NBER WORKING PAPER SERIES

GROUP LENDING WITH HETEROGENEOUS TYPES

Li Gan
Manuel A. Hernandez
Yanyan Liu

Working Paper 18847
http://www.nber.org/papers/w18847

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2013

We thank Alan de Brauw, Arun Chandrasekhar, Carlos Martins-Filho, Eduardo Nakasone, Annabel Vanroose, Ruth Vargas-Hill and seminar participants at the Winter Meetings of the Econometric Society and IFPRI for their helpful comments. We gratefully acknowledge financial support from the CGIAR Research Program on Policies, Institutions and Markets. We also thank the staff of the Center for Economics and Social Studies, particularly Prof. S. Galab, for their support and collaboration in making the data available. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Group Lending with Heterogeneous Types
Li Gan, Manuel A. Hernandez, and Yanyan Liu
NBER Working Paper No. 18847
February 2013
JEL No. C35,O16

ABSTRACT

Group lending has been widely adopted in the past thirty years by many microfinance institutions as a means to mitigate information asymmetries when delivering credit to the poor. This paper proposes an empirical method to address the potential omitted variable problem resulting from unobserved group types when modeling the repayment behavior of group members. We estimate the model using a rich dataset from a group lending program in India. The estimation results support our model specification and show the advantages of relying on a type-varying method when analyzing the probability of default of group members.

Li Gan
Department of Economics
Texas A&M University
College Station, TX 77843-4228
and NBER
gan@econmail.tamu.edu

Manuel A. Hernandez
Markets, Trade, and Institutions Division
IFPRI
Washington, DC 20006-1002
m.a.hernandez@cgiar.org

Yanyan Liu
Market, Trade, and Institutions Divisions
IFPRI
Washington, DC 20006-1002
y.liu@cgiar.org

1 Introduction

Since the establishment of Grameen Bank in Bangladesh in the mid-seventies, microfinance has boomed. As of December 2010, 3,652 microfinance institutions reported reaching over 205 million clients worldwide, and every two out of three borrowers were among the poorest when they took their first loan (Maes and Reed 2012). Such expansion can be partly attributed to the widely adopted practice of group lending in microfinance programs. In contrast to individual lending, group lending with joint liability grants a loan to a group of borrowers, and the whole group is liable for the debt of any individual member in the group.[2] This practice allows microfinance programs to rely mainly on information advantages among group members, rather than on financial collateral, to mitigate information asymmetries between lenders and potential borrowers. Given that the poor often lack appropriate financial collateral, group lending programs provide a feasible way of extending credit to poor people who are usually kept out of traditional banking systems.

Despite its rapid growth, there is an ongoing debate on whether group lending programs are sustainable and able to achieve and maintain sound repayment performance while serving poor borrowers, without the support of third parties such as international organizations. Armendariz and Morduch (2005) show, for example, that Grameen Bank has experienced losses close to eighteen percent of their outstanding loans over the period 1985-1996 after properly adjusting for their portfolio size. It is also often argued that the high transaction costs faced by micro finance institutions in identifying and screening their clients, processing applications and collecting repayments keep interest rates high and prevent them from reaching new clients and expanding their operations (Armendariz and Morduch 2004; Shankar 2006; Field and Pande 2008). Understanding the factors affecting repayment performance, which may vary by (unobserved) group types, are thus of great policy relevance. In particular, more accurate risk scoring tools can help to overcome information asymmetries by aiding lending institutions to better classify their potential clients and understand the factors driving their behavior, further promoting the development and sustainability of microcredit markets.

This paper contributes to the ongoing debate and to the literature by more explicitly dealing with the unobserved group heterogeneity. In particular, we make three contributions to the literature. First, the paper develops a basic framework with both peer selection and moral

[2] Joint liability is one of the most common varieties of group loan contracts.

hazard that shows how joint liability can lead to the coexistence of different group types, which implies the necessity to account for these group heterogeneities when modeling repayment behavior in group lending. Second, the paper proposes and applies an empirical model to explicitly deal with the problem of unobserved group heterogeneity. The paper discusses the identification and conducts a test on the specification of the empirical model proposed. Finally, the estimation results of the mixture model are more informative than standard probabilistic models about the potential factors driving repayment behavior, which may differ by group type, and the results are further shown to attain a higher predictive power.

In most group lending programs, individuals voluntarily form a group based on a set of common characteristics, which are generally observed by peers but not by lenders (and econometricians). This peer selection in the group formation process helps to lessen adverse selection as individuals screen each other when forming groups. On this matter, Ghatak (1999, 2000) and van Tassel (1999) show that in a context of individuals with heterogeneous risk types and asymmetric information (where borrowers know each other's type but lenders do not), group lending with joint liability will lead to the formation of relatively homogenous groups of either safe or risky borrowers.[3] The intuition behind is that while a borrower of any type prefers a safe partner because of lower expected joint-liability payments, safe borrowers value safe partners more than risky partners because they repay more often. This positive assortative matching is supported by empirical evidence in Ahlin (2009), who also finds that borrowers will anti-diversify risk within groups in order to lower their chances of facing liability for group members.

However, in a similar manner as self-selection, peer selection creates an omitted variable problem in the empirical literature on repayment behavior (Karlan 2007). The omitted variables may include, for example, the risk type, entrepreneurial spirit, economic opportunities, solidarity, reciprocity and trust among group members, which affect repayment performance and are likely correlated with the indicators generally used to account for group heterogeneity and social ties when modeling repayment behavior. Yet, different from the omitted variable problem due to

[3] In contrast, Armendariz de Aghion and Gollier (2000) suggest that non assortative matching equilibrium can exist in the case where a borrower knows her own type but has no ex-ante information about the other borrowers' types. Guttman (2008) indicate that negative assortative matching is possible if a riskier borrower can provide side-payments to get a safer peer. However, side-payments are usually infeasible when the group is relatively large. And group members often know each other well enough because groups are typically formed by people living in the same geographical area or in contiguous areas. In fact, the information advantage (local information) of group members over lenders is one of the main factors to justify the idea of group lending over individual lending.

self-selection, the omitted variable problem due to peer selection has largely been overlooked in the literature (Hermes and Lensink 2007). Most of the empirical studies that explore determinants of repayment in group lending programs treat the group as a decision maker and employ single-agent choice models to examine how different group characteristics, including proxies for social ties, affect the group repayment performance (e.g., Sharma and Zeller 1997; Zeller 1998; Wydick 1999; Paxton et al. 2000; Hermes et al. 2005; Ahlin and Townsend 2007; Cull et al. 2007).

In addition, groups may also differ in their effort levels and/or effectiveness of peer monitoring and peer pressure among members, which is also unobserved by lenders and have direct implications on the observed repayment performance of group members. Besides mitigating adverse selection through peer screening, group lending helps alleviate moral hazard behavior and enforce repayment because members can more closely monitor each other's use of loans and exert pressure to prevent deliberate default.[4] The success of peer monitoring and peer pressure efforts across groups may be further correlated with peer screening because individuals are more likely to select safe borrowers who are also less costly to monitor and less likely to deliberately default. Overall, group-level unobservables may result from a combination of factors, which include endogenous group formation due to ex-ante peer selection and ex-post peer monitoring and pressure efforts.

We propose and implement an empirical method to address the potential omitted variable problem in group lending resulting from unobserved types. We use a mixture model to explicitly account for unobserved group types when modeling the repayment behavior of group members. In the model, individuals make repayment decisions based on their unobserved group type as well as on observable individual and loan characteristics. Average member characteristics and other group and village characteristics help, in turn, to identify the group types. We further allow the marginal effects in the repayment equation to vary across types. We estimate the model using a rich dataset from a group lending program in Andhra Pradesh in India.[5] While the type-varying groups in the empirical model may be explained by peer selection and variations (if any) in peer efforts and the effectiveness of peer monitoring and enforcement rules, as well as by other

[4] See, e.g., Stiglitz (1990), Varian (1990), Banerjee et al. (1994), Armendariz de Aghion (1999) and Chowdury (2005) for theoretical models showing how group lending with joint liability may help solving moral hazard and monitoring problems.
[5] Group loans account for 93% of the microfinance in India (Shankar 2006).

unobserved factors like social cohesion, disentangling these effects is beyond the scope of the study.[6]

The estimation results support our model specification and show the advantages of relying on this method when analyzing the probability of default of group members. The model clearly distinguishes two group types: a first group type where members are more inclined to fulfill their credit obligations and a second group type where members are more inclined to default. We also provide evidence supporting that the group types are not simply identified by the functional form of the proposed model. We further find important differences in the marginal effects of the different individual and loan characteristics included in the repayment equation, which suggests that the underlying factors driving repayment behavior may differ across group types. In addition, the type-varying model shows a higher predictive performance than standard probabilistic models.

The remainder of the paper is organized as follows. Section 2 further discusses the implications of group lending with joint liability and heterogeneous types using a simple model of adverse selection and moral hazard. Section 3 describes in detail the group lending program considered for the study and the data. Section 4 presents the empirical model used to account for the potential omitted variable problem resulting from unobserved group types when modeling the repayment behavior of group members. Section 5 reports and discusses the estimation results. Section 6 concludes.

2 A simple model of group lending with peer selection and moral hazard

Ghatak (1999, 2000) and van Tassel (1999) develop models that describe how joint liability with heterogeneous types and local information can lead to positive assortative matching through peer selection. We extend Ghatak (1999) base model by taking into account both peer selection and moral hazard. In particular, we allow individuals to differ on their risk type (creditworthiness) and on their level of effort.

Assume borrowers are risk-neutral and endowed with one risky project, which requires one unit of capital. Individuals have no initial wealth and must borrow the required amount of

[6] For a formal evaluation of ex-post peer effects on individual repayment behavior, refer to Karlan (2007) and Li et al. (2012). Karlan (2007) exploits a unique quasi-random group formation process to isolate peer selection and examine the impact of monitoring and enforcement on repayment; Li et al. (2012) estimate a structural model that takes into account interactions across group members and incorporates group-level unobservables as random effects.

capital. Further assume that there are two types of borrowers: risky individuals of type a and safe individuals of type b.[7] The probability of success of borrower i's project (k_i) depends on her inherent probability of success ($p_i > 0$) determined by her risk type and on her effort level ($e_i \geq 0$), where $i = a, b$. In particular, a risky type borrower has a success probability of $k_a = p_a + e_a$ and a safe type has a success rate of $k_b = p_b + e_b$, with $p_a < p_b$ and $0 < k_a, k_b \leq 1$. Without loss of generality, if the project is successful the output takes the value of Y and 0 otherwise.

In the presence of local information, all borrowers know each other's risk type, but the outside lender (bank) does not. Following Ghatak (1999), in the absence of financial collateral the bank requires potential borrowers to form groups of size two where both members are jointly liable for each other. The bank offers to each group the joint liability contract (r, q), where $r > 0$ is the gross interest rate and $q > 0$ is the liability payment. Hence, r is the payment made by the individual who succeeds and q is the additional payment made by the individual when she succeeds and her partner fails. A borrower who fails pays the bank nothing. The expected payoff for type i borrower matched with type j borrower is, then, given by

$$E\pi_{ij} = (p_i + e_i)Y - (p_i + e_i)r - q(p_i + e_i)(1 - p_j - e_j) - 1/2\gamma e_i^2 \qquad (1)$$

where the disutility of the effort is captured by $-1/2\gamma e_i^2$, with parameter $\gamma > 0$.

We assume a non-cooperative game setting where each borrower maximizes her own expected payoff $E\pi_{ij}$ with respect to her effort e_i. We solve the maximization problem in Appendix B. The main results are summarized below:

1. A borrower's optimal effort level (e_{ij}^*, $i = a, b$) is higher if she is a safe type and/or if her partner is a safe type. That is, $e_{bb}^* > e_{ab}^* > e_{ba}^* > e_{aa}^*$.
2. A borrower prefers a safe partner to a risky partner, despite of her own type. That is, $E\pi_{bb}^* > E\pi_{ba}^*$ and $E\pi_{ab}^* > E\pi_{aa}^*$.

[7] In this model, we assume that the type refers to the riskiness of borrowers, but the type could also refer to other factors associated with the creditworthiness of borrowers like their entrepreneurial spirit, reciprocity, solidarity, trust or level of responsibility. In the empirical setup below, the group types may aggregate all these factors.

3. Joint liability with varying risk types and effort levels leads to a single equilibrium of positive assortative matching in group formation. More specifically, $E\pi^*_{bb} - E\pi^*_{ba} > E\pi^*_{ab} - E\pi^*_{aa}$. The net expected loss for a safe borrower of having a risky partner compared to having a safe partner is higher than the next expected gain of a risky borrower of having a safe partner compared to having a risky partner. As noted by Ghatak (1999), this equilibrium condition is similar to the optimal sorting property in Becker (1993), such that borrowers not in the same group should not be able to form a group without making one or both of them worse off.

The second and third results above are consistent with the results from Ghatak (1999). The intuition behind is that while a borrower of any type prefers a safe partner because of lower expected joint-liability payments, safe borrowers value safe partners more than risky borrowers because safe partners repay more often their loans and are more likely to realize the gains of having a safe partner. By allowing the probability of success to also depend on the effort level of borrowers, we additionally find that groups of safe partners will exhibit a higher effort, which translates into further higher repayment probabilities. This result reinforces the notion of a separating equilibrium in that borrowers of the same type will pair together and safe pairs will show an even higher likelihood of repayment than risky pairs.

We also allow for a cooperative game setting where each borrower maximizes the total payoff of her group with respect to her effort. We obtain the same key results of the non-cooperative game: a single equilibrium with positive assortative matching where groups of safe partners exhibit a higher effort than groups of risky partners. The derivation under this alternative setup is detailed in Appendix B.

Thus, a simple framework with peer selection and moral hazard helps to show how joint liability can lead to a separating equilibrium with the coexistence of two opposed groups: a group of safe borrowers with a higher probability of repayment (success) reinforced by higher effort levels, and a group of risky borrowers with a lower probability of repayment and lower efforts. The coexistence of different group types, driven by unobserved factors like risk and effort levels, implies the necessity to account for potential group types when modeling repayment behavior in group lending. Certainly, there are mechanisms other than joint liability through which group lending without financial collateral can lead to higher or lower repayment rates and varying group types; for example, the unobserved informal risk-sharing and social cohesion among group members.[8] The empirical method proposed below is flexible enough to allow for varying group types driven by a wide set of factors, which are not necessarily observable and may shape the repayment behavior of a group.

3 Data

3.1 Background and Data

The groups under study are located in Andhra Pradesh in India.[9] They are organized following a new self-help groups (SHG) model promoted by the World Bank, which targets poor women in rural areas. The model combines savings generation and micro-lending with social mobilization. In particular, women who generally live in the same village or habitat voluntarily form SHGs with the understanding of a joint liability mechanism. A typical SHG consists of 10-20 members who meet regularly to discuss social issues and activities. During the group meetings each member also deposits a small thrift payment into a joint bank account. Once enough savings have been accumulated, group members can apply for internal loans that draw from the accumulated savings at an interest rate to be determined by the group. After the group establishes a record of internal savings and repayment, it becomes eligible for loans through a commercial bank or program funds. This process of internal savings and repayments helps members to further screen each other as some individuals may leave the group prior to obtaining a formal loan.

The group as a whole, then, borrows from a commercial bank or program funds where all group members are held jointly liable for the debts of each other. The group generally allocates the loan to its members on an equal basis, and the group is not eligible for further loans unless it has made full repayment.[10] The loans may be used for labor activities or consumption smoothing. Groups also have the option of implementing non-lending programs with the support of the program funds such as in-kind credit for subsidized rice, marketing and insurance programs.

In this study, we focus on the first “expired” loan borrowed from commercial banks by each group. An “expired” loan refers to a loan that had passed its due date by the time the survey

[8] For empirical evidence on this matter see Gine and Karlan (2009) and Feigenberg et al. (2011).
[9] Andhra Pradesh is the fourth largest state in India by area and the fifth largest by population.
[10] Naturally, a woman who maintains a good record and ends in a group where not all members fulfill their loan obligations, may join another group in the future.

was conducted. In Andhra Pradesh, commercial banks carry out microfinance activities in non-overlapped territories, so groups located in contiguous villages borrowed from the same bank.

The sample includes 1,110 different group loans which were allocated to a total of 12,833 women. The data are from a SHG survey conducted between August and October 2006 in eight districts in Andhra Pradesh, which were chosen to represent the state's three macro-regions (Rayalaseema, Telangana, and Coastal AP).[11] The SHG survey contains socioeconomic characteristics of group members (households) such as education background, housing condition, land and livestock ownership, occupation, and caste. It also includes group characteristics such as age, meeting frequency of members and programs and services available within the group. More importantly, the survey directly recorded from SHG account books the information on all loans that were taken between June 2003 and June 2006. The information includes the terms of each loan, the members the loan was allocated to, and how much of the loan had been repaid by each member at the time of the survey.[12]

The SHG survey was complemented with a previous village survey that covered all the villages from which the SHGs were sampled. From the village survey, we construct four indicators to account for the economic environment of the sample groups. These indicators include availability of financial institution, public bus, telephone and post office.

Table 1 presents descriptive statistics of our full sample.[13] The top panel (Panel 1) reports member characteristics based on 12,833 observations while the bottom panel (Panel 2) reports group and loan characteristics based on 1,110 observations. Approximately twenty-three percent of the group members are literate and thirty-one percent belong to a scheduled tribe or scheduled caste. Around six percent of the members are disabled or have family members who are disabled. About sixty-five percent of households own some land, and thirty-three percent live in pucca houses, twenty-two percent in kutcha houses, and the other forty-five percent live in semi-pucca houses.[14] Similarly, about sixty-one percent are agricultural laborers who do not own land or

[11] The eight districts are Srikahulam, Adilabad, Anantapur, Kadapa, Warangal, Nalgonda, Nellore, and Visakhapatnam.
[12] The survey instrument included a separate section where the allocation of loans to members (member loans) was recorded. See Li et al. (2012) for further details on how the information on group loans and member loans was matched together.
[13] A detailed description of the variables used in the analysis is provided in Table A.1 in Appendix A.
[14] A pucca house has walls and roofs made of burnt bricks, stones, cement concrete, and timber while a kutcha house uses less sophisticated materials such as hays, bamboos, mud, and grass. A semi-pucca house uses a combination of materials from the other two types.

own such a small amount of land that they have to provide agricultural labor for others, twenty percent are self-employed agricultural workers, and the rest have other occupations (such as those self-employed and employed in non-agricultural sectors and housewives). The table also indicates that eighty percent of the group members in our sample fully repaid their loan by its due date (i.e. not defaulted). Figure A.1 in Appendix A further plots a histogram of the percentage of the loan repaid by each member. It follows that most of the data points are clustered at the endpoints, which supports the discrete treatment of the repayment (default) behavior in the empirical model.

Turning to the group and loan characteristics, the groups range from seven to twenty members and have close to thirteen members on average. The groups are from all of the three macro-regions in the state: about forty-five percent are located in Telangana, twenty-six percent in Rayalaseema, and the remaining twenty-nine percent in Coastal AP. The average group age is six years and roughly in nine of every ten groups the members meet on a regular basis (at least monthly). About twenty-eight percent of the groups have a food credit program (in-kind credit for subsidized rice), fifteen percent have a marketing program, and twenty-five percent have an insurance program. The group loan was allocated on average to twelve members and the average loan size received by a member is 3,338 rupees (about US67 dollars). The annual rate of interest is about 12.8 percent, which is much lower than the prevailing rate of moneylenders in India. The average duration of a loan is roughly one year and the majority of loans (ninety-six percent) required the groups to make repayments at least monthly.

3.2 Preliminary Analysis

A first look at the data is indicative of a separating equilibrium with apparently two group types. Table 2 shows that in more than 9 out of every 10 groups in our sample, either all of the members do not default or all of them default. In particular, in 76% of the groups (848 out of 1,110 groups) all of the group members fully repaid their loans or never defaulted and in another 17% of the groups (188 groups) all of the members defaulted. As discussed earlier, this repayment behavior may result from a combination of elements such as positive assortative matching ("matching likes") in group formation, in a context of joint liability, heterogeneous types and asymmetric information between borrowers and lenders.[15] Recall that under the SHG

[15] See Ahlin (2009) for a formal test on homogenous risk-matching in group lending.

model, groups have an initial period of internal savings and repayment, which also serves as an extended (ex-ante) screening period prior to applying for a commercial loan. This initial period also promotes social interaction among members, which may result in stronger social ties among them (see also Feigenberg et al. 2011). The observed pattern may also reflect variations (if any) in the level of effort and effectiveness of peer monitoring and peer pressure across groups, which may be correlated with peer screening. The theoretical model developed above indicates that groups composed of safe borrowers will also exhibit a higher level of effort than groups composed of risky borrowers. Hence a preliminary look at the data suggests the existence of mainly two group types: a "responsible" group of apparent "low risk" individuals with probably high efforts and/or effective monitoring and enforcement rules and strong social cohesion, and an "irresponsible" group of apparent "high risk" individuals with probably low efforts and/or ineffective monitoring and enforcement rules and lack of social cohesion.[16]

There is also the possibility of external factors, like a negative weather shock, affecting the likelihood of repayment of all members in a group, which generally live close to one another and perform similar labor activities. However, groups where all members defaulted in our sample are not concentrated at a particular location, which reduces the possibility of specific weather shocks or other contextual factors explaining inter-group variation on default behavior. In particular, Figure A.2 shows that villages with a high proportion of groups where all members default are well dispersed across the eight districts of our sample in Andhra Pradesh.[17] In addition, the estimation results presented below indicate that the variables included in the repayment equation (individual and loan characteristics) have a differentiated effect on the likelihood of default by group type, which further supports the existence of type-varying groups.

To further examine the possibility of homogenous sorting among groups, Table A.2 reports the number of groups in which the intra-group variance is less than or equal to the overall variance considering all groups in the same village and mandal for different borrower

[16] The existence of the mixed group (7% of our group sample) suggests that the observed defaults are not necessarily strategic defaults. If some members fail to repay some installments, the other members still have the incentive to repay on time because they do so in hope that the delinquent borrowers will repay their installments on a future date. In addition, individuals that maintain a good repayment record are more likely to join a "better" group in the future (if necessary). Formally addressing the dynamic aspects of installment repayments is beyond the scope of our paper.
[17] For areas with available weather data (rainfall) and vegetation information (Normalized Difference Vegetation Index or NDVI) during the period of analysis, we also did not find any significant correlation between these measures and default behavior.

characteristics.[18] The characteristics include literacy, household characteristics, land ownership, occupation and caste. The results show that individuals with similar observable characteristics appear to group together. On average, in 70-72% of the cases the intra-group variance for a given characteristic is smaller than the intra-village or intra-mandal variance. There is a relatively higher degree of homogeneity among group members in terms of belonging to a scheduled tribe or caste and being self-employed agricultural worker, and a lower level of homogeneity in terms of literacy.

Overall, a preliminary look at the data is indicative of the coexistence of different types of groups in our sample. This suggests the necessity to allow for potential unobserved group types when examining repayment behavior in group lending.

4 Empirical Model

This section develops an empirical model to address the potential omitted variable problem in group lending with unobserved types. We use a mixture model to explicitly account for unobserved group types when evaluating the repayment behavior of individual members. The unobserved types may result from peer selection as well as from variation in the level of effort and effectiveness of peer monitoring and pressure and other unobserved factors like social cohesion. The probability of default is conditional on the unobserved type and depends on observable individual and loan characteristics, while average member characteristics and other group and village characteristics (observed by lenders) may help to identify the group type the individual belongs to.

Let the default behavior of individual *i* in group *j* be given by

$$D_{ij} = 1(\alpha + X_{ij}\beta_1 + C_j\beta_2 + T_j^* + u_{ij} > 0) \qquad (2)$$

where D_{ij} is the observed binary outcome, i.e. D_{ij} equals one if the individual defaults (i.e. does not fully repay her loan) and equals zero otherwise, α is a constant, X_{ij} is a vector of observable individual characteristics, C_j is a vector of loan characteristics, T_j^* is the unobserved

[18] The comparisons exclude all villages (150 out of 457) and mandals (3 out of 97) where there is only one group in the village or mandal. A mandal is the equivalent to a sub-district in India and comprises several villages.

group type which is likely correlated with X_{ij} (and C_j), and u_{ij} is an error term. On the correlation between X_{ij} and T_j^*, we can think, for example, of a proxy for the social ties of an individual, included in X_{ij} and potentially correlated with the social ties of her peers (who generally live in the same neighborhood), which partly describe T_j^*.

If group heterogeneity is solely based on observables, the observed group characteristics (Z_j) like average member characteristics and other group controls, including social ties, would be sufficient to identify the group types, and Z_j could be used as a proxy for T_j^* to estimate equation (2) using a standard probabilistic regression (e.g., Probit, Logit). However, the unobserved group type is more accurately characterized by both observable and unobservable factors such that $T_j^* = Z_j\delta + W_j + \varepsilon_j$, where W_j is unobserved, Z_j and W_j are potentially correlated, and ε_j is an error term. Following the previous example, a proxy for the social ties or connections of a group, included in Z_j, is likely correlated with the unobserved economic opportunities and entrepreneurial spirits of the group members, which are comprised in W_j and further affect repayment.

Hence a standard probabilistic regression of equation (2) with only Z_j in the right-hand side will result in an omitted variable bias as W_j will be embedded in the error term. Another option is to incorporate the unobserved group component or type as fixed effects in a conditional logit model. Yet, a fixed-effects logistic regression mainly exploits within-group variation and will drop all groups without intra-group differences in default behavior (i.e. more than 90 percent of our sample). Further, the observed factors affecting repayment performance may vary by group type.

To address this potential omitted variable problem we propose an alternative model, where group heterogeneity can be captured by allowing groups to be one of two types with a specific probability. In particular, we assume that T_j^* can take two possible values, T_j^H if the group is "responsible" and T_j^L if the group is "irresponsible". In broader terms, we can think of the first group as a group mainly composed by "safe" borrowers with effective monitoring and enforcement efforts and high reciprocity and solidarity among members, and of the second group as a group of “risky” borrowers with less effective monitoring and enforcement efforts and low reciprocity and solidarity among members. We could easily relax this assumption to allow for a wider set of types (based on different combination of factors) but our data seems to support a two-type model. In particular, we also estimated a three-type model but the two-type model provides a better fit based on the Schwarz Bayesian Information Criterion (SBIC).[19]

Then, the repayment behavior of individual *i* in group *j* is given by

$$D_{ij} = \begin{cases} 1(\alpha_H + X_{ij}\beta_{1,H} + C_j\beta_{2,H} + u_{ij,H} > 0) & \text{if } T_j^* = T_j^H \\ 1(\alpha_L + X_{ij}\beta_{1,L} + C_j\beta_{2,L} + u_{ij,L} > 0) & \text{if } T_j^* = T_j^L \end{cases} . \quad (3)$$

In this specification, the effect of T_j^* is absorbed by the constant terms α_H and α_L, and $Cov(X_{ij}, u_{ij}) = 0$. We further allow for varying coefficients across group type, which permits to capture varying effects of different factors on repayment behavior by type.[20]

The probability of being in type-*H* group ($T_j^* = T_j^H$) can be further modeled as

$$\Pr(T_j^* = T_j^H) = \Pr(\overline{X}_j\delta_1 + G_j\delta_2 + v_j > 0) \quad (4)$$

where $\overline{X}_j$ is a vector of average characteristics of group members, G_j is a vector of group and village controls (G_j), and v_j is an error term.[21] Hence while the individual characteristics of each group member (X_{ij}) help us to approximate their default probability, the average characteristics of all group members ($\overline{X}_j$) can help us to identify their group type. The member characteristics considered for the analysis include literacy, land ownership, housing condition, occupation and caste.[22] Thus, while belonging to a certain caste, for example, may directly affect

[19]The SBIC of the two-type model is 0.838 versus 0.849 of the three-type model. Further, the predicted probability of being in the potential third type group is close to zero.
[20] This flexibility is similar to Gan and Hernandez (2013) who allow for varying coefficients across potential collusive and non-collusive regimes when modeling the pricing and occupancy rate behavior of hotels under a switching regression framework.
[21] The underlying assumption is that the probability of being a certain group type varies with some observable characteristics; in this case with $\overline{X}_j$ and G_j.
[22] This type of personal information is also generally disclosed during credit application processes.

the likelihood of repayment, the percentage of members belonging to a similar caste (included in $\overline{X}_j$) can serve as a proxy for social ties within the group, which will also have an indirect effect in the probability of default.[23] We also account for loan characteristics (C_j) in the repayment equation (e.g., loan amount, interest rate, length, repayment frequency) and we use other group and village controls (G_j) to help us identify the group type (e.g., age, number of members, location, access to programs and services).

Note that since T_j^* is likely determined by both observable ($\overline{X}_j, G_j$) and unobservable (W_j) characteristics, the parameters in equation (4) may not be consistently estimated. However, the fact that we do not observe W_j does not result in inconsistent estimates of the parameters in the repayment equation (3); we only require some but not full information about T_j^* to identify the parameters in the repayment equation. Intuitively, the identification is similar to that underlying a two-stage least squares (2SLS) procedure, where the consistency of the 2SLS estimations does not require the consistency of the first-stage regression. Mahajan (2006) refers to ($\overline{X}_j, G_j$) as instrumental-like variables (ILV). Henry et al. (2010) study the identification of this type of model. They conclude that the current model is fully identifiable if ($\overline{X}_j, G_j$) are conditionally independent of the errors in equation (3). Gan et al. (2011) also provide a discussion on the identification condition.

Formally, the key identifying assumption in the proposed model is that conditional on the group type, both observable and unobservable factors that characterize T_j^* are not related to the probability of defaulting. That is,

$$\Pr(D_{ij} = 1 \mid T_j^* = T_j^H, \overline{X}_j, G_j, W_j) = \Pr(D_{ij} = 1 \mid T_j^* = T_j^H)\,. \tag{5}$$

[23] Particularly, we generate a variable of percentage members belonging to the leading caste (defined as the caste with the largest number of members in the group) to capture social ties. Unfortunately we do not have more detailed information, like number of relatives, to more accurately control for social ties within the group.

Consequently, any association between $\bar{X}_j$, G_j and W_j and the probability of defaulting is solely driven by the association between these former variables and the probability of being of a certain group type.

The unconditional probability of default can, in turn, be written as

$$\begin{aligned}\Pr(D_{ij}=1) &= \Pr(D_{ij}=1, T_j^*=T_j^H)+\Pr(D_{ij}=1, T_j^*=T_j^L) \\ &= \Pr(D_{ij}=1\,|\,T_j^*=T_j^H)\Pr(T_j^*=T_j^H)+\Pr(D_{ij}=1\,|\,T_j^*=T_j^L)\Pr(T_j^*=T_j^L).\end{aligned} \tag{6a}$$

Similarly,

$$\Pr(D_{ij}=0)=\Pr(D_{ij}=0\,|\,T_j^*=T_j^H)\Pr(T_j^*=T_j^H)+\Pr(D_{ij}=0\,|\,T_j^*=T_j^L)\Pr(T_j^*=T_j^L). \tag{6b}$$

If we further assume that the error terms in equations (3) and (4) have a $F(\cdot)$ and $J(\cdot)$ cumulative distribution function (cdf), respectively, the log likelihood for individual *i* in group *j* is given by

$$\begin{aligned}\ln l_{ij} = D_{ij} &\ln[F(\alpha_H + X_{ij}\beta_{1,H} + C_j\beta_{2,H})J(\bar{X}_j\delta_1 + G_j\delta_2) \\ &+ F(\alpha_L + X_{ij}\beta_{1,L} + C_j\beta_{2,L})(1-J(\bar{X}_j\delta_1 + G_j\delta_2))] \\ &+ (1-D_{ij})\ln[1-F(\alpha_H + X_{ij}\beta_{1,H} + C_j\beta_{2,H})J(\bar{X}_j\delta_1 + G_j\delta_2) \\ &- F(\alpha_L + X_{ij}\beta_{1,L} + C_j\beta_{2,L})(1-J(\bar{X}_j\delta_1 + G_j\delta_2))].\end{aligned} \tag{7}$$

We approximate $F(\cdot)$ and $J(\cdot)$ with a logistic cdf.[24]

5 Results

We now turn to our estimation results. For comparison purposes, we first report the results using a standard probabilistic regression model, which does not account for unobserved types when modeling the likelihood of default. Table A.3 presents the parameter estimates (and standard

[24] We also estimated the model using a normal cdf and obtained qualitative similar results.

errors) of a Probit model using three alternative specifications.[25] The first model only accounts for member and loan characteristics. Although most of the coefficients of the member characteristics generally have the expected signs, in the sense that the variables associated with a low (high) economic status are positively (negatively) correlated with the probability of default, they are generally not statistically significant at conventional levels. We only observe a positive and significant correlation between the probability of default and belonging to a scheduled caste. The loan characteristics, in turn, show a higher correlation with repayment behavior. A larger loan amount, higher interest rate, longer duration and lower repayment frequency are all associated with a higher probability of default.

The second model adds average (leave-me-out) member characteristics and other group and village controls, which are intended to account for contextual factors that could also affect an individual's repayment decision. While the positive correlation between the probability of default and belonging to a scheduled caste disappears, a higher proportion of members of a scheduled caste in the group is associated with a lower repayment probability; the other member characteristics (and the corresponding group averages) remain not significant. The effects of most of the loan characteristics also remain intact. Several of the other group and village controls exhibit an important association with the probability of default. In particular, having a marketing and insurance program in the group, frequent meetings between group members, and the existence of a financial institution in the village, are all positively correlated with the probability of repayment. In contrast, members of groups with a food program, which is distinctive of poorer groups, show a higher probability of default. Finally, in smaller groups (less than thirteen members), an additional member in the group decreases the individual probability of default probably due to stronger peer monitoring and pressure effects while in larger groups (thirteen members or more) occurs the contrary as coordination, monitoring and enforcement efforts are probably more difficult to become effective in considerably large groups.

While in the first and second model we account for the potential correlation in the repayment decision among group members by clustering the error term by group, in the third model we explicitly control for the potential within-group correlation by estimating a Probit model with random effects. The inclusion of the random group term in the estimated regression

[25] We use a Probit model because it provides a better fit and performance than a Logit and a linear probability model. Details are available upon request.

although improves the model fit (the within-group correlation is also highly significant), it does not improve the model performance discussed below. Most of the effects of the explanatory variables also remain similar.[26]

As noted above, however, all these models do not account for the unobserved group-type component, embedded in the error term of the repayment equation and potentially correlated with some of the explanatory variables. Table 3 shows the estimation results of the alternative mixture model proposed, which explicitly accounts for unobserved group types when modeling the default behavior of group members. The model allows for two group types (type *H* and type *L*) and the repayment decision is conditional on the unobserved type, where the marginal effects of the member and loan characteristics may vary by type. The average member characteristics and other group and village controls, in turn, help to identify the group type.

Several important patterns emerge from the table. First, the conditional probability of default is considerably different between the two group types, as reported at the bottom of the table. More specifically, the estimated probability of default conditional on being in a group of type-*H* individuals is 9.5 percent versus 62.8 percent in a group of type-*L* individuals. Hence the model clearly distinguishes two group types: one type (type *H*) likely composed of "responsible" individuals with probably high levels of effort and/or effective monitoring and enforcement rules who are more likely to repay their loans, and a second type (type *L*) composed of "irresponsible" individuals with probably low levels of effort and/or less effective monitoring and enforcement rules who are less likely to repay their loans. Similarly, the average probability of being a type-*H* group is roughly 80 percent in our sample and, interestingly, groups where all members pay back their loan exhibit a higher probability of being a type-*H* group than other groups.[27] In particular, in groups where none of the members defaulted the likelihood of being a type-*H* group is 82.9 percent versus 76.4 percent in groups where some members defaulted and 66.9 percent in groups where all members defaulted. These results further support the identification of seeming "responsible" and "irresponsible" groups by our model.

An analysis of the factors used to describe the probability of being in a type-*H* group also indicates that "responsible" groups are more likely characterized, for example, by women who

[26] In this third model, individuals in groups with a higher proportion of disabled members in the household are also expected to fully repay their loans and group age is positively correlated with the probability of default (up to groups of eleven years old).

[27] Recall that in our raw data we observe full repayment by all members in 76% of the groups and in another 17% of the groups all members default.

are literate, own some portion of land, live in semi-pucca houses, are related to agricultural activities and belong to a scheduled tribe but not necessarily to a leading caste. Similarly, “responsible” groups are more likely to hold frequent meetings between its members, have a marketing and insurance program but not a food credit program for its members, and have access to additional services in the village such as a financial institution and telephone. Microfinance institutions should probably look for these characteristics when trying to identify potential “responsible” groups and/or areas where to operate or expand. Holding frequent meetings appear to be particularly important, as we further detail below. This is in line with other studies that suggest that, besides facilitating peer monitoring and enforcement, frequent group meetings may directly increase social contact and reduce lending risks (Gine and Karlan 2009; Feigenberg et al. 2011).[28] The existence of other programs in the group (like marketing and insurance programs), could also stimulate social cooperation and strengthen social ties, in addition to providing additional services to members, thereby increasing the risk-sharing among members.[29]

Figure A.3 provides additional support to the correct identification of “responsible” and “irresponsible” groups by our model, based on the observed behavior patterns in the data. For example, the probability of being a type-*H* (“responsible”) group is positively correlated with the proportion of literate women in the group; a closer look at the data shows that effectively among groups with more than half of the women in the group literate, there is a higher proportion of groups with no members defaulting (82 percent) and a lower proportion of groups with all members defaulting (13 percent), as compared to groups with less than half of the women literate (76 and 17 percent). The differences are more pronounced when comparing the distribution of intra-group default behavior between groups with high and low frequency meetings. Among groups that at least hold monthly meetings, which is also distinctive of type-*H* groups, the proportions of groups with no members defaulting and all members defaulting are 80 and 14 percent; among groups that hold less than monthly meetings, the corresponding proportions are 48 and 41 percent. Similar patterns are observed when comparing groups with and without marketing programs and a financial institution in the village, which are also correlated with the likelihood of being a type-*H* group in the model. These findings suggest that several of the

[28] Gine and Karlan (2009) find that groups with stronger social networks are less likely to experience default problems after removing joint liability. Feigenberg et al. (2011) show that repeated interactions can facilitate cooperation by allowing individuals to sustain reciprocal economic ties.

[29] Fearon et al. (2009) and Feigenberg et al. (2011) also show, in different settings, the importance of community development programs to encourage social cohesion.

factors included in the type-probability equation indeed help to identify potential group types and, in particular, that the types in the model are not purely identified by functional form.

Another important pattern that emerges from Table 3 is the difference in direction, magnitude and statistical significance of several of the parameter estimates in the default equation between the two group types. This suggests that the factors driving individual repayment behavior may vary by type. Table 4 shows the conditional marginal effects for the different individual and loan characteristics included in the repayment equation after accounting for group type.[30] We do not observe major changes in the probability of default among type-*H* group members after a change in most of the individual covariates; being a self-employed agricultural worker and living in pucca houses decrease the probability of default by roughly three and one percentage point, while owning some portion of land increases the likelihood of defaulting by less than one percent. Among type-*L* group members, in contrast, being a self-employed agricultural worker increases the probability of default by 14 percentage points; being an agricultural laborer also substantially increases the likelihood of defaulting by 29 percentage points, as well as belonging to a scheduled caste (31 percentage points). Owning some portion of land or living in either pucca or kacha houses (relative to semi-pucca houses), in turn, decrease the probability of default by 8-16 percentage points.

Regarding the loan covariates, monthly (or higher) repayment frequencies and an additional member receiving a loan decrease, for example, the likelihood of defaulting by three and 0.2 percentage points among type-*H* group members; among type-*L* group members, the corresponding decrease is of 26 and five percentage points. An increase in the loan amount, interest rate and loan duration also results in a much higher increase in the probability of default among type-*L* group members than among type-*H* group members.

These varying effects by type can help lenders to better assess their clients and understand the factors driving their behavior. Owning some portion of land, housing conditions, labor activities and belonging to a scheduled tribe seem to matter among type-*L* groups, in contrast to type-*H* groups where the effects (if any) are much more limited. The loan characteristics are also more relevant for type-*L* groups than for type-*H* groups. These differences further have important policy implications and can help lending institutions to reduce

[30] The normal-based confidence intervals reported for the estimated marginal effects are based on 200 bootstrap replications and are biased-corrected. Although not reported, the bootstrap means are very similar to the estimated marginal effects, which support the bootstrap procedure implemented.

their transaction costs. Field and Pande (2008), for example, point out the important tradeoff between imposing higher repayment frequencies (a standard practice among microfinance institutions to encourage fiscal discipline and reduce default risk) and the substantial increase in transaction costs of installment collection. The authors find that switching to lower frequency repayment schedules could allow lenders to significantly reduce their transaction costs with virtually no increase in client default, particularly among first-time borrowers. Our results suggest that the fiscal discipline imposed by frequent repayment is critical among groups suspected (or with a higher probability) of being type-*L* groups, but not on type-*H* groups where less costly repayment schedules could be implemented; the cost savings are likely higher than the (marginal) increase in the default rate in this type of groups. Encouraging longer term investments through higher loan terms also seems more reasonable among type-*H* groups, which could improve the borrowers' repayment capacity in the long run (in a similar way as a more flexible repayment schedule).

The parameter estimates in the two-type model are also different from those obtained under a standard probabilistic regression, which does not allow for unobserved consumer types. To better appreciate these differences, Table 5 reports the unconditional marginal effects on the probability of default for all the variables included in the regression analysis for the Probit and two-type model specifications.[31] In the full two-type model (last column), the average member characteristics and other group and village characteristics affect the likelihood of defaulting through the probability of being in a type-*H* group or "responsible" group. A direct comparison between the full Probit model and the two-type model reveal that the two models produce different marginal effects.[32] For example, being an agricultural laborer or belonging to a scheduled caste increases the overall probability of default by roughly four percentage points in the two-type model (all else equal), while in the Probit model the change in the probability is not significant; a similar pattern is observed for the condition of living in pucca houses or being self-employed agricultural workers, which decrease the overall probability of default by three and one percentage points in the type-varying model and are not significant in the Probit model. Similarly, monthly (or higher) repayment frequencies will decrease the likelihood of defaulting

[31] The marginal effects of the Probit model with random effects, excluded from the table, are qualitatively similar (although smaller) to those of the full Probit model. For comparison purposes, the confidence intervals of the marginal effects for all models were derived using 200 bootstrap replications.
[32] Note that the marginal effects decrease as we move across the two Probit-model specifications, for the variables they can be compared.

by six percentage points in the two-type model and by seven percentage points in the Probit model, while an additional year in the length of the loan will increase the likelihood of defaulting by four percentage points in the first model and by more than eight percentage points in the second model. Interestingly, an additional member in a group seems to increase the probability of default in the type-varying model while in the Probit model is the converse, at least in smaller groups; it seems that the stronger peer monitoring and pressure effects do not necessarily outweigh the higher coordination costs of having additional members in the group.

From the two models, however, it is also clear the importance of frequent meetings among group members, for individuals to not fall behind in their loan repayments (probably resulting in better peer monitoring and pressure and/or higher social interactions). In particular, in groups where members meet at least on a monthly basis, the individual probability of default is 30 percentage points lower in the Probit model and 45 percentage points lower in the type-varying model than in groups where members meet less frequently. Both models also suggest the importance of promoting marketing and insurance programs among group members, which are negatively correlated with defaulting, and the inverse for subsidized food credit programs, which are also distinctive of poorer groups. The existence of a financial institution and a telephone in the village is also highly correlated with a positive repayment behavior under the two models.

Overall, the results indicate the importance of having a flexible, type-consistent model, which allows for varying effects by type and provides better insight about the possible factors affecting the members' repayment behavior. The proposed model can also help lenders to better identify and screen their potential clients, as we further discuss below.

5.1 Model Identification

Next, we further evaluate the identification of our empirical model. As noted above, a formal implication of the type-varying model is that we require some but not full information about the factors describing group heterogeneity (T_j^*) to identify the parameters in the main repayment equation.[33] Our model setup allows for both the presence of observable ($\overline{X}_j, G_j$) and unobservable (W_j) characteristics. Hence, even a subset of the observed factors used to identify

[33] See also Gan et al. (2011) for further details.

the group types may produce consistent estimates of the parameters in the main repayment equation.

Tables A.4 through A.6 report the estimation results of the two-type model when excluding different subsets of the variables used to identify the type-*H* group. In particular, we separately exclude the average member characteristics, group size and age, group programs, if group has frequent meetings, group location, and village characteristics. We observe that the coefficients of both the individual and loan characteristics, included in the repayment equation, are generally not much sensitive to the inclusion or exclusion of different variables in the group-type equation. In our full sample estimations in Table 3, for example, the coefficients for self-employed agricultural worker is -0.593 (0.184) among type-*H* groups and 1.173 (0.266) among type-*L* groups, while the coefficients for interest rate is 0.083 (0.013) among type-*H* groups and 0.277 (0.034) among type-*L* groups. When excluding different subsets of variables in the group-type equation, the corresponding coefficients fluctuate between -0.521 (0.113) – -0.644 (0.074), 0.979 (0.317) – 1.451 (0.331), 0.082 (0.013) – 0.094 (0.011), and 0.234 (0.040) – 0.284 (0.039). The Hausman tests reported in Table A.7 further indicate that in most cases there are not systematic differences between the coefficients in the repayment equation of the baseline model and the corresponding coefficients in these alternative specifications, at least at a 5 percent level of significance. This exercise provides additional support for the robustness of the mixture model proposed.

5.2 Predictive Performance

We now analyze whether allowing for different group types yields better out-of-sample predictions for the probability of default. We want to examine if the proposed type-varying model has a higher predictive power than standard probabilistic methods, which can further help to reduce information asymmetries in micro lending and aid lenders to correctly identify and select their current and future clients (groups). To conduct the performance assessment, we follow a standard cross-validation procedure and randomly partition our dataset into a design sample for model estimation (60% of the observations) and a test sample for further analysis (40% of the observations). The partition is conducted at the group level and both samples maintain the population proportions of default and non-default actions.

Table 6 provides performance indicators for the different models estimated.[34] The indicators include the average predicted default probability, the mean square predicted error and several performance indicators based on converting the estimated default probabilities to a binary regime prediction using the standard 0.5 rule (i.e. if the estimated default probability is greater or equal to 0.5 the individual is predicted to default, while if the estimated probability is less than 0.5 the individual is predicted to not default). For the two-type model, the performance assessment is based on two alternative calculations of the probability of default. Generally speaking, a lender could evaluate granting a loan based on the estimated unconditional probability of default or based on the conditional probability of default, depending on the likelihood of being in a group of a certain type. Hence different mixtures for estimating the probability of default could be used.

The two approaches considered are:

(1) A “naïve” type-consistent approach that only uses the unconditional probability of default such that,

$$\begin{aligned}\Pr(D_{ij}=1) &= F(\alpha_H + X_{ij}\beta_{1,H} + C_j\beta_{2,H})J(\overline{X}_j\delta_1 + G_j\delta_2) \\ &\quad + F(\alpha_L + X_{ij}\beta_{1,L} + C_j\beta_{2,L})(1 - J(\overline{X}_j\delta_1 + G_j\delta_2)).\end{aligned}$$

(2) A “conservative” type-consistent approach which takes into account the likelihood of being in a type-*H* group. In particular,

$$\Pr(D_{ij}=1) = \begin{cases} F(\alpha_H + X_{ij}\beta_{1,H} + C_j\beta_{2,H}) & \text{if } \hat{\Pr}(T_j^* = T_j^H) \text{ in upper quintile} \\ F(\alpha_H + X_{ij}\beta_{1,H} + C_j\beta_{2,H})J(\overline{X}_j\delta_1 + G_j\delta_2) & \\ + F(\alpha_L + X_{ij}\beta_{1,L} + C_j\beta_{2,L})(1 - J(\overline{X}_j\delta_1 + G_j\delta_2)) & \text{if } \hat{\Pr}(T_j^* = T_j^H) \text{ in 2nd - 4th quintile} \\ F(\alpha_L + X_{ij}\beta_{1,L} + C_j\beta_{2,L}) & \text{if } \hat{\Pr}(T_j^* = T_j^H) \text{ in lower quintile} \end{cases}$$

[34] The results are based on 200 repeated 60-40% partitions. The results are also not sensitive to alternative data partitions (70-30% and 50-50%).

where $\hat{\Pr}(T_j^* = T_J^H)$ is the estimated probability of being in a type-*H* group.[35]

As shown in the table, the “naive” approach produces a mean default probability (19.9%) closer to the observed sample mean of 21% than the full Probit model (18.6%) and the “conservative” approach (23.7%). The “naïve” and “conservative” approach also report a lower mean squared prediction error than the Probit model (0.145 and 0.156 versus 0.159). The two type-consistent approaches also show a higher overall predictive performance based on McFadden et al. (1977) standard measure.[36] In particular, the “naïve” approach has a predictive performance of 76.4% and the “conservative” approach has a predictive performance of 76% versus 74.7% of the Probit model. The poorer performance of the Probit model is largely explained by its lower correct default classification rate (i.e. identification of “bad” borrowers): 17.2% versus 21.9% of the “naïve” approach and 31.3% of the “conservative” approach. Regarding the correct non-default classification rate (i.e. identification of “good” borrowers), the Probit model performs better than the “conservative” approach, but poorer than the “naïve” approach.

An alternative way to evaluate the out-of-sample performance consists in examining the number of “good” clients the model rates as “bad” (Type I error) and the number of “bad” clients the model rates as “good” (Type II error) for varying cutoff values of the probability of default. In Table 6, we used the standard 0.5 rule for the performance assessment. Figures 1 and 2 compare the percentage of “good” borrowers rejected and the percentage of “bad” borrowers accepted across the Probit, “naïve” and “conservative” type-consistent approaches for different cutoff values. In the case of Type I errors, the “naive” approach and the Probit model outperform the “conservative” approach for most of the cutoff values. More specifically, for cutoff values above 0.1 the lending institution will do better in identifying “good” clients by relying on the “naïve” approach or Probit model. In the case of Type II errors, however, both the “naïve” and “conservative” approach outperform the Probit model for basically the entire range of cutoff values, and for values above 0.3 the “conservative” approach has a considerably higher (and

[35] This approach is in line with Gan and Mosquera (2008) who examine unobserved consumer types in the Ecuadorian credit card market.

[36] McFadden et al. (1977) overall performance measure is equal to $p_{11} + p_{22} - p_{12}^2 - p_{21}^2$, where p_{ij} is the *ij*th entry (expressed as a fraction of the sum of all entries) in the 2x2 confusion matrix of actual versus predicted (0,1) outcomes using the 0.5 rule.

increasing) performance than the “naïve” approach. For sufficiently lenient acceptance rules (cutoff values above 0.5), the differences in the percentage of “bad” accepted between the “conservative” approach and the other models are in the order of 10-23 percentage points.

Hence, we generally attain a higher predictive power when allowing for unobserved group types when modeling the probability of default of group members, as compared to a standard probabilistic regression model. If the lending institution is more interested in minimizing the number of “bad” clients (classified as “good” by the model), the lender should probably follow a “conservative” approach, while if the lender is more interested in identifying “good” clients (classified as “bad” by the model) it should follow a “naïve” approach; the Probit model will also perform well for the latter. Yet, for more lenient acceptance rules using a “naïve” approach or Probit model will also result in a much higher acceptance rate of “bad” clients relative to the “conservative” approach. For example, for a cutoff value of 0.4 the “naïve” approach outperforms the “conservative” approach by three percentage points in terms of the rejection rate of “good” clients, while the “conservative” approach outperforms the “naïve” approach by a similar degree in terms of the acceptance rate of “bad” clients; but for a cutoff value of 0.6, the “naïve” approach outperforms the “conservative” approach by four percentage points when identifying “good” clients, while the “conservative” approach outperforms the “naïve” approach by fourteen percentage points when identifying “bad” clients.

6 Concluding Remarks

This paper proposes an empirical model to address the potential omitted variable problem resulting from group lending with unobserved types. We use a mixture model to explicitly account for group types when modeling the repayment behavior of group members. In the model, individuals make repayment decisions based on their unobserved group type as well as on observable individual and loan characteristics. Average member characteristics and other group and village characteristics help, in turn, to identify the group types. We also allow the marginal effects in the repayment equation to vary across types.

The estimation results support our model specification and show the advantages of relying on a type-consistent method when examining the probability of default of group members. First, the model clearly distinguishes two group types: an apparent “responsible” group with a low probability of default among group members and another “irresponsible” group with a high probability of default. Second, we find important differences in the marginal effects of the different individual and loan characteristics included in the repayment equation across group types. Third, the type-varying model shows a higher predictive performance than standard probabilistic models. From a policy perspective, our model helps to better understand the underlying factors driving repayment behavior, which appear to differ across groups. These differences can aid lenders when designing loan contracts for different “types” of clients. Similarly, the model can help to attenuate information asymmetries in micro lending by aiding lenders to correctly classify their potential clients. A more accurate risk scoring tool is essential to reduce the high transaction costs faced by micro finance institutions. It can also prevent including potential “bad” borrowers and excluding “good” borrowers from sensitive microcredit markets in developing regions.

Finally, it is worth noting that the analysis has focused on a two-type model given the nature of our data. The apparent two types may result from a combination of factors, including peer selection, peer monitoring and pressure and other unobserved factors like social cohesion, but disentangling these effects is beyond the scope of the study. Certainly, there can be a wider set of types in other contexts, and the proposed method can be easily adapted to allow for additional types. Considerably increasing the number of types, however, may require imposing restrictions on the value of the coefficients in the repayment equation (for example, not necessarily allowing for different marginal effects across all types) in order to avoid a highly parameterized model, which could be difficult to estimate in practice. Our analysis also follows a discrete treatment of the repayment decision given the observed behavior of most of the borrowers in the sample (either full repayment or no payment). Yet, the model can be adapted to examine instead the percentage of loan repaid by members. Future research should further attempt to incorporate dynamic aspects in the repayment decision of members under a type-varying setting.

References

Ahlin, C. (2009). Matching for credit: risk and diversification in Thai microcredit groups. BREAD Working Paper No. 251, December.

Ahlin, C. and Townsend, R.M. (2007). Using repayment data to test across models of joint liability lending. *Economic Journal* 117: F11-F51.

Armendariz de Aghion, B. (1999). On the design of a credit agreement with peer monitoring. *Journal of Development Economics* 60: 79-104.

Armendariz de Aghion, B. and Gollier, C. (2000). Peer group formation in an adverse selection model. *Economic Journal* 110: 632-643.

Armendariz de Aghion, B. and Morduch, J. (2004). Microfinance: Where do we stand? In Goodhart, Ch. (Ed.) *Financial development and economic growth: Explaining the links*. Palgrave Macmillan, Basingstoke, UK.

Armendariz de Aghion, B. and Morduch, J. (2005). The economics of microfinance. MIT Press, Cambridge, MA.

Banerjee, A., Besley, T. and Guinnane, T. (1994). Thy neighbor's keeper: the design of a credit cooperative with theory and a test. *Quarterly Journal of Economics* 109: 491-515.

Becker, G. (1993). A Treatise on the Family. Harvard University Press, Cambridge, MA.

Chowdury, P.R. (2005). Group lending: sequential financing, lending monitoring and joint liability. *Journal of Development Economics* 77: 415-439.

Cull, R., Demirguc-Kunt, A. and Morduch, J. (2007). Financial performance and outreach: A global analysis of leading microbanks. *Economic Journal* 117: F107-F133.

Fearon, J.D., Humphreys, M. and Weinstein, J.M. (2009). Can development aid contribute to social cohesion after civil war? Evidence from a field experiment in post-conflict Liberia. *American Economic Review* 99: 287-219.

Feigenberg, B., Field, E. and Pande, R. (2011). The economic returns to social interaction: Experimental evidence from microfinance. Working Paper.

Field, E. and Pande, R. (2008). Repayment frequency and default in microfinance: Evidence from India. *Journal of the European Economic Association* 6: 501-509.

Gan, L. and Hernandez, M.A. (2013). Making friends with your neighbors? Agglomeration and tacit collusion in the lodging industry. *Review of Economics and Statistics,* forthcoming.

Gan, L., Huang, F. and Mayer, A. (2011). A simple test of private information in the insurance markets with heterogeneous insurance demand. NBER Working Paper 16738, January.

Gan, L. and Mosquera, R. (2008). An empirical study of the credit market with unobserved consumer types. NBER Working Paper 13873, March.

Ghatak, M. (1999). Group lending, local information, and peer selection. *Journal of Development Economics* 60: 27-50.

Ghatak, M. (2000). Screening by the company you keep: joint liability lending and the peer selection effect. *Economic Journal* 110: 601-631.

Ghatak, M. and Guinnane, T. (1999). The economics of lending with joint liability: theory and practice. *Journal of Development Economics* 60: 195-228.

Gine, X. and Karlan, D. (2009). Group versus individual liability: Long term evidence from Philippine microcredit lending groups. Working Paper.

Guttman, J.M. (2008). Assortative matching, adverse selection, and group lending. *Journal of Development Economics* 87: 51–56.

Henry, M., Kitamura, Y. and Salanie, B. (2010). Identifying finite mixtures in econometric models. Cowles Foundation Discussion Paper #1767.

Hermes, N. and Lensink, R. (2007). The empirics of microfinance: What do we know? *Economic Journal* 117: 1-10.

Hermes, N., Lensink, R. and Mehrteab, H. (2005). Peer monitoring, social ties and moral hazard in group lending programmes: evidence from Eritrea. *World Development* 33: 149-169.

Karlan, D. (2007). Social connections and group banking. *Economic Journal* 117: 52-84.

Li, S., Liu, Y. and Deininger, K. (2012). How important are endogenous peer effects in group lending? Estimating a static game of incomplete information. *Journal of Applied Econometrics*, forthcoming.

Maes, J. and Reed, L. (2012). State of the microcredit summit campaign report 2012. Microcredit Summit Campaign.

Mahajan, A. (2006). Identification and estimation of regression models with misclassification. *Econometrica* 74: 631-665.

McFadden, D., Puig, C. and Kirschner, D. (1977). Determinants of the long-run demand for electricity. *Proceedings of the American Statistical Association* (*Business and Economics Section*): 109-117.

Paxton, J., Graham, D. and Thraen, C. (2000). Modeling group loan repayment behavior: New insights from Burkina Faso. *Economic Development and Cultural Change* 48: 639-655.

Shankar, S. (2006). Transaction costs in group micro credit in India: Case studies of three microfinance institutions. Centre for Microfinance, Institute for Financial and Management Research Working Paper, August.

Sharma, M. and Zeller, M. (1997). Repayment performance in group-based credit programs in Bangladesh: An empirical analysis. *World Development* 25: 1731-1742.

Stiglitz, J. (1990). Peer monitoring and credit markets. *World Bank Economic Review* 4: 351-366.

van Tassel, E. (1999). Group lending under asymmetric information. *Journal of Development Economics* 60: 3-25.

Varian, H. (1990). Monitoring agents with other agents. *Journal of Institutional and Theoretical Economics* 146: 153-174.

Wydick, B. (1999). Can social cohesion be harnessed to repair market failure? Evidence from group lending in Guatemala. *Economic Journal* 109: 463-475.

Zeller, M. (1998). Determinants of repayment performance in credit groups: The role of program design, intragroup risk pooling, and social cohesion." *Economic Development and Cultural Change* 46: 599-620.

Table 1
Summary statistics

Variable	Mean	Std. Dev.	Min	Max
Panel 1: Individual characteristics (12,883 observations)				
If defaulted	0.20	0.40	0.00	1.00
If literate	0.23	0.42	0.00	1.00
If disabled member in household	0.06	0.24	0.00	1.00
If owns land	0.65	0.48	0.00	1.00
If lives in pucca house	0.33	0.47	0.00	1.00
If lives in kacha house	0.22	0.42	0.00	1.00
If self-employed agricultural worker	0.20	0.40	0.00	1.00
If agricultural laborer	0.61	0.49	0.00	1.00
If belongs to scheduled tribe/caste	0.31	0.46	0.00	1.00
If belongs to leading caste	0.92	0.27	0.00	1.00
Panel 2: Group and loan characteristics (1,110 groups)				
Average member characteristics				
% literate	0.22	0.21	0.00	0.94
% disabled member in household	0.05	0.10	0.00	0.94
% own land	0.59	0.31	0.00	0.95
% live in pucca house	0.32	0.31	0.00	0.95
% live in kacha house	0.21	0.26	0.00	0.95
% self-employed agricultural worker	0.18	0.30	0.00	0.95
% agricultural laborer	0.56	0.36	0.00	0.95
% belong to scheduled tribe/caste	0.31	0.43	0.00	1.00
% belong to leading caste	0.91	0.14	0.36	1.00
Other group and village characteristics				
Age of group (years)	6.44	2.49	1.00	25.00
If group has food credit program	0.28	0.45	0.00	1.00
If group has marketing program	0.15	0.35	0.00	1.00
If group has insurance program	0.25	0.43	0.00	1.00
If group meets at least monthly	0.89	0.31	0.00	1.00
If located in Telangana	0.45	0.50	0.00	1.00
If located in Rayalaseema	0.26	0.44	0.00	1.00
If located in Coastal AP	0.29	0.45	0.00	1.00
Number of group members	12.52	2.37	7.00	20.00
If financial institution in village	0.34	0.47	0.00	1.00
If public bus in village	0.66	0.48	0.00	1.00
If telephone in village	0.75	0.43	0.00	1.00
If post office in village	0.63	0.48	0.00	1.00
Loan characteristics				
Amount of loan (rupees)	3,338	2,685	400	25,000
Number of members with loan	11.61	3.24	2.00	20.00
Annual interest rate (%)	12.83	3.10	6.00	25.00
Length of loan (years)	1.11	0.46	0.17	5.00
If repayment at least monthly	0.96	0.19	0.00	1.00
If loan due in 2004	0.11	0.31	0.00	1.00
If loan due in 2005	0.49	0.50	0.00	1.00
If loan due in 2006	0.40	0.49	0.00	1.00

Table 2
Intra-group default behavior

Default behavior	Groups	
	#	%
If none of the members defaulted	848	76.4
If all of the members defaulted	188	16.9
If some of the members defaulted	74	6.7
Total	1,110	100.0

Table 3
Probability of default, Two-type model

Variable	Type *H*		Type *L*	
	Coeff.	Std. Error	Coeff.	Std. Error
	Dependent variable: If default			
Constant	-3.399	0.629	7.775	28.740
If literate	0.160	0.105	0.540	0.206
If disabled member in household	0.258	0.163	-0.263	0.383
If owns land	0.180	0.119	-0.556	0.181
If lives in pucca house	-0.198	0.122	-0.997	0.186
If lives in kacha house	0.022	0.124	-0.844	0.209
If self-employed agricultural worker	-0.593	0.184	1.173	0.266
If agricultural laborer	0.120	0.140	1.748	0.155
If belongs to scheduled tribe/caste	0.082	0.110	2.736	0.279
If belongs to leading caste	-0.092	0.163	0.260	0.383
Amount of loan (1,000 rupees)	0.068	0.016	0.462	0.049
Number of members with loan	-0.062	0.090	-0.338	0.151
Number of members with loan squared	0.001	0.004	0.003	0.007
Annual interest rate (%)	0.083	0.013	0.277	0.034
Length of loan (years)	0.508	0.081	0.963	0.193
If repayment at least monthly	-0.497	0.244	-10.989	30.416
If loan due in 2005	-1.267	0.435	-0.128	0.287
If loan due in 2006	1.052	0.189	1.229	0.286
Probability of type-*H* Group				
Constant	-2.901	2.501		
% literate	1.921	0.409		
% disabled member in household	1.630	0.777		
% own land	0.707	0.212		
% live in pucca house	-1.124	0.276		
% live in kacha house	-1.052	0.228		
% self-employed agricultural worker	0.697	0.323		
% agricultural laborer	1.902	0.318		
% belong to scheduled tribe/caste	0.623	0.167		
% belong to leading caste	-1.020	0.496		
Age of group (years)	0.025	0.066		
Age of group squared	-0.004	0.004		
If group has food credit program	-0.951	0.115		
If group has marketing program	1.688	0.277		
If group has insurance program	0.443	0.139		
If group meets at least monthly	3.105	0.223		
If located in Telangana	2.320	0.255		
If located in Rayalaseema	0.652	0.211		

(Cont.)

Variable	Type *H*		Type *L*	
	Coeff.	Std. Error	Coeff.	Std. Error
	Dependent variable: If default			
Number of group members	0.132	0.360		
Number of group members squared	-0.014	0.014		
If financial institution in village	0.979	0.139		
If public bus in village	0.139	0.117		
If telephone in village	1.076	0.168		
If post office in village	-0.684	0.130		
Predicted probability of being Type-*H* group				
Average				79.8%
Group, no members defaulting				82.9%
Groups, all members defaulting				66.9%
Groups, some members defaulting				76.4%
Predicted individual default probability				
Average				19.6%
Conditional on being in Type-*H* group				9.5%
Conditional on being in Type-*L* group				62.8%
# observations				12,883
Log-likelihood				-5,111.6

Table 4
Conditional marginal effects (percentage points)

Variable	Type *H*			Type *L*		
	Mg. Effect	[95% Conf. Interv.]		Mg. Effect	[95% Conf. Interv.]	
Individual characteristics						
If literate	0.84	-0.14	1.81	7.33	2.39	11.57
If disabled member in household	1.44	-0.54	3.53	-4.21	-24.12	11.92
If owns land	0.89	0.23	1.69	-7.87	-13.13	-2.19
If lives in pucca house	-0.97	-1.91	-0.06	-16.44	-21.08	-9.58
If lives in kacha house	0.11	-0.78	1.19	-14.47	-21.46	-8.02
If self-employed agricultural worker	-2.57	-3.91	-1.19	13.95	7.65	18.10
If agricultural laborer	0.60	-0.72	1.82	29.16	19.65	36.86
If belongs to scheduled tribe/caste	0.42	-0.18	1.14	31.20	24.78	36.05
If belongs to leading caste	-0.48	-2.48	1.18	4.15	-8.23	14.55
Loan characteristics						
One thousand rupees increase in loan	0.36	0.22	0.50	5.92	4.08	6.88
One more member with loan	-0.23	-0.32	-0.13	-4.77	-7.24	-1.04
One-percent increase interest rate	0.44	0.32	0.52	3.77	2.39	4.68
One more year in length of loan	3.23	2.27	3.95	10.39	6.79	12.36
If repayment at least monthly	-3.08	-5.08	-1.11	-26.28	-35.23	-13.69
If loan due in 2005	-6.60	-8.33	-4.97	-1.91	-6.85	4.88
If loan due in 2006	6.03	4.10	7.43	17.05	12.08	20.68

Note: The marginal effects are calculated at the means of the covariates. For continuous variables, the corresponding change is indicated in the table. For discrete variables, the change is from 0 to 1. The confidence intervals reported are normal-based and biased-corrected using 200 bootstrap replications.

Table 5
Unconditional marginal effects (percentage points)

Variable	Probit model			Full Probit model			Two-type model		
	Mg. Effect	[95% Conf. Interv.]		Mg. Effect	[95% Conf. Interv.]		Mg. Effect	[95% Conf. Interv.]	
Individual characteristics									
If literate	-0.81	-2.01	0.51	-0.18	-1.84	1.58	1.56	0.54	2.50
If disabled member in household	-1.62	-4.01	0.72	-0.04	-3.15	3.21	0.82	-1.45	2.76
If owns land	-0.84	-1.71	0.27	0.18	-1.37	2.18	-0.08	-0.80	0.76
If lives in pucca house	-0.37	-1.48	0.64	-0.73	-2.86	1.22	-2.68	-3.67	-1.50
If lives in kacha house	2.82	1.43	4.26	-0.11	-2.30	2.19	-1.50	-2.74	-0.20
If self-employed agricultural worker	-0.37	-2.09	1.02	0.04	-3.25	2.76	-0.74	-2.13	0.51
If agricultural laborer	0.76	-0.67	2.02	0.59	-2.12	3.16	3.76	2.30	5.03
If belongs to scheduled tribe/caste	6.10	5.40	6.83	-1.98	-5.35	1.17	3.83	2.38	5.33
If belongs to leading caste	3.12	1.06	4.76	-0.23	-3.37	2.05	0.03	-1.94	1.51
Loan characteristics									
One thousand rupees increase in loan	1.60	1.46	1.76	1.45	1.30	1.63	0.97	0.77	1.11
One more member with loan	0.01	-0.14	0.16	0.15	-0.06	0.34	-0.74	-0.95	-0.37
One-percent increase interest rate	1.19	1.13	1.26	1.37	1.30	1.45	0.81	0.65	0.89
One more year in length of loan	7.90	7.47	8.26	8.31	7.90	8.69	4.02	3.21	4.48
If repayment at least monthly	-14.03	-15.83	-12.55	-6.78	-8.28	-5.51	-5.65	-7.60	-3.39
If loan due in 2005	-6.01	-6.59	-5.36	-5.84	-6.44	-5.14	-6.08	-7.17	-4.85
If loan due in 2006	9.52	8.90	10.18	10.64	9.97	11.35	7.25	5.55	8.39
Average member characteristics									
10-% increase literate				0.00	-0.21	0.21	-1.34	-1.66	-1.04
10-% increase disabled member				-0.94	-1.35	-0.56	-1.15	-1.64	-0.56
10-% increase own land				-0.51	-0.74	-0.33	-0.52	-0.80	-0.28
10-% increase pucca house				-0.12	-0.33	0.12	0.88	0.60	1.13
10-% increase kacha house				0.45	0.20	0.68	0.82	0.46	1.25
10-% increase self-employed ag. worker				0.12	-0.19	0.48	-0.51	-0.92	-0.05
10-% increase agricultural laborer				0.18	-0.11	0.47	-1.33	-1.64	-1.01
10-% increase scheduled tribe/caste				0.75	0.42	1.11	-0.46	-0.72	-0.28
10-% increase leading caste				0.49	0.24	0.85	0.80	0.29	1.53
Other group and village characteristics									
One more year of age of group				1.19	1.03	1.36	0.06	-0.21	0.37
If group has food credit program				8.08	7.67	8.57	8.46	4.94	13.33
If group has marketing program				-6.12	-6.49	-5.76	-8.36	-9.43	-7.51
If group has insurance program				-5.29	-5.75	-4.88	-3.07	-4.50	-2.20
If group meets at least monthly				-30.11	-30.88	-29.49	-44.59	-47.40	-42.51
If located in Telangana				-9.58	-10.03	-9.13	-18.01	-22.78	-13.68
If located in Rayalaseema				-2.79	-3.32	-2.28	-4.27	-5.33	-3.02
One more member in group				-1.41	-1.63	-1.15	1.27	0.60	1.73
If financial institution in village				-6.01	-6.39	-5.65	-6.59	-8.22	-5.45
If public bus in village				1.19	0.83	1.59	-1.06	-1.72	-0.12
If telephone in village				-3.43	-3.83	-3.01	-9.96	-11.56	-8.18
If post office in village				0.97	0.66	1.34	4.85	3.89	6.31

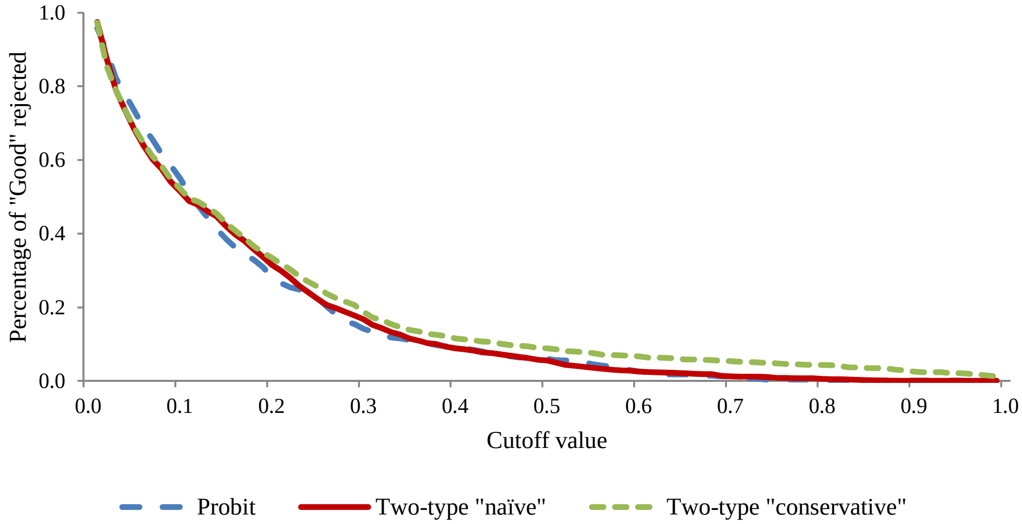
Note: The marginal effects are calculated at the means of the covariates. For continuous variables, the corresponding change is indicated in the table. For discrete variables, the change is from 0 to 1. The confidence intervals reported are normal-based and bias-corrected using 200 bootstrap replications.

Table 6
Predictive performance of alternative models

Indicator	Probit model	Full Probit model	Two-type "naïve"	Two-type "conservative"
	Out-of-sample performance (5,068 obs.)			
Average predicted default probability (observed=0.210)	0.185	0.186	0.199	0.237
Mean Square Predicted Error	0.160	0.159	0.145	0.156
Predictive performance	73.7%	74.7%	76.4%	76.0%
Correct default/non-default classification	77.9%	77.9%	79.2%	78.6%
Correct default classification (sensitivity), 1,062 defaults	2.2%	17.2%	21.9%	31.3%
Correct non-default classification (specificity), 4,006 non-defaults	98.0%	94.0%	94.4%	91.2%

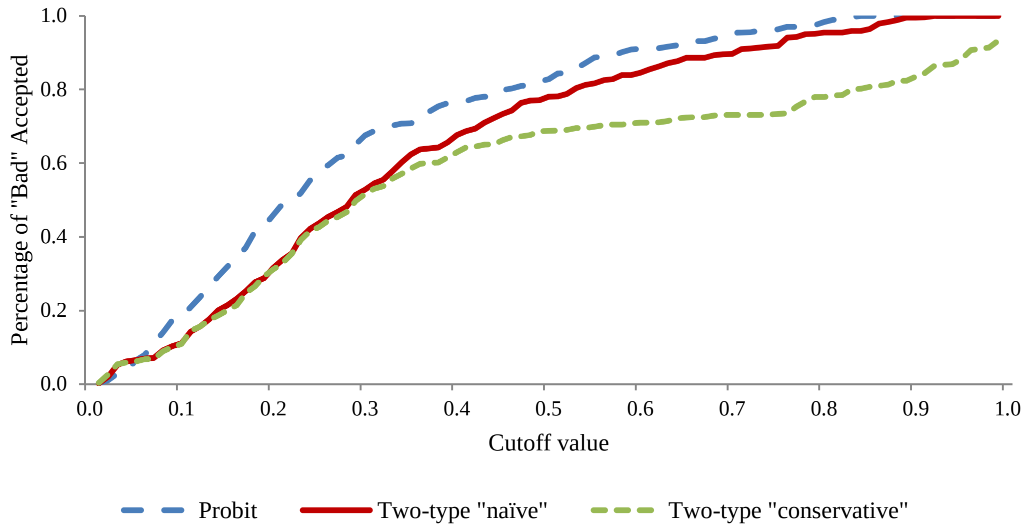
Note: The "naïve" approach is based on the unconditional probability of default of each individual. The "conservative" approach uses the probability of default based on the probability of individual of being in a particular group type. The performance and classification rates are based on converting the estimated default probabilities to a binary regime prediction using the standard 0.5 rule. The predictive performance measure is based on McFadden, Puig, & Kirschner (1977); the measure is equal to $p_{11} + p_{22} - p_{12}^2 - p_{21}^2$ where p_{ij} is the *ij*th entry in the standard 2x2 confusion matrix of actual versus predicted (0,1) outcomes in which the entries are expressed as a fraction of the sum of all entries. Sensitivity accounts for the percentage of cases in which individuals defaulting are also predicted to default, while specificity measures the percentage of cases in which individuals not defaulting are also predicted to not default. The results are based on 200 repeated 60-40% data partitions (averages reported).

Figure 1
Comparison of Type I errors



Note: The “naïve” approach is based on the unconditional probability of default of each individual. The “conservative” approach uses the probability of default based on the probability of individual of being in a particular group type. The results are based on 200 repeated 60-40% data partitions (averages reported).

Figure 2
Comparison of Type II errors



Note: The “naïve” approach is based on the unconditional probability of default of each individual. The “conservative” approach uses the probability of default based on the probability of individual of being in a particular group type. The results are based on 200 repeated 60-40% data partitions (averages reported).

Appendix A

Table A.1
Data description

Variable	Description
Default	If member failed to fully repay loan
Literate	If member can read and write
Disabled	If any household member has a disability
Own land	If member owns any land
Pucca house	If member lives in a house made of stone, bricks, concrete or timber
Kacha house	If member lives in a house made of hay, grass, mud or bamboo
Self-employed	If member is self-employed agricultural worker
Agricultural laborer	If member provides agricultural labor for someone else
Scheduled tribe/caste	If member belongs to a scheduled tribe or caste
Leading caste	If member belongs to a leading caste
Age of group	Group age in years
Food credit program	If group members receive a food credit program
Marketing program	If group members are provided with a marketing program
Insurance program	If group members are provided with an insurance program
Group meets at least monthly	If group members meet at least on a monthly basis
Located in Telangana	If the group is located in Telangana
Located in Rayalaseema	If the group is located in Rayalaseema
Located in Coastal AP	If the group is located in Coastal Andhra Pradesh
Number of group members	Number of members in the group
Financial institution in village	If there is a financial institution in the village
Public bus in village	If public bus service is available in the village
Telephone in village	If telephone service is available in the village
Post office in village	If there is a post office in the village
Amount of loan	Amount of loan borrowed by member in rupees
Number of members with loan	Number of members in the group who borrowed loan
Annual interest rate	Annual interest rate of the loan
Length of loan	Length of the loan in years
Monthly repayment frequency	If repayment frequency of the loan at least monthly
Loan due in 2004	If loan is due in 2004
Loan due in 2005	If loan is due in 2005
Loan due in 2006	If loan is due in 2006

Table A.2
Sorting based on observables

If intra-group variance is less than or equal to intra-village or intra-mandal variance by member characteristic				
	Intra-village		Intra-mandal	
	# groups	% total groups	# groups	% total groups
If literate	538	56.0	646	58.4
If disabled member in household	636	66.3	727	65.7
If owns land	606	63.1	755	68.2
If lives in pucca house	591	61.6	746	67.4
If lives in kacha house	627	65.3	768	69.4
If self-employed agricultural worker	761	79.3	866	78.2
If agricultural laborer	703	73.2	863	78.0
If belongs to scheduled tribe/caste	863	89.9	1018	92.0
If belongs to leading caste	680	70.8	763	68.9
Average	667	69.5	795	71.8

Note: The intra-village comparisons exclude 150 villages where there is only one group in the village, while the intra-mandal comparisons exclude 3 mandals.

Table A.3
Probability of default, One-type model

Variable	Probit model		Probit model full		Random-effects Probit model	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
	Dependent variable: If default					
Constant	-1.827	0.562	0.502	1.370	0.646	2.121
If literate	-0.032	0.055	-0.007	0.020	0.039	0.118
If disabled member in household	-0.065	0.080	-0.002	0.032	-0.030	0.194
If owns land	-0.033	0.071	0.008	0.026	0.010	0.139
If lives in pucca house	-0.015	0.082	-0.031	0.037	-0.189	0.149
If lives in kacha house	0.107	0.086	-0.005	0.047	0.056	0.152
If self-employed agricultural worker	-0.014	0.119	0.002	0.077	0.212	0.221
If agricultural laborer	0.030	0.089	0.025	0.056	0.247	0.173
If belongs to scheduled tribe/caste	0.229	0.094	-0.085	0.113	0.300	0.298
If belongs to leading caste	0.128	0.081	-0.009	0.036	0.288	0.204
Amount of loan (1,000 rupees)	0.061	0.016	0.059	0.018	0.071	0.021
Number of members with loan	0.009	0.072	0.065	0.095	0.126	0.195
Number of members with loan squared	0.000	0.003	-0.003	0.005	-0.003	0.010
Annual interest rate (%)	0.046	0.014	0.056	0.015	0.182	0.024
Length of loan (years)	0.274	0.108	0.304	0.113	0.867	0.168
If repayment at least monthly	-0.460	0.249	-0.256	0.269	-0.463	0.423
If loan due in 2005	-0.235	0.156	-0.247	0.163	-0.803	0.286
If loan due in 2006	0.359	0.158	0.430	0.170	1.138	0.275
% literate			0.000	0.237	-0.034	0.448
% disabled member in household			-0.405	0.426	-1.730	0.987
% own land			-0.220	0.178	-0.467	0.325
% live in pucca house			-0.051	0.190	0.050	0.353
% live in kacha house			0.189	0.216	0.584	0.380
% self-employed agricultural worker			0.050	0.259	-0.143	0.471
% agricultural laborer			0.074	0.188	-0.310	0.337
% belong to scheduled tribe/caste			0.310	0.160	0.337	0.356
% belong to leading caste			0.206	0.334	0.338	0.583
Age of group (years)			0.076	0.074	0.299	0.109
Age of group squared			-0.003	0.004	-0.012	0.006
If group has food credit program			0.319	0.108	0.910	0.172
If group has marketing program			-0.288	0.144	-0.775	0.236
If group has insurance program			-0.238	0.118	-0.513	0.193
If group meets at least monthly			-0.952	0.144	-2.935	0.203
If located in Telangana			-0.409	0.141	-1.140	0.227
If located in Rayalaseema			-0.122	0.154	-0.490	0.247

(Cont.)

Variable	Probit model		Probit model full		Random-effects Probit model	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
	Dependent variable: If default					
Number of group members			-0.346	0.204	-0.933	0.331
Number of group members squared			0.013	0.008	0.033	0.013
If financial institution in village			-0.266	0.116	-0.765	0.189
If public bus in village			0.051	0.103	0.152	0.166
If telephone in village			-0.140	0.119	-0.250	0.186
If post office in village			0.041	0.109	0.147	0.171
$\ln(\sigma^2 u)$					2.836	0.103
Rho					0.945	0.005
Predicted default probability		19.5%		19.5%		7.6%
# observations		12,883		12,883		12,883
Log likelihood		-5776.26		-5237.50		-1121.56

Note: The standard errors reported in the Probit model are robust, clustered by group. The $\ln(\sigma^2 u)$ term in the random-effects model represents the group-level variance component and Rho captures the proportion of the total variance contributed by the group-level variance component.

Table A.4
Two-type model exclusion tests: average member characteristics and group size and age

Variable	Excluding member characteristics				Excluding group size and age			
	Type *H*		Type *L*		Type *H*		Type *L*	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
	Dependent variable: If default				Dependent variable: If default			
Constant	-3.477	0.216	7.782	21.388	-3.492	0.339	7.570	29.829
If literate	-0.020	0.107	0.176	0.212	0.158	0.097	0.602	0.244
If disabled member in household	0.161	0.173	-0.563	0.334	0.253	0.155	-0.454	0.365
If owns land	-0.015	0.025	-0.787	0.235	0.073	0.058	-0.581	0.153
If lives in pucca house	-0.069	0.076	-0.713	0.212	-0.054	0.015	-1.092	0.180
If lives in kacha house	0.205	0.117	-0.542	0.207	0.102	0.116	-0.908	0.203
If self-employed agricultural worker	-0.592	0.164	1.451	0.331	-0.644	0.074	1.302	0.316
If agricultural laborer	-0.150	0.119	1.383	0.222	0.061	0.107	1.807	0.155
If belongs to scheduled tribe/caste	-0.060	0.103	2.454	0.196	-0.047	0.035	2.825	0.637
If belongs to leading caste	0.103	0.149	0.145	0.313	-0.132	0.125	0.231	0.433
Amount of loan (1,000 rupees)	0.101	0.018	0.564	0.091	0.090	0.016	0.463	0.050
Number of members with loan	-0.045	0.105	-0.344	0.125	-0.031	0.050	-0.318	0.242
Number of members with loan squared	0.001	0.005	0.004	0.004	0.001	0.002	0.003	0.010
Annual interest rate (%)	0.085	0.012	0.261	0.047	0.082	0.013	0.284	0.039
Length of loan (years)	0.585	0.086	1.012	0.351	0.555	0.020	0.825	0.358
If repayment at least monthly	-0.444	0.241	-10.982	21.541	-0.567	0.225	-11.081	31.719
If loan due in 2005	-1.269	0.546	-0.117	0.084	-1.351	0.409	-0.066	0.044
If loan due in 2006	0.730	0.160	1.607	0.346	0.894	0.157	1.233	0.289
Probability of type-*H* Group								
Constant	-2.777	5.297			-3.006	0.402		
% literate					2.055	0.530		
% disabled member in household					1.699	0.117		
% own land					0.469	0.230		
% live in pucca house					-1.005	0.136		
% live in kacha house					-0.992	0.215		
% self-employed agricultural worker					0.518	0.063		
% agricultural laborer					1.748	0.350		
% belong to scheduled tribe/caste					0.473	0.170		
% belong to leading caste					-1.149	0.509		
Age of group (years)	-0.036	0.064						
Age of group squared	0.000	0.003						
If group has food credit program	-0.814	0.129			-1.011	0.109		
If group has marketing program	1.397	0.218			1.567	0.306		
If group has insurance program	0.526	0.122			0.300	0.125		
If group meets at least monthly	3.036	0.303			2.963	0.262		
If located in Telangana	2.341	0.306			2.060	0.093		
If located in Rayalaseema	0.809	0.420			0.912	0.025		

(*Cont.*)

Variable	Excluding member characteristics				Excluding group size and age			
	Type *H*		Type *L*		Type *H*		Type *L*	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
	Dependent variable: If default				Dependent variable: If default			
Number of group members	0.156	0.747						
Number of group members squared	-0.010	0.029						
If financial institution in village	0.961	0.054			0.870	0.119		
If public bus in village	0.166	0.078			0.138	0.057		
If telephone in village	0.829	0.107			1.135	0.074		
If post office in village	-0.571	0.115			-0.535	0.104		
# observations				12,883				12,883
Log-likelihood				-5173.4				-5153.9

Table A.5
Two-type model exclusion tests: group programs and frequency of meetings

Variable	Excluding group programs				Excluding group meetings			
	Type *H*		Type *L*		Type *H*		Type *L*	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
	Dependent variable: If default				Dependent variable: If default			
Constant	-3.355	0.557	7.795	21.128	-3.461	0.635	7.572	35.427
If literate	0.095	0.106	0.426	0.200	0.081	0.113	0.429	0.218
If disabled member in household	0.254	0.161	-0.113	0.377	0.228	0.173	-0.167	0.444
If owns land	0.161	0.096	-0.278	0.125	0.100	0.111	-0.406	0.194
If lives in pucca house	-0.211	0.114	-0.992	0.203	-0.109	0.139	-1.043	0.204
If lives in kacha house	-0.130	0.127	-0.588	0.186	0.087	0.131	-0.948	0.219
If self-employed agricultural worker	-0.621	0.184	1.214	0.270	-0.638	0.204	1.103	0.285
If agricultural laborer	0.087	0.143	1.574	0.218	0.081	0.141	1.729	0.215
If belongs to scheduled tribe/caste	0.082	0.103	2.683	0.254	0.071	0.117	2.895	0.297
If belongs to leading caste	-0.086	0.154	0.149	0.412	-0.052	0.149	0.272	0.445
Amount of loan (1,000 rupees)	0.062	0.017	0.470	0.056	0.076	0.018	0.430	0.058
Number of members with loan	-0.050	0.069	-0.350	0.195	-0.067	0.077	-0.356	0.267
Number of members with loan squared	0.001	0.003	0.002	0.006	0.004	0.003	0.005	0.011
Annual interest rate (%)	0.087	0.012	0.270	0.038	0.083	0.013	0.283	0.040
Length of loan (years)	0.472	0.079	0.987	0.148	0.472	0.082	0.826	0.208
If repayment at least monthly	-0.548	0.237	-11.009	20.329	-0.681	0.222	-11.231	37.088
If loan due in 2005	-1.059	0.109	-0.070	0.204	-1.145	0.092	-0.207	0.243
If loan due in 2006	0.998	0.188	1.405	0.241	0.887	0.154	1.153	0.293
Probability of type-*H* Group								
Constant	-3.065	2.244			-2.674	0.881		
% literate	1.624	0.263			2.228	0.425		
% disabled member in household	1.353	0.584			1.881	0.637		
% own land	0.777	0.194			0.456	0.243		
% live in pucca house	-1.047	0.193			-0.971	0.229		
% live in kacha house	-0.907	0.223			-0.849	0.233		
% self-employed agricultural worker	1.018	0.258			0.448	0.324		
% agricultural laborer	1.585	0.245			2.053	0.241		
% belong to scheduled tribe/caste	0.707	0.167			0.634	0.172		
% belong to leading caste	-1.236	0.440			-0.885	0.516		
Age of group (years)	0.051	0.058			0.038	0.073		
Age of group squared	-0.004	0.003			-0.005	0.004		
If group has food credit program					-0.890	0.130		
If group has marketing program					1.441	0.185		
If group has insurance program					0.588	0.135		
If group meets at least monthly	2.972	0.117						
If located in Telangana	2.301	0.135			2.146	0.117		
If located in Rayalaseema	0.865	0.095			0.975	0.154		

(*Cont.*)

Variable	Excluding group programs				Excluding group meetings			
	Type *H*		Type *L*		Type *H*		Type *L*	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
	Dependent variable: If default				Dependent variable: If default			
Number of group members	0.148	0.312			0.080	0.144		
Number of group members squared	-0.014	0.013			-0.001	0.006		
If financial institution in village	0.769	0.135			1.237	0.129		
If public bus in village	0.209	0.108			0.358	0.123		
If telephone in village	1.163	0.099			1.322	0.162		
If post office in village	-0.627	0.075			-0.569	0.157		
# observations				12,883				12,883
Log-likelihood				-5223.0				-5418.5

Table A.6
Two-type model exclusion tests: group location and village characteristics

Variable	Excluding group location				Excluding village characteristics			
	Type *H*		Type *L*		Type *H*		Type *L*	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
	Dependent variable: If default				Dependent variable: If default			
Constant	-3.536	0.745	7.808	30.180	-3.423	0.537	7.796	49.901
If literate	0.153	0.094	0.679	0.226	0.212	0.102	0.344	0.233
If disabled member in household	0.177	0.125	-0.411	0.445	0.235	0.159	-0.409	0.409
If owns land	0.059	0.075	-0.738	0.098	0.180	0.094	-0.514	0.216
If lives in pucca house	-0.020	0.078	-0.814	0.370	-0.040	0.102	-0.921	0.193
If lives in kacha house	0.201	0.063	-0.875	0.220	0.070	0.117	-0.651	0.223
If self-employed agricultural worker	-0.521	0.113	1.319	0.312	-0.554	0.181	0.979	0.317
If agricultural laborer	-0.037	0.114	1.912	0.423	0.167	0.128	1.611	0.292
If belongs to scheduled tribe/caste	0.002	0.148	2.519	0.889	-0.047	0.103	2.825	0.473
If belongs to leading caste	-0.116	0.131	0.338	0.451	-0.044	0.156	0.246	0.441
Amount of loan (1,000 rupees)	0.104	0.025	0.519	0.199	0.056	0.016	0.512	0.048
Number of members with loan	-0.013	0.265	-0.328	0.587	-0.060	0.080	-0.336	0.164
Number of members with loan squared	0.000	0.013	0.007	0.019	0.001	0.004	0.002	0.006
Annual interest rate (%)	0.094	0.011	0.234	0.040	0.082	0.013	0.265	0.038
Length of loan (years)	0.583	0.053	0.596	0.300	0.611	0.096	0.881	0.277
If repayment at least monthly	-0.548	0.222	-10.967	33.708	-0.533	0.263	-10.956	50.313
If loan due in 2005	-1.237	0.234	-0.102	0.240	-0.992	0.170	-0.221	0.241
If loan due in 2006	0.811	0.140	0.998	0.514	0.910	0.180	1.141	0.288
Probability of type-*H* Group								
Constant	-2.742	9.305			-2.637	0.603		
% literate	1.759	0.206			2.138	0.370		
% disabled member in household	1.858	0.352			1.843	0.577		
% own land	0.881	0.210			0.479	0.243		
% live in pucca house	-1.300	0.580			-0.820	0.241		
% live in kacha house	-1.322	0.215			-0.992	0.230		
% self-employed agricultural worker	0.857	0.095			0.424	0.308		
% agricultural laborer	1.722	0.435			1.925	0.233		
% belong to scheduled tribe/caste	0.618	0.508			0.327	0.159		
% belong to leading caste	-0.957	0.174			-0.869	0.475		
Age of group (years)	0.033	0.097			0.022	0.068		
Age of group squared	-0.006	0.006			-0.003	0.004		
If group has food credit program	-1.205	0.066			-0.934	0.124		
If group has marketing program	1.813	0.108			1.622	0.187		
If group has insurance program	0.132	0.064			0.687	0.160		
If group meets at least monthly	2.822	0.449			3.220	0.154		
If located in Telangana					1.994	0.180		
If located in Rayalaseema					0.826	0.079		

(*Cont.*)

Variable	Excluding group location				Excluding village variables			
	Type *H*		Type *L*		Type *H*		Type *L*	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
	Dependent variable: If default				Dependent variable: If default			
Number of group members	0.090	1.566			0.153	0.166		
Number of group members squared	-0.003	0.067			-0.014	0.007		
If financial institution in village	0.711	0.131						
If public bus in village	0.421	0.258						
If telephone in village	1.207	0.057						
If post office in village	-0.411	0.118						
# observations				12,883				12,883
Log-likelihood				-5236.6				-5191.6

Table A.7
Hausman tests: Baseline model versus alternative specifications

Variables excluded	H_0: Difference in coefficients of repayment equation between baseline model and alternative specifications not systematic
Average member characteristics	16.610
	(0.165)
Group size and age	31.648
	(0.084)
Group programs	12.402
	(0.574)
Frequency of group meetings	32.087
	(0.076)
Group location	11.307
	(0.662)
Village characteristics	45.828
	(0.000)

Note: Hausman Chi-squared statistics reported and *p*-values in parenthesis.

Figure A.1
Histogram of percentage of loan repaid by member

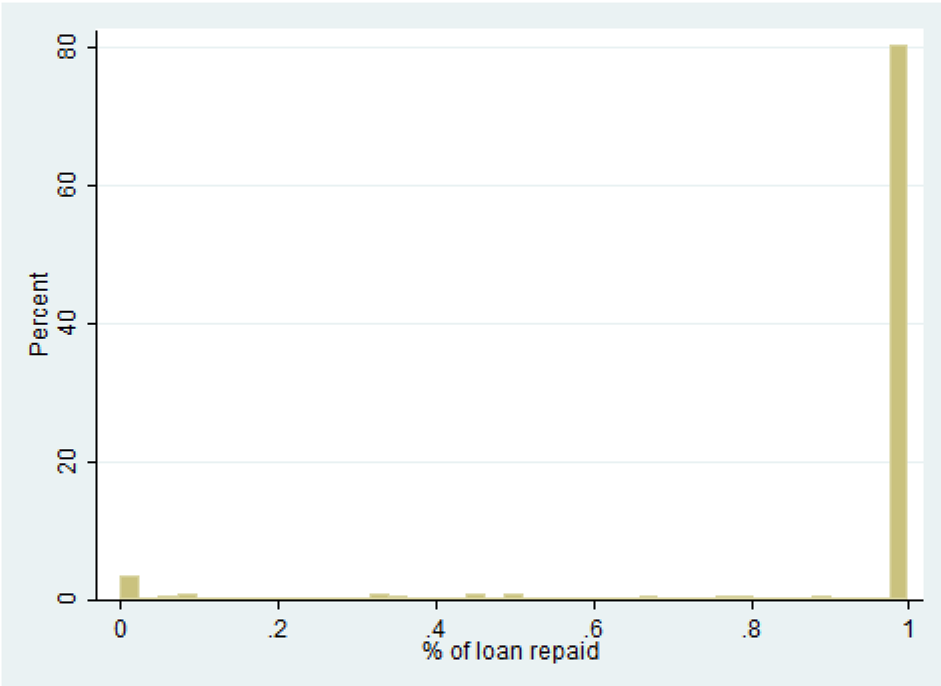
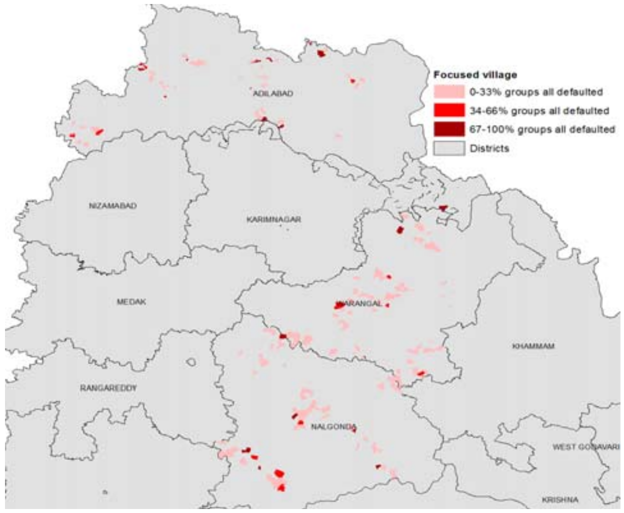
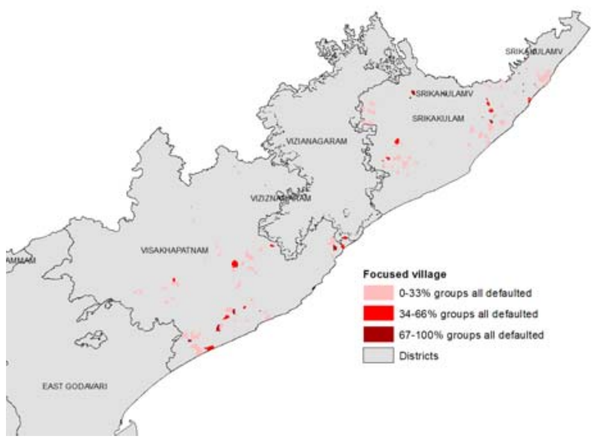


Figure A.2
Location of villages in Andhra Pradesh and group default behavior

North



North East



South

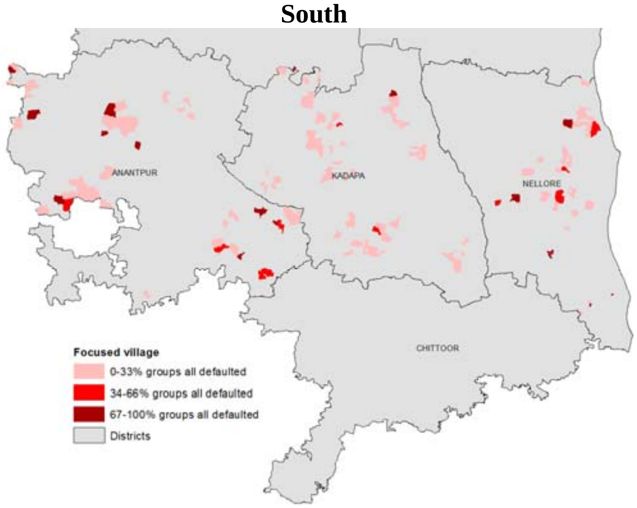
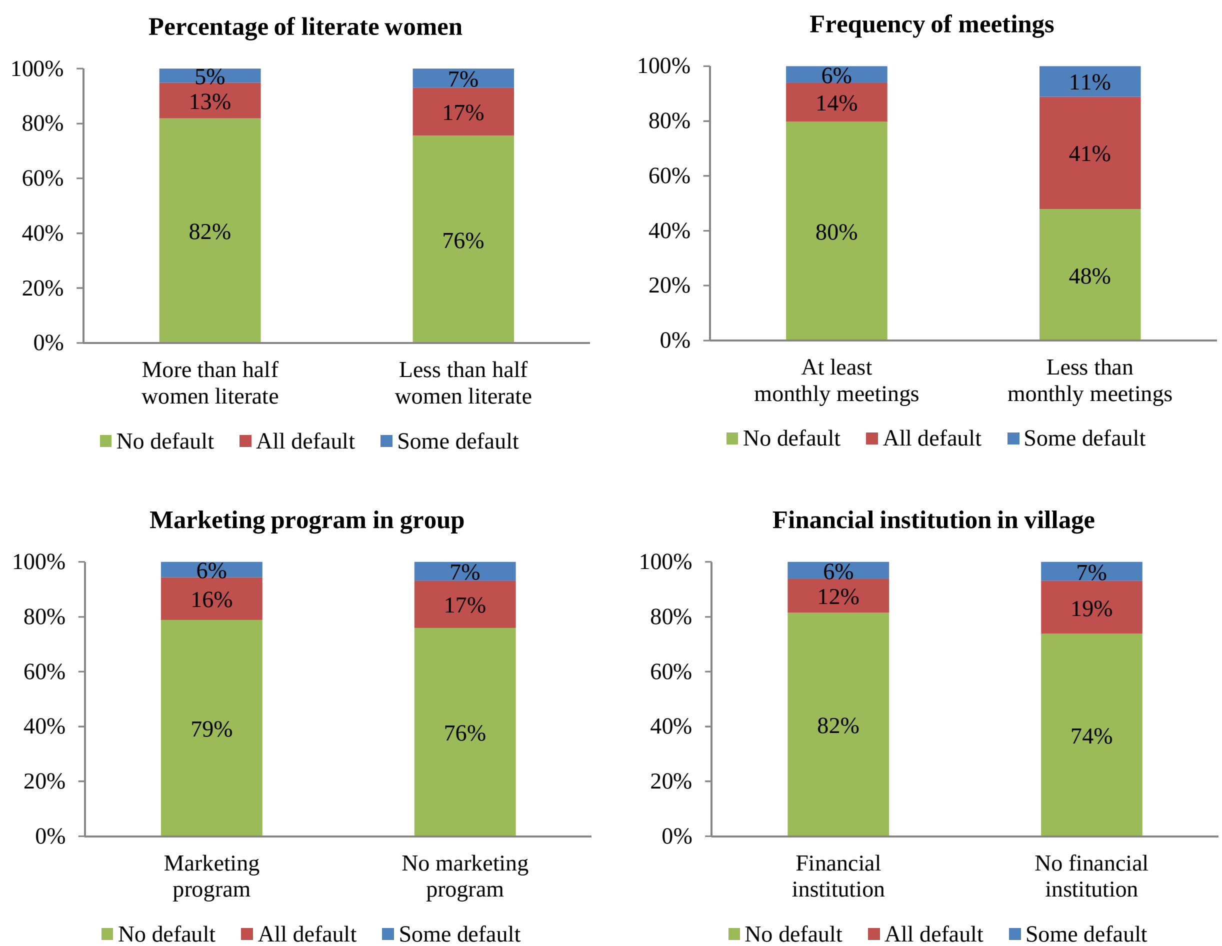


Figure A.3
Distribution of intra-group default behavior by different group characteristics



Appendix B: Solution to model with peer selection and moral hazard

In this appendix, we solve the model with peer selection and moral hazard outlined in Section 2. The model is an extension of Ghatak (1999) basic model, where we allow borrowers to vary on their risk type and effort level. The model setup is presented in Section 2. We also assume that $Y > r + q$, i.e. a successful borrower can make a profit even when her partner loses. This assumption is innocuous because if it does not hold, a borrower with a failed project may have a higher payoff than one with a successful project, which is an unreasonable scenario. We consider both a non-cooperative game scenario where each borrower maximizes her own payoff and a cooperative game scenario where matched borrowers maximize the total payoff of their group.

In the non-cooperative game setting, the maximization problems of the matched borrowers are given by

$$\max_{e_i} E\pi_{ij} = (p_i + e_i)Y - (p_i + e_i)r - q(p_i + e_i)(1 - p_j - e_j) - 1/2\gamma e_i^2$$

$$\max_{e_j} E\pi_{ji} = (p_j + e_j)Y - (p_j + e_j)r - q(p_j + e_j)(1 - p_i - e_i) - 1/2\gamma e_j^2$$

$$s.t.\ \ e_i \geq 0, \quad e_j \geq 0\,.$$

The first order conditions (FOCs) are:

$$\partial E\pi_{ij} / \partial e_i = Y - r - q(1 - p_j - e_j) - \gamma e_i \leq 0$$

$$\partial E\pi_{ji} / \partial e_j = Y - r - q(1 - p_i - e_i) - \gamma e_j \leq 0$$

$$e_i \geq 0$$

$$e_j \geq 0$$

$$e_i[Y - r - q(1 - p_j - e_j) - \gamma e_i] = 0$$

$$e_j[Y - r - q(1 - p_i - e_i) - \gamma e_j] = 0\,.$$

Solving the FOCs, we have

$$e_{ij}^* = 0, \text{ if } \ \gamma \leq q$$

$$e_{ij}^* = \frac{(\gamma + q)(Y - r) - q[q(1 - p_i) + \gamma(1 - p_j)]}{\gamma^2 - q^2} \text{ if } \gamma > q .$$

We change the subindex of effort from i to ij because the optimal effort of borrower i depends not only on her own type but also on the type of her partner. To eliminate the corner solution under which the second order condition (SOC) is violated, we assume $\gamma > q$. Hereafter we only consider the interior solution. We note that the SOC of the internal solution is satisfied and we have

$$e_{bb}^* > e_{ab}^* > e_{ba}^* > e_{aa}^* .$$

The above result suggests that a borrower's optimal effort level is higher if she is a safe type and/or if her partner is a safe type.

Substituting e_{ij}^* into $E\pi_{ij}$ and denoting *M*, *A* and *B* as

$$M = Y - r - q ,$$

$$A = e_{bb}^* - e_{ba}^* = e_{ab}^* - e_{aa}^* = \frac{\gamma q(p_b - p_a)}{\gamma^2 - q^2},$$

$$B = e_{ba}^* - e_{aa}^* = e_{bb}^* - e_{ab}^* = \frac{q^2(p_b - p_a)}{\gamma^2 - q^2},$$

we obtain

$$\begin{aligned} E\pi_{bb}^* - E\pi_{ba}^* &= AM + qp_b(p_b - p_a) + qp_b B + q(e_{bb} p_b - e_{ba} p_a + e_{bb}^2 - e_{ba} e_{ab}) - 0.5\gamma A(e_{bb} + e_{ba}) \\ &> AM + qp_b(p_b - p_a) + qp_b B + qe_{bb}(p_b - p_a) - (\gamma - q) A e_{bb} \\ &= AM + qp_b(p_b - p_a) + qp_b B + q^2(p_b - p_a) e_{bb} /(\gamma + q) > 0, \end{aligned}$$

and

$$
\begin{aligned}
E\pi^*_{ab} - E\pi^*_{aa} &= AM + qp_a(p_b - p_a) + qp_a B + q(e_{ab} p_b - e_{aa} p_a - e^2_{aa} + e_{ba} e_{ab}) - 0.5\gamma A(e_{ab} + e_{aa}) \\
&> AM + qp_a(p_b - p_a) + qp_a B + qe_{ab}(p_b - p_a) - (\gamma A - qB)e_{ab} \\
&= AM + qp_b(p_b - p_a) + qp_b B > 0.
\end{aligned}
$$

The above results suggest that a borrower prefers a safer partner despite of her own type.

We then examine if positive assortative matching is the only equilibrium. Following Ghatak (2009, such equilibrium must satisfy the optimal sorting property (Becker, 1993). That is, the net expected loss for a safe borrower of having a risky borrower is higher than the net expected gain for a risky borrower of having a safe partner. Therefore, a risky borrower does not have sufficient incentives to pay enough money to a safe borrower to match with her. We find

$$
\begin{aligned}
(E\pi^*_{bb} - E\pi^*_{ba}) - (E\pi^*_{ab} - E\pi^*_{aa}) &= q(p_b - p_a)^2 + 2q(p_b - p_a)B - \gamma AB + q(e^2_{bb} + e^2_{aa} - 2e_{ba} e_{ab}) \\
&= q\gamma^4 (p_b - p_a)^2 / (\gamma^2 - q^2)^2 > 0.
\end{aligned}
$$

Consistent with Proposition 1 in Ghatak (1999), this result suggests positive assortative matching is the only equilibrium.

Next, we keep the same model setup but assume a cooperative game setting where matched borrowers maximize their joint payoff given by

$$
\begin{aligned}
\max_{e_i, e_j}(E\pi_{ij} + E\pi_{ji} &= (p_i + e_i)Y - (p_i + e_i)r - q(p_i + e_i)(1 - p_j - e_j) - 1/2\gamma e_i^2 \\
&+ (p_j + e_j)Y - (p_j + e_j)r - q(p_j + e_j)(1 - p_i - e_i) - 1/2\gamma e_j^2
\end{aligned}
$$

$$s.t.\ \ e_i \geq 0, \quad e_j \geq 0\,.$$

The FOCs are:

$$\partial E\pi_{ij} / \partial e_i = Y - r - q + 2q(p_j + e_j) - \gamma e_i \leq 0$$

$$\partial E\pi_{ji} / \partial e_j = Y - r - q + 2q(p_i + e_i) - \gamma e_j \leq 0$$

$$e_i \geq 0$$

$$e_j \geq 0$$

$$e_i[Y - r - q + 2q(p_j + e_j) - \gamma e_i] = 0$$

$$e_j[Y - r - q + 2q(p_i + e_i) - \gamma e_j] = 0 .$$

Solving the FOCs, we have

$$e_{ij}^* = 0 \text{ if } \gamma \le 2q$$

$$e_{ij}^* = \frac{(\gamma + 2q)(Y - r - q) + 2q[2qp_i + \gamma p_j)]}{\gamma^2 - 4q^2} \text{ if } \gamma > 2q .$$

We impose the assumption $\gamma > 2q$ to eliminate the corner solution. For the interior solution, the SOC is satisfied. Similar to the non-cooperative game, we obtain

$$e_{bb}^* > e_{ab}^* > e_{ba}^* > e_{aa}^* .$$

We next prove that a group with two safe borrowers has a higher joint payoff than a group with one safe borrower and one risky borrower. Plugging e_{ij}^* into $E\pi_{ij}$, we have

$$E\pi_{bb}^* - E\pi_{ba}^* = A'M + qp_b(p_b - p_a) + qp_b B' + q(e_{bb}p_b - e_{ba}p_a) + q(e_{bb}^2 - e_{ba}e_{ab}) - 0.5\gamma(e_{bb}^2 - e_{ba}^2),$$

$$E\pi_{bb}^* - E\pi_{ab}^* = B'M + (p_b - p_a)M + qp_b(p_b - p_a) + q(2p_b e_{bb} - p_a e_{ba} - p_a e_{ab}) + q(e_{bb}^2 - e_{ba}e_{ab}) - 0.5\gamma(e_{bb}^2 - e_{ab}^2),$$

where $A' = e_{bb}^* - e_{ba}^* = e_{ab}^* - e_{aa}^* = \dfrac{2\gamma q(p_b - p_a)}{\gamma^2 - 4q^2}$, $B' = e_{ba}^* - e_{aa}^* = e_{bb}^* - e_{ab}^* = \dfrac{4q^2(p_b - p_a)}{\gamma^2 - 4q^2}$. We note

$$\begin{aligned}
&q(e_{bb}^2 - e_{ba}e_{ab}) - 0.5\gamma(e_{bb}^2 - e_{ba}^2) + q(e_{bb}^2 - e_{ba}e_{ab}) - 0.5\gamma(e_{bb}^2 - e_{ab}^2) \\
&> 2q(e_{bb}^2 - e_{ba}^2 - e_{ab}^2) - \gamma e_{bb}^2 + 0.5\gamma(e_{ab}^2 + e_{ba}^2) \\
&= -0.5(\gamma - 2q)A(e_{bb} + e_{ba}) - 0.5(\gamma - 2q)B(e_{bb} + e_{ab}) \\
&> -(\gamma - 2q)(A + B)e_{bb} = -(A + B)M - 2qp_b(A + B).
\end{aligned}$$

Then,

$$
\begin{aligned}
&2E\pi_{bb}^{*} - (E\pi_{ab}^{*} + E\pi_{ba}^{*}) \\
&> (A+B)M + 2qp_b(p_b - p_a) + (p_b - p_a)M + qp_bB + q(e_{bb}p_b - e_{ba}p_a) + \\
&\quad +q(2p_be_{bb} - p_ae_{ba} - p_ae_{ab}) - (A+B)M - 2qp_b(A+B) \\
&= 2qp_b(p_b - p_a) + (p_b - p_a)M + q(e_{bb}p_b - e_{ba}p_a) + q(2p_be_{bb} - p_ae_{ba} - p_ae_{ab}) - 2qp_bA - qp_bB \\
&> 2qp_b(p_b - p_a) + (p_b - p_a)M + qp_bA + qp_bA + qp_bB - 2qp_bA - qp_bB \\
&= 2qp_b(p_b - p_a) + (p_b - p_a)M > 0
\end{aligned}
$$

Therefore, a safe borrower will prefer a safe to a risky borrower.

We finally examine if positive assortive matching is the only equilibrium, which is implied by

$$2E\pi_{bb} + 2E\pi_{aa} - 2(E\pi_{ba} + E\pi_{ab}) > 0\,.$$

We have

$$
\begin{aligned}
E\pi_{bb} + E\pi_{aa} - (E\pi_{ba} + E\pi_{ab}) &= (E\pi_{bb}^{*} - E\pi_{ba}^{*}) - (E\pi_{ab}^{*} - E\pi_{aa}^{*}) \\
&= q(p_b - p_a)^2 + 2q(p_b - p_a)B' - \gamma A'B' + q(e_{bb}^2 + e_{aa}^2 - 2e_{ba}e_{ab}) \\
&= q(p_b - p_a)^2 + B'[2q(p_b - p_a) - \gamma A'] + \frac{4q^5(\gamma^2 + 4q^2)(p_b - p_a)^2}{(\gamma^2 - 4q^2)^2} \\
&= q(p_b - p_a)^2 - \frac{32q^5(p_b - p_a)^2}{(\gamma^2 - 4q^2)^2} + \frac{4q^5(\gamma^2 + 4q^2)(p_b - p_a)^2}{(\gamma^2 - 4q^2)^2} \\
&= q(p_b - p_a)^2\left[1 + \frac{-32q^4 + (\gamma^2 + 4q^2)4q^2}{(\gamma^2 - 4q^2)^2}\right] \\
&= q\gamma^2(p_b - p_a)^2/(\gamma^2 - 4q^2) > 0
\end{aligned}
$$

This result indicates that the model also leads to positive assortative matching in the cooperative game setting.