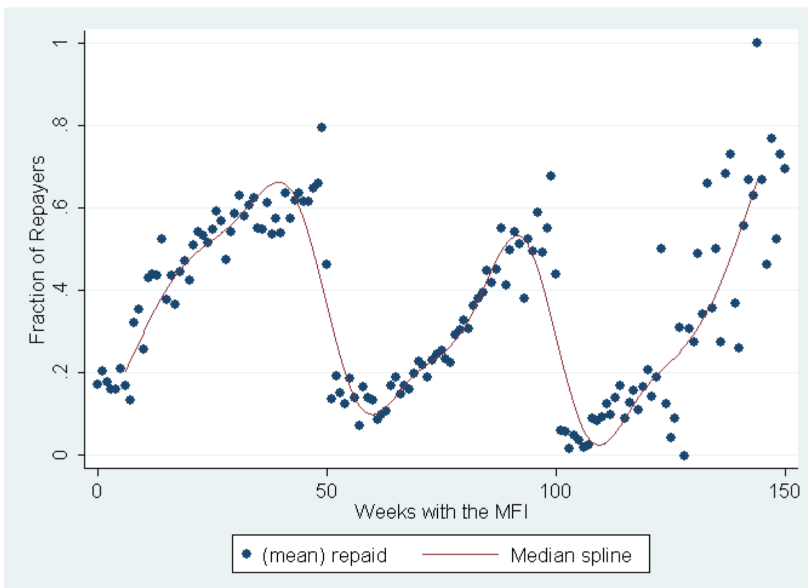

Figure 3: Fraction of Repayers by Weeks with the MFI

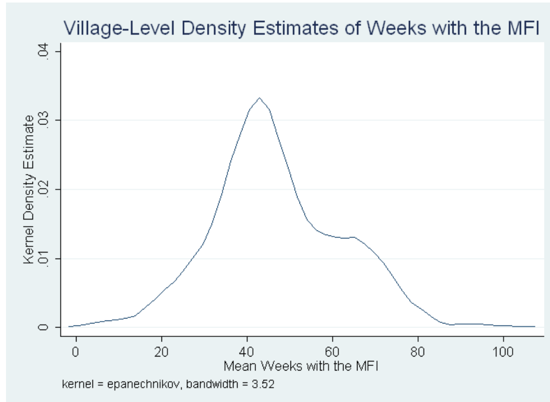

Figure 4: Density of Weeks with the MFI at the Average Village Level

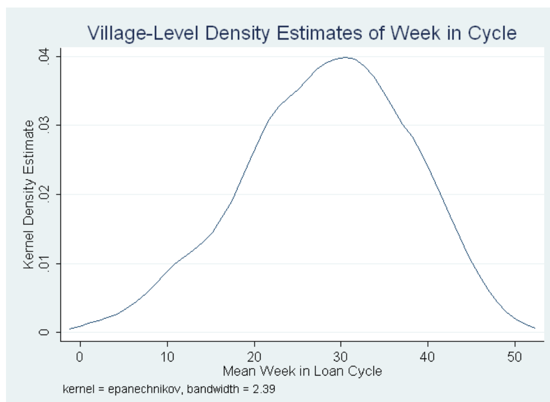

Figure 5: Density of Weeks in the Loan Cycle at the Average Village Level

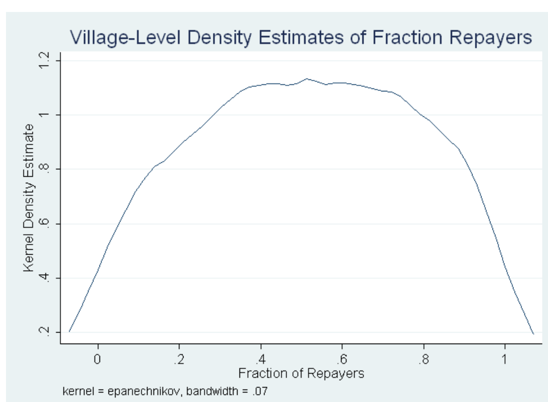

Figure 6: Density of Fraction of Repayers at the Village Level

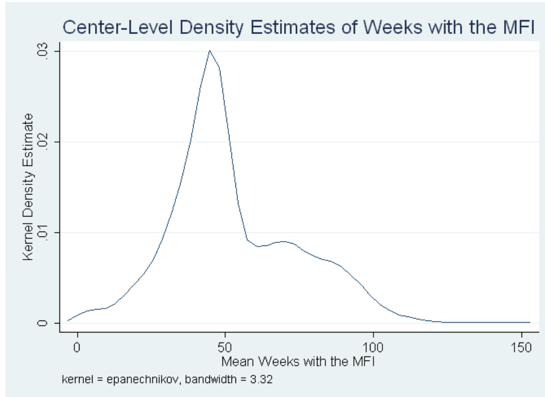

Figure 7: Density of Weeks with the MFI at the Average Center Level

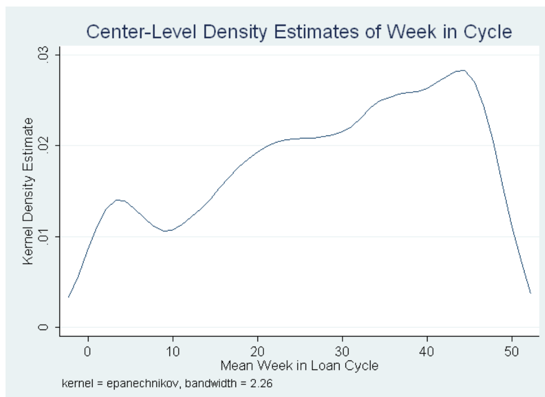

Figure 8: Density of Weeks in the Loan Cycle at the Average Center Level

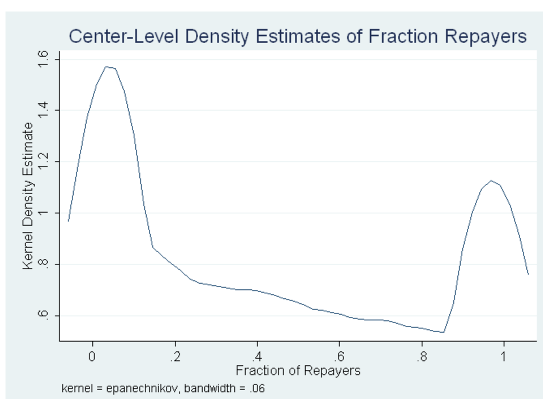

Figure 9: Density of Fraction of Repayers at the Village Level

	Full Sample	Full Sample		Maj. 0-5 and 45-50
	Week in Cycle	Fraction in Weeks 0-5	Fraction in Weeks 45-50	Indicator for Maj. Weeks 0-5
Population	-56.86**	2167	-1100	3423***
	(25.87)	(1361)	(1066)	(1028)
Cultivation Area per Capita	0.00195	-0.0380	0.0474	-0.0923
	(0.00125)	(0.0659)	(0.0516)	(0.119)
Irrigated Area per Capita	0.00189	-0.0919	0.0304	-0.137
	(0.00121)	(0.0638)	(0.0500)	(0.121)
Distance to Town	-0.152*	14.72***	8.199**	1.319
	(0.0911)	(4.694)	(3.678)	(5.681)
Education Facilities	0.000551	0.0128	0.0123	
	(0.000366)	(0.0193)	(0.0151)	
Primary Schools per Capita	1.26e-05**	-0.000101	0.000499**	-0.000562
	(5.65e-06)	(0.000297)	(0.000233)	(0.000597)
Medical Facilities	-0.00148	0.0698	-0.0491	0.210
	(0.00225)	(0.118)	(0.0927)	(0.140)
Health Centers per Capita	3.18e-07	1.45e-05	-7.65e-06	3.62e-05***
	(3.35e-07)	(1.76e-05)	(1.38e-05)	(1.29e-05)
Health SubCenters per Capita	8.62e-07	4.47e-05	4.48e-05	3.84e-05
	(1.16e-06)	(6.10e-05)	(4.78e-05)	(6.62e-05)
Number of Banks per Capita	-8.99e-07	1.24e-05	-2.03e-05	1.24e-05
	(6.31e-07)	(3.32e-05)	(2.60e-05)	(2.91e-05)
Railway	-0.000363	-0.0753	-0.0233	
	(0.00129)	(0.0678)	(0.0531)	
Paved Roads	0.000324	0.0561	0.0440	0.0312
	(0.00138)	(0.0726)	(0.0569)	(0.0459)

Table 2: Week in loan cycle vs. Village Census Characteristics

Table 3: OLS Results - Individual Repayment on Peer Repayment

Individual Repayment	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
Village Repayment ex Group	0.941***		0.816***			
	(0.00461)		(0.0216)			
Village Repayment ex Center		0.326***		0.265***	0.256***	0.273***
		(0.0151)		(0.0202)	(0.0206)	(0.0221)
Center Repayment ex Group		0.629***		0.601***	0.485***	0.503***
		(0.0150)		(0.0154)	(0.0169)	(0.0165)
Individual Controls	No	No	Yes	Yes	Yes	Yes
Peer Group Controls	No	No	No	No	No	Yes
Fixed Effects	No	No	Yes	Yes	Yes	Yes
Observations	115636	108381	115636	108381	108381	108381
R^2	0.231	0.323	0.256	0.341	0.386	0.391

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

Clustered standard errors at the village level in parentheses

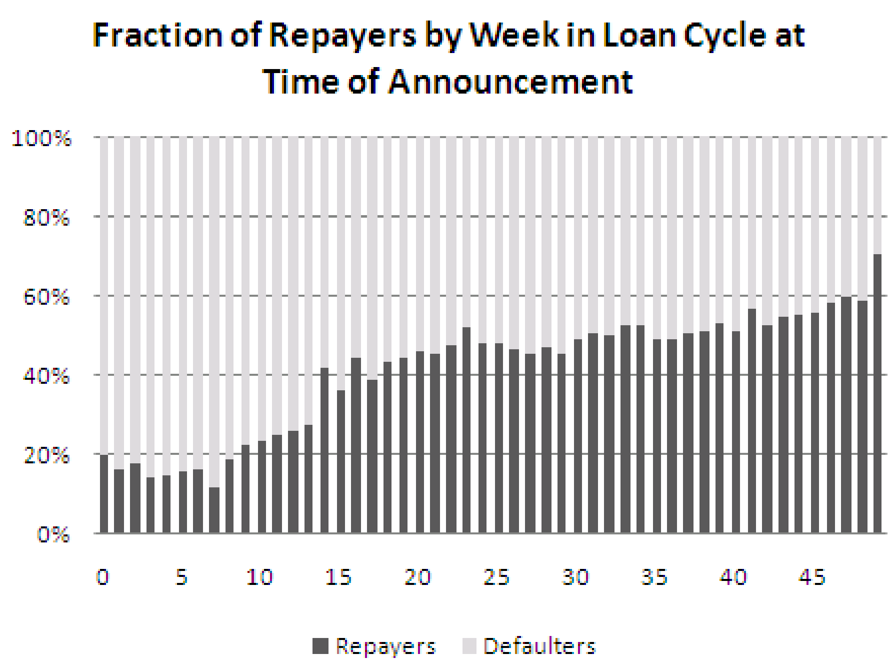

Figure 10: Collapsed Relationship Between Week in the Loan Cycle and Repayment Rate

Table 4: Individual First Stage Results

Individual Repayment	(1)	(2)	(3)	(4)	(5)
Weeks in Cycle	0.0118***	0.0110***	0.0106***	0.0106***	0.0107***
	(0.000430)	(0.000436)	(0.000412)	(0.000392)	(0.000379)
Weeks with MFI	-0.00264***	-4.66e-05	2.67e-05	-0.000874*	-0.00130***
	(0.000255)	(0.000147)	(0.000206)	(0.000478)	(0.000414)
Weeks with MFI^2				8.72e-06**	1.26e-05***
				(3.98e-06)	(3.57e-06)
Loan Size	-5.31e-06*	1.84e-06	2.21e-06	-8.54e-06	-0.0102
	(2.81e-06)	(2.32e-06)	(2.53e-06)	(8.85e-06)	(0.00875)
Loan Size^2				7.02e-10	0.000677
				(5.26e-10)	(0.000509)
Village Peer Controls	No	Yes	No	Yes	Yes
Fixed Effects	No	No	Branch	Branch	Village
Observations	115636	115636	115636	115636	115636
R^2	0.168	0.212	0.281	0.288	0.347

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

Clustered standard errors at village level in parentheses

Table 5: Aggregate First Stage Results - Average Weeks

	(1)	(2)	(3)	(4)
	Village	Village	Center	Center
Weeks in cycle (Village ex Group)	0.0110***			
	(0.00106)			
Weeks in cycle (Village ex Center)		0.0112***	-7.11e-06	
		(0.000940)	(0.000966)	
Weeks in cycle (Center ex Group)		-0.000114	0.0110***	0.0107***
		(0.000192)	(0.000414)	(0.000468)
Peer and Individual Controls	Yes	Yes	Yes	Yes
Fixed Effects	Branch	Branch	Branch	Village
Constant	2.214***	1.996***	1.827***	0.185*
	(0.392)	(0.360)	(0.371)	(0.100)
Observations	115636	108381	108381	108381
R^2	0.670	0.654	0.486	0.661

Clustered standard errors at the village level in parentheses.

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

Table 6: Aggregate First Stage Results - Extreme Weeks

	(1)	(2)	(3)	(4)
	Village	Village	Center	Center
Weeks 0-5 Fraction (Village ex Group)	-0.271***			
	(0.0506)			
Weeks 45-50 Fraction (Village ex Group)	0.235***			
	(0.0446)			
Weeks 0-5 Fraction (Village ex Center)		-0.272***	-0.0675	
		(0.0431)	(0.0457)	
Weeks 45-50 Fraction (Village ex Center)		0.262***	0.0418	
		(0.0381)	(0.0398)	
Weeks 0-5 Fraction (Center ex Group)		-0.00555	-0.237***	-0.214***
		(0.00753)	(0.0152)	(0.0153)
Weeks 45-50 Fraction (Center ex Group)		-0.000402	0.210***	0.197***
		(0.00690)	(0.0168)	(0.0180)
Peer and Individual Controls	Yes	Yes	Yes	Yes
Fixed Effects	Branch	Branch	Branch	Village
Constant	2.546***	2.626***	2.467***	0.541***
	(0.405)	(0.361)	(0.370)	(0.107)
Observations	115636	108381	108381	108381
R^2	0.646	0.621	0.438	0.632

Clustered standard errors at the village level in parentheses.

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

Table 7: Reduced Form Results - Average Weeks Instruments

	(1)	(2)	(3)	(4)
Weeks in cycle (Village ex Group)	0.00124			
	(0.00112)			
Weeks in cycle (Village ex Center)		0.000299		
		(0.00102)		
Weeks in cycle (Center ex Group)		0.00100***	0.00119***	0.000768**
		(0.000334)	(0.000387)	(0.000338)
Individual and Peer Controls	Yes	Yes	Yes	Yes
Fixed Effects	Branch	Branch	Branch	Village
Observations	115636	108381	108381	108381
R^2	0.289	0.292	0.286	0.401

Clustered standard errors at the village level in parentheses

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

Table 8: Reduced Form Results - Extreme Weeks Instruments

	(1)	(2)	(3)	(4)	(5)
Weeks in cycle (Center ex Group)					0.00101
					(0.000614)
Weeks 0-5 Fraction (Village ex Group)	-0.0464				
	(0.0529)				
Weeks 45-50 Fraction (Village ex Group)	0.0316				
	(0.0465)				
Weeks 0-5 Fraction (Village ex Center)		-0.0337			
		(0.0471)			
Weeks 45-50 Fraction (Village ex Center)		-0.00356			
		(0.0419)			
Weeks 0-5 Fraction (Center ex Group)		-0.00649	-0.0159	-0.000599	0.00346
		(0.0138)	(0.0164)	(0.0128)	(0.0219)
Weeks 45-50 Fraction (Center ex Group)		0.0346**	0.0298**	0.0374**	0.0176
		(0.0136)	(0.0150)	(0.0145)	(0.0171)
Individual and Peer Controls	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Branch	Branch	Branch	Village	Branch
Observations	115636	108381	108381	108381	108381
R^2	0.289	0.292	0.286	0.402	0.286

Clustered standard errors at the village level in parentheses

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

Table 9: Instrumental Variables Results - Full Sample

	(1)	(2)	(3)	(4)	(5)	(6)
Instruments	Average	Average	Average	Extremes	Extremes	Extremes
Village Repaid (ex G)	0.112			0.162		
	(0.0928)			(0.111)		
Village Repaid (ex C)		0.0273			0.0464	
		(0.0814)			(0.0949)	
Center Repaid (ex G)		0.0936***	0.110***		0.137***	0.137***
		(0.0303)	(0.0339)		(0.0481)	(0.0497)
Week in Cycle	0.0104***	0.00997***	0.0101***	0.0103***	0.00969***	0.00989***
	(0.000379)	(0.000378)	(0.000379)	(0.000410)	(0.000457)	(0.000454)
Weeks w/ MFI	-0.000845*	-0.000524	-0.000753*	-0.000809*	-0.000426	-0.000685
	(0.000449)	(0.000385)	(0.000456)	(0.000449)	(0.000403)	(0.000470)
Weeks w/ MFI Sq.	8.76e-06**	7.30e-06**	8.76e-06**	8.55e-06**	6.82e-06**	8.42e-06**
	(3.79e-06)	(3.32e-06)	(3.97e-06)	(3.77e-06)	(3.40e-06)	(4.03e-06)
Loan Size (1000s)	-0.0108	-0.00653	-0.00923	-0.0115	-0.00747	-0.00962
	(0.00854)	(0.00747)	(0.00760)	(0.00859)	(0.00754)	(0.00760)
Loan Size Sq.	0.000781	0.000443	0.000618	0.000808	0.000438	0.000606
	(0.000499)	(0.000453)	(0.000464)	(0.000499)	(0.000451)	(0.000461)
Weeks w/ MFI (VexC)	-0.00654*			-0.00598		
	(0.00363)			(0.00364)		
Loan Size (VexG)	-0.506***			-0.487***		
	(0.104)			(0.105)		
Weeks w/ MFI (VexC)		-0.00447			-0.00393	
		(0.00322)			(0.00316)	
Loan Size (VexC)		-0.403***			-0.383***	
		(0.0888)			(0.0893)	
Weeks w/ MFI (CexG)		-0.00191**	-0.00250***		-0.00197**	-0.00251***
		(0.000843)	(0.000966)		(0.000818)	(0.000942)
Loan Size (CexG)		-0.0166	-0.0347*		-0.0146	-0.0328*
		(0.0162)	(0.0187)		(0.0159)	(0.0187)
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Branch	Branch	Branch	Branch	Branch	Branch
Observations	115636	108381	108381	115636	108381	108381
R^2	0.303	0.323	0.320	0.309	0.337	0.327

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

Clustered standard errors at the village level in parentheses

Table 10: Instrumental Variables Results - Restricted Sample

	(1) 0.5 Cutoff	(2) 0.5 Cutoff	(3) 0.5 Cutoff	(4) 0.75 Cutoff
Village Repaid (ex G)	0.117			
	(0.128)			
Village Repaid (ex C)		-0.0633		
		(0.101)		
Center Repaid (ex G)		0.251***	0.128**	0.124**
		(0.0891)	(0.0527)	(0.0599)
Week in Cycle	0.0106***	0.00947***	0.0103***	0.0107***
	(0.000880)	(0.00107)	(0.000565)	(0.000643)
Observations	11469	10409	30836	21986
R^2	0.305	0.357	0.398	0.407

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

Clustered standard errors at the village level in parentheses

Table 11: Fuzzy RD Interpretation

	(1) 0.75	(2) 0.80	(3) 0.85
Repaid (CexG)	0.129**	0.102*	0.0992
	(0.0605)	(0.0610)	(0.0636)
Week in Cycle	0.0106***	0.0110***	0.0111***
	(0.000643)	(0.000666)	(0.000687)
Controls	0.175	0.185	-0.274
Fixed Effects	Branch	Branch	Branch
Observations	21986	21391	20666
R^2	0.408	0.401	0.399

Clustered standard errors at the village level in parentheses

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

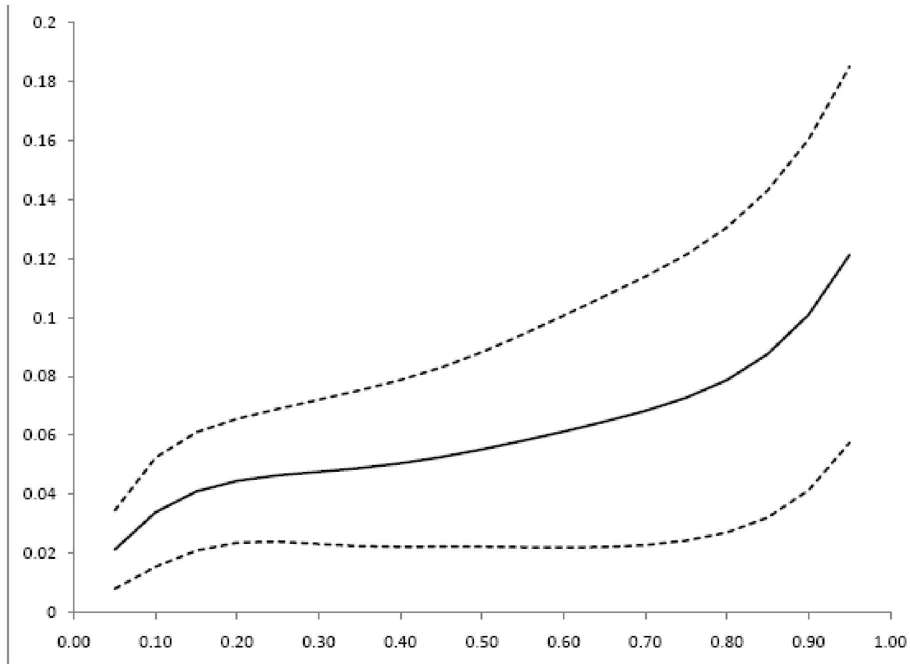

Figure 11: Individual repayment as a non-parametric function of center-level repayment with 95% confidence band.