Risk-Matching in Credit Groups: Evidence from Guatemala.

Loïc Sadoulet* and Seth Carpenter†

First Version: June 1999
This Version: October 1999‡

Abstract

With widely publicized high repayment rates, microfinance is gaining a great deal of attention. Using data from Guatemala, this paper examines risk matching in credit groups. The literature often *assumes* that joint-liability will lead groups to form homogeneously in risk, and that risk heterogeneity emerges only as a second-best. We find they do not, even accounting for matching frictions. Data on mutual-help within groups provides evidence consistent with the hypothesis that group lending provides insurance among borrowers. This intra-group insurance suggests that current credit contracts can be improved by incorporating insurance provisions. We discuss one possibility of such a contract briefly.

*Corresponding author. Address: ECARE, Université Libre de Bruxelles, Ave. F. Roosevelt 39, B-1050 Brussels, Belgium. email: sadoulet@ulb.ac.be. The latest version of this paper is available from http://homepages.ulb.ac.be/~sadoulet/papers.html

†College of William and Mary, and Federal Reserve Board. email: seth.b.carpenter@frb.gov

‡We thank Anne Case, Robert Shimer, and the participants at seminars at ECARE, Université de Liège, INSEAD, and the participants at the Third Annual Seminar on New Development Finance in Frankfurt for very valuable comments and suggestions. Funding for the survey in Guatemala was generously provided by the Mellon and the Ford Foundations. The Financial Research Center at Princeton University funded the costs of additional data entry and cleaning costs. The authors are solely responsible for the views expressed and for any eventual mistakes remaining in this paper.

1 Introduction

Impressive repayment rates and successful outreach of programs such as the *Grameen Bank* in Bangladesh or *BancoSol* in Bolivia have given microfinance an increased amount of attention as a mechanism to extend credit to the working poor.[1] Not all programs have enjoyed the same success, however. The varying records are arguably due to the setup of the programs themselves rather to their operational environments (Huppi and Feder, 1990; Braverman and Guasch, 1993). Therefore, there is an acute interest, in assessing the mechanisms at work in microfinance contracts to design the most appropriate programs.

Microfinance institutions recognize that the poor, by definition, have few assets for collateral. Instead, borrowers are offered a sequence of loans that are cheaper than their alternative credit sources, typically local money-lenders. Repayment of loans gives access to further loans; default excludes the borrower permanently from future loans from the lending institution. The incentive to repay relies on the benefit of future loans, relative to the cost of repaying the current one.[2]

Group lending with joint liability has gained particular attention to circumvent the risk and uncertainty of lending to the poor. Borrowers find partners, each partner receives a loan, and all partners are *jointly liable* for all the loans in their group. If *any* member defaults, the *entire* group is considered in default.[3] By using local information to form their groups, borrowers have better information on their partners' risks than the lending institution does. Risk has thus become a central theme of academic research into group lending mechanisms. In spite of this, both theoretical and empirical work have assumed, explicitly or implicitly, that groups will be composed of members of equal risk and focus on group performance defined by such things as repayment rates, sustainability or outreach. In the theoretical literature, the

[1] Morduch (1997) reports several examples of multi-million dollar funds and programs established. In 1993, the World Bank created a program entitled *Sustainable Banking with the Poor* with the aim of producing a "best practices" manual for microfinance institutions.

[2] To increase the attractiveness of future loans, lending institutions typically increase loan sizes, reduce interest rates, or offer letters of recommendation upon graduation to the formal banking sector.

[3] Though neither the academic literature nor the popular press has distinguished them from one another, there exist two prevalent group-lending methodologies: the "ACCION model," in which all borrowers of a credit group receive their loans simultaneously; and the "Grameen model," according to which borrowers of credit receive their loans sequentially, conditional on the good repayment of their partners (see Hossain, 1988). All the models that we are familiar with have only examined the case in which all members of a credit group receive their loans simultaneously, and this paper continues in that line since our data comes from an institution lending according to the "ACCION model." Modeling rigorously the mechanisms at play in the "Grameen model" would be an interesting project.

results are mixed. Some authors find that joint liability increases repayment rates (Stiglitz, 1990; Ghatak, 1995; Armendáriz, 1995; Wydick, 1995; Sadoulet, 1997), while others find that it can reduce repayment rates (Besley and Coate, 1995; Diagne, 1998; Sadoulet, 1997). Similarly, empirical papers have tried to estimate the importance of screening, monitoring, and social sanctions for repayment rates, and have reached mixed conclusions. Some find that proximity and social cohesiveness are linked with higher repayment rates (Sharma and Zeller, 1997; Zeller, 1998), while others find the opposite result (Wydick, 1999).

In both of these literature, rarely have authors taken into account that borrowers *choose* the members of their credit groups.[4] Ignoring this endogeneity can lead researchers to flawed conclusions and misguided policy recommendations. In this paper, we examine borrowers' choices of partners using data from a 1995 survey of 250 credit groups in Guatemala. In particular, we test the hypothesis of homogeneous matching prevalent in the literature. Risk-heterogeneity is often assumed to emerge only as a second-best due to *matching frictions* that keep borrowers from finding their first-best partners. These frictions include limited partner availability, informational problems which restrict borrowers' monitoring ability, social codes restricting enforcement sanctions, or characteristics which impede borrowers' credibility in promising or requiring transfers. After accounting for matching frictions, we test if groups *choose* risk heterogeneity, and suggest an explanation as to *why*.

We estimate the relationship between individual risk and the level of risk heterogeneity in the individual's group, explicitly accounting for the endogeneity of group formation and of borrowers' choice of project risk. The results show that, even when allowing for matching frictions, a strong, systematic relationship exists between risk and the risk heterogeneity. Safe borrowers tend to choose partners substantially riskier than themselves, rather than borrowers of risk similar to their own. Risky borrowers manage to attract partners safer than themselves to their group. We show that this result is not due to the inherent availability of potential partners, unless the distribution of types were extremely skewed locally – a characteristic we find very unlikely in our context. Our results suggest that borrowers appear to be *choosing* risk-heterogeneity when forming or joining groups.

As a possible explanation of heterogeneity in groups, we find evidence of intra-group insurance. Theoretical models of Sadoulet (1997, 1998) show the joint liability and repeated

[4] Three exceptions being Armendáriz and Gollier (1999), Ghatak (1995), and Sadoulet (1998).

nature of these loans allows for credible insurance arrangements. Insurance can encourage groups of heterogeneous risks, not through a reduction of covariance risk, but by a transfer of the safe borrowers' lower absolute risk to risky partners in exchange for a payment.

The results in the current paper are interesting in several dimensions. First, the fact that borrowers choose their group partners emphasizes the fact that group characteristics are endogenous. Group composition and repayment rates are simultaneously determined, so using group characteristics to explain repayment performance leads to inconsistent estimation. Empirical studies need to account for this endogeneity.

Second, this paper's matching results are consistent with the idea that group lending programs can provide important insurance mechanisms for borrowers. Better lending contracts could be designed, with explicit insurance provisions. In particular, contracts incorporating credit records (such as credit card contracts in the US) could allow insurance to be provided by the better diversified lending institution.

The results are also interesting beyond the development finance literature. The endogeneity of agents' choice has received much attention recently. The matching literature has added market imperfections and other matching frictions to the classic model to examine how these affect equilibrium matching patterns (e.g.: Kremer and Maskin, 1995; Legros and Newman, 1998; and Shimer and Smith, 1998). Ackerberg and Botticini (1998) examine empirically how risk aversion, and not simply the inherent riskiness of crops, affected the contractual choice in 15th century Italy. The current paper complements that literature by documenting a non-trivial matching pattern and by careful treatment of the endogeneity.

The remainder of the paper is organized as follows. Section 2 discusses microfinance contracts and summarizes the issues in credit group formation. Section 3 elaborates an econometric model and the methodology to test the hypothesis of homogeneous risk-matching. Section 4 presents the data and constructs the empirical measures of the variables in the theoretical model. Section 5 presents the results, suggesting a strong rejection of homogeneous matching, even accounting for matching frictions. In section 6, we document of intragroup risk sharing as a possible explanation of group heterogeneity and explore intra-group characteristics. The last section concludes.

2 Group formation in microfinance contracts

2.1 Microfinance contracts

Lending institutions typically have limited information on the quality and behavior of their clients, especially when the borrowers are small informal businesses, so-called *microentreprises*. As such, collateral is a standard mechanism to avoid adverse selection and moral hazard problems in loan markets. Collateralized loans, however, exclude resource-poor entrepreneurs, and these microentrepreneurs either do not borrow or rely on the informal money-lending sector, often at usurious interest rates.[5]

Financial institutions aiming at integrating the poor into the formal financial sector have been relying on new lending contracts to circumvent collateral requirements. These lending contracts have been offered in two variants: group loans and individual loans. Both types of contracts rely on the following two fundamental principles: borrowers are offered a sequence of uncollateralized loans, at a much lower interest rate than offered by their alternative credit sources (typically the afore-mentioned money-lenders); and borrowers are guaranteed continued access to these loans upon repayment, but lose access upon default. Borrowers' incentive to repay is induced by their desire to maintain access to this source of cheap loans, rather than by the threat of losing collateral as in traditional lending contracts. As explained in the introduction above, the difference between group loans and individual loans resides in the definition of what constitutes a default. In an individual loan, a borrower is in default when s/he does not repay the loan. In group lending, however, the members of a credit group are *jointly liable*: if *any* part of the loans to the group is not repaid, *all* members of the group are considered in default and lose access to future loans.[6]

For financial institutions, one motivating factor behind the practice of group loans is that local borrowers have better information than they on the quality and actions of other borrowers, and borrowers may have access to local enforcement mechanisms (often referred to as *social collateral* in the literature) to sanction opportunistic behavior in their credit groups. Joint liability in group contracts ties borrowers' expected returns to their partners' behavior, inducing a selection process, peer monitoring, and modifications in behavior among

[5]See Aleem (1990) for evidence on the costs of money-lender loans, and the Associated Press release (1998) on "payday" loans in the US.

[6]Typically, these successive loan amounts are increasing to discourage borrowers from defaulting and subsequently self-financing.

borrowers. The selection process and the emergence of possible insurance among groups may increase repayment rates, as compared to loan provision through individual contracts solely.[7]

The impact of joint liability on repayment behavior has been studied in the group lending literature. In the theoretical branch of the literature, Stiglitz (1990) posits that peer pressure and joint liability force the coordination of choice of projects and of repayment behavior among borrowers. Joint liability in group loans entices borrowers to choose lower risk projects because of the additional risk imposed on them by their partners' loans, which increases borrowers' *ability* to repay. With similar assumptions on borrowers' ability to control strategic defaults within groups, Ghatak (1995) shows that joint liability can increase repayment rates by attracting safe borrowers back into the credit market. Besley and Coate (1995) examine how the availability of sufficiently severe local sanctions can allow borrowers to punish partners who willfully default on their loans. Joint liability, in combination with such social sanctions, decreases the proportion of defaults originating from borrowers' *unwillingness* to repay. Armendáriz de Aghion (1995) examines the incentives to incur costly monitoring in credit contracts, and how the relative benefit of peer monitoring on repayment rates can be affected by the covariance structure of projects, the size of credit groups, and the monitoring technology.

Similarly, the empirical literature has examined the impact of group characteristics on repayment rates. Studies have characterized groups by their differential screening, monitoring, risk diversification, and enforcement abilities. The conclusions have been that groups that screen have lower delinquency rates (Wenner, 1995). Repayment rates increase with the proximity and social cohesiveness of groups (Zeller, 1998), and with the size and diversification of activities in groups (Sharma and Zeller, 1997). The ability to exert peer pressure is also found to be an important determinant of repayment rates (Sharma and Zeller, 1997), although Wydick (1999, p. 474) finds that social ties within credit groups affect group performance to a "surprisingly small degree".

However, none of these papers, theoretical or empirical, addresses the endogeneity of group composition. In particular, borrowers should self-select out of problematic groups if, for example, they are faced with partners who strategically default as in Besley and Coate (1995).

[7] Another motivating factor for financial institutions is the possible reduction of transactions costs. If financial institutions can reduce processing costs when offering group rather than individual contracts, it allows them to serve clients to whom it would be prohibitively costly to extend small loans to otherwise. Whether there are scale economies in loan processing is still subject to virulent debate among practitioners.

Or borrowers wanting to default would not join a group in which other members could pressure them into repayment; they would form a group with other borrowers wanting to default and collude against the financial institution. The endogeneity of group formation can, in particular, explain what Wydick (1999, p.471) refers to as a "mysterious" negative relationship between social ties and the intra-group insurance provided to prevent a default. If enforcement through social collateral is more expensive and less effective than direct enforcement mechanism (as is documented in Diagne, 1998), groups having to resort to such methods will be faced with higher default rates. The endogenous choice of group membership and of repayment (and possibly insurance) strategies is hence clearly important for the analysis of repayment performance.

2.2 Do credit groups match homogeneously?

Overwhelmingly, papers assume that credit groups will form homogeneously. Even in the papers that take group composition as given, there is often an implicit assumption that groups would form homogeneously if they could, as we will explain below. The two main lines of argument behind this assumption are that peer selection and peer pressure leads to the alignment of partner objectives and behaviors.

The peer selection argument for the assumption of homogeneous matching is that the joint liability in credit groups will encourage each borrower to find the safest partners as possible, such as to be responsible for their loans the least frequently possible. In equilibrium, this will lead to the safest borrowers forming groups among themselves, the "second-safest" borrowers grouping together, and so on down the risk distribution. The equilibrium matching is therefore homogeneous (Armendáriz de Aghion, 1996). The peer pressure argument for the assumption of homogeneous matching is that borrowers will control their partners' behavior, as any extra risk-taking by their partners imposes a cost on them through the joint-liability. Borrowers will hence punish any opportunistic behavior, or any deviation from the agreed-upon behavior. The leads to homogeneous risks in credit groups in equilibrium (Stiglitz, 1990).

The question is then why groups would display any level of risk heterogeneity in equilibrium. The answer suggested by this literature is that borrowers are faced with matching frictions, which leads to second-best groups. Borrowers cannot find their optimal partners, and this causes dis-alignment of preferences in credit groups. Borrowers might differ in their willingness to repay (Besley and Coate, 1995; Diagne, 1998), or might need to monitor their partners to control moral hazard behavior (Wydick, 1995; Armendáriz de Aghion, 1996). The second-best matches would disappear if borrowers had the choice of partners of all types that they could costlessly select, monitor, and sanction.

The hypothesis we seek to test in this paper is whether groups *do* form homogeneously in risk, once we control for matching frictions, and if there is evidence that they do not, to try to provide a possible explanation as to why borrowers might *choose* to match heterogeneously in credit groups. In providing this explanation, we will refer to the literature on the endogenous choice of partners (see section 6).

3 The empirical model

A simple, naive test of the homogeneous matching hypothesis would be to generate a measure of risk heterogeneity within groups, and test whether or not it is equal to zero. However, frictions in matching, due to informational or enforcement problems, or to partner availability problems for instance, could lead to heterogeneity even if the underlying desired matching pattern is homogeneous. Therefore, we specify an empirical model that incorporates these frictions explicitly.

In a frictionless world, a borrower's choice of partners and risk would be simultaneously determined by the borrower's individual characteristics that govern his risk preferences. If borrowers are subject to matching frictions, these can pull them away from their first-best level of risk heterogeneity and from their first-best level of risk. Nonetheless, if the desired underlying matching pattern were one of homogeneous matches, we show that it would remain that no systematic relationship between borrowers desired (frictionless) risk and the level of observed risk heterogeneity in their group would exist. This leads to a natural test of the homogeneous matching hypothesis in environments with matching frictions. We precisely define our measures of risk and risk heterogeneity in section 4.

No frictions To simplify, let us start from a context in which there are no matching frictions. Denote borrower i's first-best risk choice by r_i^*, and the heterogeneity in i's group by h_i^*. The simultaneous determination of these can be written as:

$$h_i^* = H(r_i^*, 0) \tag{1}$$

$$r_i^* = R(X_i, h_i^*), \tag{2}$$

where H and R are general functions, and X is the set of exogenous borrower characteristics which determine the optimal risk. Borrower i's first best risk r_i^* depends on h^*, as his risk choice depends not only on his own characteristics, but also on the risk levels of his partners.[8] The second argument of H makes explicit the lack of matching frictions. In reduced form, the r_i^* equation becomes:

$$r_i^* = R(X_i, H(r_i^*, 0)) = g(X_i),$$

and, under homogeneous matching, h_i should be zero in all groups.

Adding matching frictions Matching frictions may prevent perfect homogeneous matching, in which case h_i will not be zero. When we refer to matching frictions, we mean any imperfect information, borrowers' inability to control moral hazard or enforce contracts, or the unavailability of borrowers' optimal partners that would cause borrowers to match with second-best partners. This affects risk heterogeneity directly by changing the difference between members from the optimum. It also affects risk heterogeneity indirectly, by influencing borrowers' choice of project risk. Borrowers having to accept a higher risk partner than desired could change their own risk by choosing a safer project than they would otherwise (to maintain a high repayment rate), or a riskier project because of the joint liability (choosing a safer project would increase the probability of having to repay the loan of the risky partner). Denoting frictions by f_i, the observed risk-heterogeneity h_i and risk level r_i under matching

[8] The full insurance-motive model would in fact be:

$$\begin{aligned} h_i^* &= h(r_i^*, t_i^*) \\ t_i^* &= T(h_i^*, 0) \\ r_i^* &= r(X_i, h_i^*, t_i^*), \end{aligned}$$

where t_i^* denotes the net transfers i pays to partners in his group, and h, T, and r are general functions. Substituting $T(h_i^*, 0)$ for t_i^* in the other expressions gives us the equations (1) and (2) in the text.

frictions are thus given by:

$$\begin{aligned} & h_i &&= H\left(r_i^*, f_i\right) \\ \text{and} & && \\ & r_i &&= R\left(X_i, h_i\right) \\ & &&= R\left(X_i, H\left(g\left(X_i\right), f_i\right)\right) \quad \equiv k\left(X_i, f_i\right). \end{aligned}$$

The complete system under matching frictions then becomes:

$$\begin{cases} \text{observed risk-heterogeneity:} & h_i = H\left(r_i^*, f_i\right) \\ \text{observed risk:} & r_i = k\left(X_i, f_i\right) \\ \text{first-best risk:} & r_i^* = k\left(X_i, 0\right). \end{cases} \tag{3}$$

Matching frictions can lead to deviations from perfect homogeneity of risks in groups. However, if frictions are the only reason for this observed heterogeneity and that borrowers are trying to match as homogeneously as possible, rejecting the null hypothesis that $h_i = 0$ would lead to an erroneous rejection of the homogeneous matching hypothesis. In appendix A.1, we show that, with a sufficiently large number of potential borrowers and a not overly skewed distribution of borrowers, frictions will add only random heterogeneity and not systematic heterogeneity. Under the hypothesis of homogeneous matching, we should see no systematic relationship between the level of first-best (frictionless) risk r_i^* and the level of risk heterogeneity h_i in a borrower's group. The test of homogeneous matching becomes:

$$\frac{\partial h_i}{\partial r_i^*} = 0. \tag{4}$$

4 The data and estimation strategy

4.1 The data

We use a 1995 survey of credit group members in Guatemala, that collected data on all members of 250 randomly selected groups that borrowed from *Génesis Empresarial*. *Génesis Empresarial* is a Guatemalan Non-Governmental Organization[9], part of the *ACCION International* network, that started lending programs for microentreprises in 1988. We use a sub-sample of "retail" groups, leaving 780 members from 210 groups.[10] The groups come from 8 locations: Guatemala City, Escuintla, an urban center in the coastal plains, the tourist

[9] *Génesis* was in phase of becoming a full-fledged bank as of 1998.
[10] We exclude producers and mixed groups because they have very different loan cycles. Doing so allows us to focus more closely on similar groups.

markets in Antigua, the small city of Chimaltenango at the base of the highlands, and three rural towns – Tecpan, Santa Lucia and La Gomera.[11]

Group members were surveyed individually, with data collected on characteristics such as the borrower's principal activity, sales, asset position, and other sources of credit. In addition, questions about the other members of the group – prior mutual knowledge, proximity, mutual help within the group – give information on screening, monitoring, and mutual-insurance behavior.

The group members are small vendors, typical of developing economies' markets, with activities ranging from sales of perishable goods to clothing items. Some borrowers are specialized (e.g. tomato sellers), while others are diversified (tiny drugstores or basic grocery stands). Business sizes are also quite variable, ranging from a blanket on a sidewalk or a palette around the neck, to stalls in covered markets. Inventories are in general very low, with many sellers holding their stocks at night in small locked storage spaces outside the market.

Table 1 provides summary statistics on the characteristics of the sample. Median average weekly sales are less than $400 a week.[12] Capital turnover is high, with 50% of borrowers buying merchandise 2 to 3 times per week, and 19% of borrowers buying merchandise daily. A very small fraction of borrowers has access to formal sector loans, and close to 30% have no other source of credit but family or friends. About half of the borrowers report having access to money lenders, with interest rates up to 25% a week, some as high as 25% daily.

Génesis offers two types of loans: individual and group loans. Borrowers receive a sequence of low interest rate loans, and maintain access to these loans if they do not default. Defaulting borrowers are denied access to future loans. Group loans are issued to partners with whom they share liability for the total group loan. *Génesis* requires credit groups be between four (which can be reduced to three after the first loan) and eight members. Most groups have three or four members.

Loans are small, short-term loans for working capital, typically starting at $50, with eight weekly (or four fortnightly) payments. Upon repayment, the loan amount grows, typically by 10-30% from one loan to the next. Group with good repayment record are given longer term

[11] The size of the sample interviewed in each city was proportional to the number of *Génesis* employees in that city as they served as enumerators.

[12] There are two wholesalers with high daily sales in excess of $4000. Their profit margins, however, stay comparatively low – around $50 to $100 per day.

loans and the possibility of monthly payments. Groups with problems repaying their loans, however, face penalties. One late payment disqualifies the group from an increase for the next loan. Two late payments reduces the loan size. Three late payments in a year is grounds for permanent exclusion. All loans have a monthly interest rate of 2.5%, which was the typical interest rate in the commercial banking sector in 1995. Annual inflation in Guatemala was an annual 10.7% in 1995 (International Finance Statistics, 1998), giving a real monthly interest rate of 1.65%.

Most loans in the sample are between $200 to $1500, representing one to three weeks worth of inventory.[13] The maximal loan size given to any borrower is about $2500, at which point the borrower is expected to graduate to the formal banking sector with "letters of recommendation" for groups that repay the last loan.

4.2 How to measure observed risk (r_i)

We construct a measure of risk based on the repayment strategy borrowers report to measure the likelihood an individual will repay. Early accumulation of money for the payment reduces the probability of default on a payment. However, because of high turnover of working capital, money set aside for repayment has a high opportunity cost. Delaying saving to repay yields a higher return, albeit with a higher risk of default.

Our measure of risk is the percentage of the last three days of sales before the repayment date that a borrower relies on to make his/her payment. Loan payments are due on Wednesdays, and borrowers were asked how much they had saved for the Wednesday payment by Sunday night. The survey also has data on daily sales. Specifically, we measure risk as:

$$r_i = \frac{P_i - S_i}{E_i}$$

where P_i is the total loan payment due on Wednesday, S_i is the amount borrower i reports having saved by Sunday night, and E_i is the sum of expected sales on a "good" Monday, Tuesday, and Wednesday. The definition of "good" was left to the borrower's discretion.[14]

[13] "Loan Size" in Table 1 includes loans to 17 groups in Escuintla whose businesses were destroyed by a fire in their market in December 1994. These groups were refinanced by Génesis with large loans (up to $5000) at a term of 18-24 months. "Loan Size if no fire" excludes these groups. In the analysis, we kept these groups, however, as the refinanced loans had the same modalities as the other loans in terms of interest payments and payment frequency.

[14] We chose "good" weeks rather than some weighted average of "good" and "bad" weeks as the weight would have been arbitrarily chosen.

The higher the fraction of expected sales E_i needed to cover the amount left to save $P_i - S_i$, the higher the risk of the borrower's liquidity strategy r_i.

Table 1 provides summary statistics for r_i. The median risk strategy relies on about 20% of the last three days of sales. Just over five percent of borrowers reported relying on over 100% of those sales (Figure 1 shows the distribution of the risk measure). We emphasize that the risk measure is the percent of total expected *revenues*. Most borrowers have only one business, and 40% are the sole provider of income to their household, so borrowers must also take out other expenses from these revenues. Further, the denominator is expected sales in "good weeks." Given that sales in bad weeks average 60% of the sales in good weeks (see table 1), relying on 50% of expected sales is not as safe as one might presume.

4.3 How to measure risk heterogeneity (h_i)

The other measure central to our analysis is the measure of risk heterogeneity in a group. We use two different measures, the second one to mitigate possible artificial heterogeneity from measurement error in r_i.

The first measure is a direct generalization of the measure used in theoretical models. In most theoretical work, credit groups have two borrowers and risk heterogeneity is simply the difference between the risk of the two members, $r_i - r_j$. Because we have groups of different sizes, we use the average euclidean distance between a borrower's risk and the risk of his partners. Like in the groups of two, heterogeneity is negative if the borrower is below the mean risk of the group, and positive if the borrower is above:[15]

$$h_i = \left[\sum_{r_j \in G_i} \frac{(r_i - r_j)^2}{(N_i - 1)} \right]^{1/2} \cdot \operatorname{sign}(r_i - \overline{r}_i) \tag{5}$$

where r_i is defined above, $r_j \in G_i$ are i's partners, $\overline{r_i}$ is the average risk in i's group G_i, and N_i is the number of borrowers in the group.[16] A distance of -0.2 means that borrower i relies on 20% less of his expected sales than his partners do.

Table 1 provides summary statistics for h_i, and the estimated (kernel) density of h_i is plotted in Figure 2. A large proportion of the distribution is concentrated between -1 and

[15] Signing with respect to the mean rather than the median group risk is more robust to errors such as the third borrower in the group $\{0, 0, 0, 1\}$ getting a positive wrong sign.

[16] The data only contain one group in which we have $r_i = \overline{r}_i$. In that case, the group is perfectly homogeneous in risk, so that the sign of $r_i - \overline{r}_i$ is not an issue.

1. However, there are a few outliers, including values up to 3.23. These high values of h_i are all in groups composed of very high risks ($r_i > 1.5$), which suggests possible measurement problems. For example, it is not clear that a reported risk of four is very different from a value of five. However a risk of zero is very different from a risk of one. The distance metric in equation (5) treats these differences as the same.

The second measure of risk heterogeneity we use attempts to reduce the effect of imprecision in r_i by calculating the difference in risk in a group based on risk deciles:[17]

$$h_i^d = \left[\sum_{d_j \in G_i} \frac{(d_i - d_j)^2}{(N_i - 1)} \right]^{1/2} \cdot \operatorname{sign}(r_i - \overline{r}_i).$$

The variables d_i and d_j are the risk deciles to which borrowers r_i and r_j belong, respectively. As for h_i, h_i^d is signed by a borrower's position with respect to the average risk in the group. This heterogeneity measure is less sensitive to measurement error in r_i. However, as the highest risk decile is extremely wide (.72 to 5.23), it might over-smooth differences for high risk borrowers.

Figure 3 shows the estimated density of h_i^d. Twenty percent of borrowers are still over four deciles away from their partners, and 13 groups have one borrower in the lowest decile and another in the top decile. These groups seem, in our judgement, realistically heterogeneous, rather than artificially so due to measurement error.

Both of these heterogeneity measures reject the naive assertion that groups match homogeneously, as some groups display substantial levels of heterogeneity. This may solely be due to matching frictions, so we turn to an empirical test of a *systematic* link between risk and risk heterogeneity.

4.4 Estimation strategy

For estimation, we assume a linear approximation to system (3):

$$\begin{cases} h_i = \eta + \gamma r_i^* + \delta f_i + \varepsilon_i^h \\ r_i = X_i \alpha + \beta f_i + \varepsilon_i^r. \end{cases} \tag{6}$$

Because we do not observe borrowers' matching frictions, we let f_i be a function of observable determinants W_i, and re-write the first equation as:

$$r_i = X_i \alpha + (W_i \phi)\, \beta + \varepsilon_i^r. \tag{7}$$

[17] We also did the analysis using a distance measure that weighted down distances at high risk levels. The results were between the results we will obtain with the risk-level and risk-decile based distance measures.

An error ε_i^r is added to equation (7) to account for variation in risk levels not captured by the model.[18] Estimation of (7) gives fitted values

$$\begin{aligned} \widehat{r_i^*} &= X_i \widehat{\alpha} \\ \widehat{\beta f}_i &= W_i \widehat{\phi\beta}, \end{aligned}$$

which we then substitute into equation (6) to estimate γ:

$$h_i = \eta' + \gamma \widehat{r_i^*} + \delta' \widehat{\beta f}_i + \varepsilon_i^h. \tag{8}$$

The coefficients η' and δ' are transformations of the original coefficients η and δ, as neither coefficient is identifiable.[19] The error ε_i^h is added to equation (8) to capture unexplained matching variations and to absorb the effect of measurement error in h_i.[20] Testing the hypothesis of homogeneous group formation becomes testing the null hypothesis $\gamma = 0$.

As the estimated values $\widehat{r_i^*}$ and $\widehat{\beta f_i}$ are measured with error, we estimate equation (8) by instrumental variables (IV). The estimation of the equations (7) and (8) presents various econometric difficulties, which we turn to next.

4.4.1 Identification and omitted X or W

Inspection of equation (7) reveals that identification of the parameters relies on correct specification of the variables which determine risk and frictions. In our first specification below, we err on the side of parsimony in choosing the variables which determine an individual's optimal risk. Omitted variables bias is a commonly known phenomenon. However, as we do not directly care about the parameters in equation (7) and are only interested in the fitted values $\widehat{r_i^*}$ and $\widehat{\beta f_i}$, appendix A.2 demonstrates that we obtain a consistent estimates, even if we had omitted a variable from either X or W (but not from both).

4.4.2 Identification and shared X and W variables

If there were a variable which determined both risk and matching frictions, we would be unable to identify the separate effects, and our methodology would not yield us a consistent

[18] For simplicity of exposition, we assume that $W_i\phi$ is exactly f_i, not f_i plus some error ε_i^f. In fact, adding an error term to f_i would not bias the estimates, as the error term would remain uncorrelated with W_i (and X_i). The estimates of α and β are hence unaffected by its inclusion or not.

[19] The constant is not identified as it contains the effect of the constant terms in both X and W. The parameter δ is not identifiable as we only get an estimate of δ/β and have no other information to identify δ and β separately.

[20] Note that the variance-covariance structure of ε^r and ε^h are not independent as the measurement error in r_i goes directly into h_i.

estimate of γ. Appendix A.3 gives a 2-step procedure which allows us to recover γ consistently in this case. Essentially, we rely on the Frisch-Waugh theorem, and partial out the effects of the shared terms from all of the other variables. We can thus re-capture the coefficients we need. We use this two-step procedure in section 5.2 to check the robustness to specification of our results.

5 Testing homogeneous matching

5.1 Estimation of $\widehat{r}_i^*$ and $\widehat{\beta f}_i$

The results from the estimation of equation (7), namely

$$r_i = X_i \alpha + (W_i \phi)\, \beta + \varepsilon_i^r,$$

are reported in Table 2. We choose set of variables X_i, related to borrowers' risk preferences, to predict first-best risk $\widehat{r}_i^*$. We choose the set of variables W_i, related to borrowers' geographical location, their integration into the market, and a measure of their outside options, to measure the impact of frictions. The observed risk r_i is technically censored at zero, so we report both the OLS estimates (in column 1), and the Censored Least Absolute Deviations (CLAD) estimates (in column 2).[21] Robust standard errors are calculated to account for heteroskedasticity and correlation within credit groups. The F-statistics are based on the robust variance-covariance matrix for the OLS estimates, and on bootstrapped estimates corrected for clustering for the CLAD estimates.

The factors X we use to predict borrower i's first-best risk are related to his risk preferences only, and not those related to the strength of matching frictions. Secondary school completion, whether from a traditional or a professional track, is positively correlated with lower first-best risk choice in our sample.[22] Similarly, borrowers with restricted sales activities during the weekends, often because of their participation in religious activities, appear to have lower risk preferences than others. Wealthy borrowers, as captured by house-owning status, are significantly less averse to risk. Although not significant, we include the effect of gender because many of the *Génesis* staff believe that women are more averse to risk than

[21] Heteroskedasticity, typically present in survey data, makes maximum likelihood Tobit estimates biased and inconsistent.

[22] We maintain the disaggregated categories, though the effects are not very significantly different from each other, in an effort to gain more information in the predictions of $\widehat{r}_i^*$. The reference level of education is secondary school completion.

men. Larger weekend sales indicate less risky businesses in terms of loan repayment. Categorical variables classifying the businesses by types were excluded as they were insignificant for all specifications.

The results for the friction determinants W are difficult to interpret, as they predict βf_i rather than simply f_i since β and ϕ are not separately identifiable. Nonetheless, the size of the city seems to affect the severity of frictions (cities are ordered according to size in Table 2, with the largest city, Guatemala City, as the control). The matching frictions are significantly higher in the rural areas of Tecpan, La Gomera, and Santa Lucia. Tenure in the market and access to alternative credit sources are also significantly correlated with frictions. Geography and market tenure potentially affect matching frictions through partner availability. Alternative emergency credit sources, such as friends, family, or money-lenders, might affect borrowers' response to matching frictions through their role as a source of insurance. In addition, the monthly interest rates charged by moneylenders[23] includes information the money-lender might have on borrower risk. Similarly, these interest rates reflect borrowers' outside option; fewer options imply matching frictions can have stronger effects.

Judging by the F-statistics and the R-squared, the fits are admittedly not very strong, particularly for the determinants first-best risk r_i^*. This is partly due to the noisy data, but also due to our parsimonious specification of first-best risk. Nonetheless, as discussed below, a poor fit would cause a tendency to fail to reject the null hypothesis of $\gamma = 0$.

The same equation was re-estimated using CLAD, to account for the censoring in r_i and for the potential heteroskedasticity. The estimated effects are smaller in general, and the effect of house ownership disappears. Matching frictions generally remain higher in the rural areas. The effect of access to alternative credit remains significant (with a p-value for joint-significance of less than 1%), but smaller in magnitude. This estimation allows for a sensitivity test of our results to the estimation method.

5.2 The test: estimating γ

Estimation of equation (8), namely

$$h_i = \eta + \gamma r_i^* + \widetilde{\delta}\left(\beta f_i\right) + \varepsilon_i^h,$$

[23] We include a dummy variable for borrowers reporting no access to moneylenders, and a dummy for borrowers reporting daily interest rates rather than monthly rates.

yields our test of homogeneous matching. Unfortunately, $\widehat{r_i^*}$ and $\widehat{\beta f_i}$ are estimated with error. If the measurement error in $\widehat{r^*}$ and $\widehat{\beta f}$ is correlated, the bias in the un-instrumented OLS estimate is not necessarily the usual attenuation bias (see Deaton, 1997, p.100). In our case, without further assumptions, the bias cannot be signed.

We therefore estimate the equation (8) with instrumental variables. We need instruments that are not direct determinants of r_i^* and f_i, i.e. not included in X and W, but correlated with r_i^* and f_i. We use as instruments the observed risk levels of *similar* borrowers. *Similarity* is measured by migrant status, educational level, years of experience, discount rate, whether the individual works on Sundays, the number of family members in the group, and the number of times a borrower needed help to make a payment in the past year. The average observed risk levels were calculated excluding the borrower and members of his or her group, so as not to introduce a correlation with the measurement error. These average risks are highly correlated with borrowers' first-best risk – similar borrowers have similar preferences –, and with the effect of matching frictions – similar borrowers are faced with similar matching frictions. As additional instruments, we use the monthly interest rates borrowers report from moneylenders[24]. These outside options help determine how strong matching frictions are.

The first column of Table 3 reports the first-stage regressions for the estimates of $\widehat{r_i^*}$ and of $\widehat{\beta f_i}$ (top and bottom respectively) from the OLS estimation of equation (7). The second column reports the first-stage results using the CLAD estimates of $\widehat{r_i^*}$ and $\widehat{\beta f_i}$. The F-statistics on the instruments in the risk-measure equations are 22.50 and 19.52 for $\widehat{r_i^*}$ predicted using the least-squares and CLAD estimates, respectively. The instrumentation is weaker on the friction equation, with F-statistics of 5.07 and 6.37. We kept the same instrument set for $\widehat{r_i^*}$ and $\widehat{\beta f_i}$ such as not to impose arbitrary restrictions; estimating with equation-specific instrument sets improves the fits, but doesn't otherwise change the qualitative nature of the results. The over-identification tests, reported in Table 4, do not warrant the rejection of the orthogonality of the instruments.

The first column of Table 4 reports the second-stage coefficients for the main regression (8). The top half of the table shows the coefficients from the un-instrumented regression. The bottom half comes from IV estimation. The second column reports the results using the CLAD estimates. While the coefficient on the frictions cannot be separately identified,

[24]We include a dummy variable for borrowers reporting no access to moneylenders, and a dummy for borrowers reporting daily interest rates rather than monthly rates.

the coefficient of interest, γ, is estimated to be positive and significantly different from zero, rejecting the homogeneous matching hypothesis.

To check the robustness of our result, we re-estimate the risk-difference equation using decile-based heterogeneity measure h_i^d. Table 5 shows the results are similar, with the estimates of γ still suggesting rejection of the homogeneous matching hypothesis.

As a specification test, we run the two-step procedure outlined in section 4.4.2, regrouping any variable from X and W which might include shared effects into a set V. The results are reported in Table 6. (xx)

Using either measure of heterogeneity (h_i or h_i^d), either set of predicted values for first-best risk and for frictions (by OLS or CLAD), adjusting for frictions, and running a specification test, we uniformly find a positive relationship between a borrower's preferred risk level and the heterogeneity in the group. Safe borrowers tend to match with riskier partners.

We note that the positive γ is not due to a poorly measured r_i^*, with measurement error correlated with the measurement error in h_i. By construction, $\widehat{r}_i^*$ is orthogonal to the measurement error in r_i, and hence orthogonal to the measurement error in h_i. So, even if r_i^* were so poorly measured that $\widehat{r}_i^*$ is in fact random noise, the estimated γ would be equal to zero. We can be confident that we estimate a positive γ, rejecting the hypothesis of homogeneous matching.

6 Insurance in Credit Groups

Our results refute the common assumption of homogeneous risk-matching, even controlling for matching frictions. We now explore one explanation why borrowers might want to form heterogeneous groups. Homogeneous matching is usually based on the following argument. If borrowers are jointly liable, each will want the safest partner possible (assuming away covariance). The safest borrowers will match together, the next safest will match together, and so on. In equilibrium, groups will be homogeneous in risk. However, the joint liability aspect may allow borrowers to use the credit groups to set up insurance arrangements where such arrangements are otherwise not possible. Joint liability makes promises of insurance credible, as reneging borrowers would be considered in default (Sadoulet, 1997). Risky borrowers can then gain from matching with safe borrowers, can in turn receive a transfer payment to compensate (and possibly remunerate) them. Consider a borrower i who never fails. Matching with an identical partner brings no benefit over individual loan. However, if he were to match with a partner j who is riskier, then j would benefit from the extra insurance over matching homogeneously, and i could extract some transfer from j as a payment from that insurance. Sadoulet (1998) shows that this leads to some heterogeneous matches in equilibrium. Safer borrowers essentially sell some of their low risk to their riskier partners.

Like the argument for homogeneous matching, this insurance motivation is independent of covariance risk. The heterogeneous matching does, however, depend on borrowers being able to extract transfers in heterogeneous groups. We will return to this point later.[25]

To measure the insurance in groups, we rely on the number of times a borrower received help to make a payment. Table 7 reports insurance statistics. Almost two-thirds of the groups report some insurance in the past year, half reporting more than four.[26] The amounts are relatively small, \$20 to \$30, which represents about 20 to 25% of borrower payments, though in some cases amounts range up to two-thirds of payments. Most of the insurance comes from within the group, though 20% of borrowers receive assistance from outside. Borrowers report that the insurance is mainly to offset low sales or poor planning, directly related to our risk measure.

Examining intragroup insurance flows in Table 8, we note that there are many borrowers who provide insurance more often than they receive it.[27] The insurance is provided by the borrowers who are relatively safe (as measured by $\widehat{r_i^*}$) to their higher risk partners. However, if insurance is part of the group formation process, the net providers of insurance are the safer borrowers in the population (and not only within their group), and receivers are the riskier ones. The bottom table of Table 8 shows net insurance provision by quantiles of $\widehat{r^*}$. The χ^2 statistic implies that insurance providers come disproportionately from the safe quantiles, the recipients from the highest risk quantiles.

In heterogeneous groups, there need to be risk premium transfer payments to account for

[25] Besides Sadoulet (1998), only Ghatak (1996), and Armendáriz de Aghion and Gollier (1999) also endogenize group formation. These two other papers find that groups match homogeneously in equilibrium. However the models are static so that the gain from insurance is solely lower equilibrium interest rates, which is not sufficient to sustain heterogeneous groups. In Sadoulet (1998), the repeated game gives borrowers the additional cost of loss to future loans in case of default. The gain from insurance is therefore much greater and can lead borrowers to choose heterogeneous groups.

[26] The differences in number of groups reporting not receiving any insurance (38%) and groups reporting not giving any insurance (39%) is due to turnover in group membership.

[27] We only have information on how many times a borrower provided and received insurance in the past year, and not the amounts provided.

the asymmetric insurance. Unfortunately, we have no data on these payments. However, an anecdote from conducting the survey illustrates one way in which transfers can be paid in exchange for insurance. One group interviewed in the survey was composed of four borrowers. The group leader was a 50 year-old man with a well-established cloth business. He had been selling in the market for 26 years and was well-respected. The other 3 members of the group were sold shoes, were around 25 years old, and in their second year of business. It became clear by talking to each member of the group, that the younger members were repaying part or all of the group leader's loan. The leader only repaid when the younger members could not repay the group loan.

This apparent free-riding can be seen as transfers as payment for insurance to the group. The young shoe sellers pay part or all of the group leader's loan, in exchange for insurance when their projects fail. They do not evict him for not paying his share because they gain from the insurance that he provides them when their project fails. The cloth seller, on his part, gets the loan for free.

So far, our results have shown that groups display more risk heterogeneity than matching frictions would imply, and that there is substantial intragroup insurance. There are, however, some very safe, homogeneous groups in the sample, which seem to conflict with the insurance model. Nonetheless, if borrowers are unable to extract transfers, assess their partners' risk, or monitor their behavior, transfers in heterogeneous groups are impossible. Failing this, borrowers matching with riskier partners would be worse off than if they matched homogeneously.

Table 9 compares the borrowers in safe, homogeneous groups to the rest of borrowers. Surprisingly, these groups display slightly more insurance activity than average, despite the fact that they are of much lower risk.[28] These groups have fewer late payments, with one in six groups reporting late payments compared to 50% in the full sample. Screening and monitoring seem to be more important in these homogenous groups. Partners are physically closer together, within sight in most cases. Borrowers all knew their partners before they joined the group, and tend to believe that visits to partners are important for the group. The other striking difference between these safe homogeneous borrowers and the rest is the difference in access to outside credit. A smaller proportion of these borrowers report having

[28] Calculating average insurance per member still reveals more insurance activity.

access to credit from friends, and a much smaller proportion report being able to borrower from friends and money-lenders.

This evidence is consistent with greater costs of losing access to credit from *Génesis*. These borrowers may hence require larger transfers, all else being equal, in heterogeneous groups. This could push them towards homogeneous matching in equilibrium. Such clients, however, could be made strictly better off with mechanisms for self-insurance, which would then allow them to match heterogeneously and extract some surplus from their low risk.

7 Conclusion

We have tested whether credit groups form homogeneously, as is widely assumed in the literature. Using survey data from Guatemala, we consistently find evidence that, even accounting for matching frictions, some borrowers *choose* to form heterogeneous groups.

The repercussions from these results are twofold. Methodologically, they underline the importance of endogeneity of group formation for analysis of credit group performance. When borrowers choose partners,[29] they may do so to respond to differing contractual needs. Using group characteristics as explanatory variables without accounting for this endogeneity can lead to erroneous conclusions.

On the policy side, the analysis in the last section suggests that group lending may be important to borrowers as a source of insurance. There is room to design better insurance mechanisms within these financial contracts.

Lastly, we underline the possible importance of contractual failures among partners of credit groups. Group lending does not necessarily empower the disenfranchised members of societies, as is often heralded in the popular press. There is even strong evidence that microfinance contracts do not reach the "poorest of the poor." Individual loans with self-insurance mechanisms may yield better results than second-best groups joined because of contractual failures. Microcredit organizations may better serve borrowers by adapting their financial products to the needs of the clientele.

[29] See Diagne (1998) for an interesting case in which peer selection is limited by social hierarchy.

References

[1] Ackerberg, D.A. and M. Botticini (1998). "Endogenous Matching and the Empirical Determinants of Contract Form," mimeo, Boston University.

[2] Aleem, Irfan. (1990). "Imperfect Information, Screening, and the Costs of Informal Lending: A Study of a Rural Credit Market in Pakistan", *World Bank Economic Review*, 4, pp. 351-366.

[3] Armendáriz de Aghion, B. (1995) "On the Design of a Credit Agreement with Peer Monitoring", typescript, University College London, February.

[4] Armendáriz de Aghion, B. and C. Gollier (1999). "Peer Group Formation in an Adverse Selection Model", typescript, University College London, January.

[5] Associated Press (1998). "Illegal Small Loan Activity Rising," The New York Times, November 11.

[6] Becker, G. (1973). "A Theory of Marriage: Part I," *Journal of Political Economy*, pp. 813-846.

[7] Besley, T. and S. Coate (1995). "Group Lending, Repayment Incentives and Social Collateral", *Journal of Development Economics*, Vol. 46, pp.1-18.

[8] Braverman, A. and L. Guasch. (1993)."Administrative Failures in Government Credit Programs," in *The Economics of Rural Organization: Theory, Practice and Policy,* edited by K. Hoff, A. Braverman, and J. Stiglitz. Published for the World Bank by Oxford University Press.

[9] Che, Y-K. (1999)."Joint Liability and Peer Sanction under Group Lending," typescript, University of Wisconsin, March 1.

[10] Conning, J. (1997). "Joint Liability, Peer Monitoring, and the Creation of Social Collateral," typescript, November 19. Available from http://wso.williams.edu/~jconning

[11] Davidson, R. and J.G. MacKinnon (1993). *Estimation and Inference in Econometrics*, Oxford University Press, New York and Oxford.

[12] Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, published for the World Bank, The Johns Hopkins University Press, Baltimore and London.

[13] Diagne, A. (1997). "Default Incentives, Peer Pressure, and Equilibrium Outcomes in Group-based Lending Programs", mimeo, IFPRI, November 1997.

[14] Gourieroux, C. (1989). *Économétrie des Variables Qualitatives*, Édition Économica.

[15] Ghatak, M. (1995). "Group Lending and the Peer Selection Effect", typescript, University of Chicago.

[16] Huppi, M. and G. Feder (1990). "The Role of Groups and Credit Cooperatives in Rural Lending," *World Bank Research Observer* 5 (2), pp. 187-204.

[17] International Finance Statistics (1998).

[18] Kremer, M. and Maskin (1995). "Segregation by Skill and the Rise in Inequality," NBER Working Paper 5718.

[19] Madajewicz, M. (1997). "Capital for the Poor: the Role of Monitoring," typescript, Columbia University, November.

[20] Madajewicz, M. (1998)."Capital for the Poor: the Effect of Wealth on the Optimal Loan Contract and the Implications for Growth," typescript, Columbia University, October.

[21] Morduch, J. (1997). "The Microfinance Revolution", typescript, Princeton University, April 5.

[22] Rhyne, E. and M. Otero eds. (1994). *The New World of Microentreprise Finance: Building Healthy Financial Institutions for the Poor*, West Hartford, CT: Kumarian.

[23] Rhyne, E. and L. Rotblatt (1994). *What Makes Them Tick? Exploring the Anatomy of Major Microentreprise Finance Organizations*, Cambridge, Ma: ACCION International.

[24] Sadoulet, L. (1997). "The Role of Mutual Insurance in Group Lending", typescript, December. Available from http://homepages.ulb.ac.be/~sadoulet

[25] Sadoulet, L. (1998). "Non-monotone Matching in Group Lending", typescript, revised April 99. Available from http://homepages.ulb.ac.be/~sadoulet

[26] Sharma, M. and M. Zeller (1997). "Repayment Performance in Group-based Credit Programs in Bangladesh: An Empirical Analysis", *World Development*, October 1997.

[27] Shimer and Smith (1998) "Assortative Matching and Search", forthcoming in *Econometrica.*

[28] Stiglitz, J. (1990). "Peer Monitoring and Credit Markets", *World Bank Economic Review*, 4, pp. 351-366.

[29] Udry, C. (1990). "Credit Markets in Northern Nigeria: Credit as Insurance in a Rural Economy," *The World Bank Economic Review,* Volume 4, Number 3, September 1990.

[30] Van Tassel, E. (1998). "Group Lending under Incomplete Information," typescript, University of Queensland, Australia.

[31] Varian, H. (1990). "Monitoring Agents with Other Agents", *Zeitschrift für die Gesamte Staatswissenschaft,* 146, pp. 153-174.

[32] Wenner, M. (1995). "Group Credit: A Means to Improve Information Transfer and Loan Repayment Performance", *Journal of Development Studies,* Vol.32, No.2, December 1995, pp.263-281.

[33] Wydick, W.B. (1995). "Group Lending as a Credit Delivery Mechanism in Guatemala", mimeo, University of San Francisco, October 1995.

[34] Wydick, W.B. (1999). "Can Social Cohesion be Harnessed to Repair Market Failures," *The Economic Journal, 109 (July), 463-475.*

[35] Zeller, M. (1998). "Determinants of Repayment Performance in Credit Groups: The Role of Program Design, Intra-group Risk Pooling, and Social Cohesion in Madagascar," *Journal of Economic Development and Cultural Change,* January.

A Appendix

A.1 Matching Frictions

As described in section 3, matching frictions, due to contractual problems or to limited partner availability, will engender risk heterogeneity in credit groups, even if borrowers want to match as homogeneously as possible. In this appendix, we will show that, if borrowers want to match homogeneously and are faced with matching frictions, there is still no systematic relationship between borrowers' first-best risk r_i^* and the level of observed risk heterogeneity in their group h_i.

We must distinguish two types of matching frictions: "contractual frictions", by which we mean issues of imperfect information, enforcement, etc.; and search frictions. Search frictions arise when there are costs in finding optimal partners. These costs can lead borrowers to match with second-best partners, and if costs are sufficiently large, random matching ensues. In random matching models, this usually leads to mean-reversion in the choice of partners (see Shimer and Smith, 1999): it is more likely for a safe borrower to encounter a partner riskier than himself, as there are more borrowers riskier than borrowers safer than he. In our context, however, because borrowers have access to individual loans, the set of second-best groups can be found to be a symmetric set around each borrower type (Sadoulet, 1997). The mean-reversion in the choice of partners does not hold, therefore, unless the distribution of types is very skewed on this symmetric matching set. As the survey was conducted mainly in large urban centers and that credit needs of microentrepreneurs in the markets surveyed were vastly underserved, there are large numbers of available partners of each type. The distribution of types on the matching set is essentially uniform. It is thus unlikely that search frictions would lead to a mean reversion. "Contractual frictions" also lead borrowers to pick second-best partners. However, unless there is a negative correlation between the effect of these frictions and borrowers' risk preferences (i.e. that borrowers have systematically better contractual relationships with partners very different from them), these types of matching frictions would not lead to a positive relationship between individual risk and the level of observed risk heterogeneity in their groups.

In our context, therefore, matching frictions do not lead to a systematic deviation between borrowers' risk preferences and the risk preferences of their partners. While second-best partners can enduce an adjustment of borrowers' risk strategy, this implies, in particular, that matching frictions do not enduce a systematic relationship between borrowers' risk preferences and the level of observed risk heterogeneity in their group.

A.2 Misspecification of X and W

This appendix will demonstrate that in the event of omitted variables from either the optimal risk specification or the friction specification (but not both), although the parameter estimates are biased, the predicted value will be consistent.

Let the true model be (combining X and W in equation (7) into a single variable X):

$$r = X\alpha + Z\zeta + \varepsilon$$

If we omit Z from the regression, the standard omitted variables bias obtains for the parameter estimate:

$$\widehat{\alpha} = \alpha + \left(X'X\right)^{-1} X'Z\zeta + \left(X'X\right)^{-1} X'\varepsilon$$

However, if we use this (biased) coefficient estimate to predict y, we get:

$$E\left(X\widehat{\alpha}\right) = X\alpha + X\left(X'X\right)^{-1} X'Z\zeta.$$

The expected forecast error is hence

$$E\left[r - X\widehat{\alpha}\right] = X\alpha + Z\zeta - X\alpha - X\left(X'X\right)^{-1} X'Z\zeta.$$

Defining $\omega = Z\zeta$, we get:

$$\begin{aligned} E\left[r - X\widehat{\alpha}\right] &= \omega - X\left(X'X\right)^{-1} X'\omega \\ &= \left(I - X\left(X'X\right)^{-1} X'\right)\omega. \end{aligned} \tag{9}$$

Equation (9) defines the residuals from a least squares regression of ω on X. Assuming that X contains a constant term, the first order conditions imply that this will sum to zero for the sample. Therefore the average value of the difference between the predicted value and the actual value (the sum divided by the sample size) will be zero, and the predicted value is a consistent estimator, even in the presence of omitted variables.

A.3 Consistent γ with shared risk/friction variables

This appendix will demonstrate a means of consistently estimating γ, the parameter of interest, even if there is a variable which belongs in both the risk and friction specification. The procedure follows directly from the Frisch-Waugh theorem. Take equation (7) from section 4.4. Suppose a variable V belonged in both X and W:

$$r_i = (X_i\alpha + V_i\sigma_1) + ((W_i\beta)\,\phi + V_i\sigma_2) + \varepsilon_i.$$

If we were to estimate this regression, we could not separately identify σ_1 and σ_2. However, if we apply the Frisch-Waugh theorem, we can regress r_i on V_i, regress X_i on V_i, and regress W_i on V_i. Denoting the residuals of these regressions by $r_i^{\perp}$, $X_i^{\perp}$, and $W_i^{\perp}$ respectively, we can then regress them in the following manner:

$$r_i^{\perp} = X_i^{\perp}\alpha + W_i^{\perp}\beta\phi + \epsilon_i$$

This will recover consistent estimates of $\widehat{\alpha}$ and $\widehat{\phi\beta}$, by the Frisch-Waugh theorem.

In a second stage, equation (8) can be re-written as:

$$\begin{aligned} h_i &= \eta' + \gamma\,(X_i\alpha + V_i\sigma_1) + \delta\,((W_i\beta)\,\phi + V_i\sigma_2) + \varepsilon \\ &= \eta' + \gamma X_i\alpha + \delta\,(W_i\beta)\,\phi + (\gamma\sigma_1 + \delta\phi\beta\sigma_2)\,V_i + \varepsilon \end{aligned}$$

From the first stage, we have consistent estimates of $\widehat{X_i\alpha}$ and $\widehat{(W_i\beta)}\,\phi$. We can therefore regress h_i on these fitted values and V_i and from the coefficient on $\widehat{X_i\alpha}$ we can recover γ, the parameter of interest.

A.4 Overidentification Tests

The overidentification tests we use to test the validity of the instruments proceed as follows:

1. We estimated equation (7) and predicted $\widehat{r^*}$ and $\widehat{\beta f}$, which are noisy estimates of r^* and βf, respectively. Equation (7) was estimated by two alternative methods: OLS and CLAD. The results are reported in Table 2.

2. We regressed the predicted $\widehat{r^*}$ and $\widehat{\beta f}$ on a set of instruments Z. The results of these first stages are reported in Table 3 (for both sets of predictions). We predicted $\widehat{\widehat{r^*}}$ and $\widehat{\widehat{\beta f}}$.

3. The second stage consists in the regression of h on $\widehat{\widehat{r^*}}$ and $\widehat{\widehat{\beta f}}$, which give the IV estimates for the parameters. The standard errors of the parameters were calculated using standard IV formulas, allowing for clustering effects within groups. The results corresponding to the OLS and CLAD predictions are reported in Table 4 for the heterogeneity measure h_i based on risk levels, and in Table 5 for the heterogeneity measure h_i^d based on risk deciles.

4. We then compute the IV residuals $h - \gamma_{IV} \cdot \widehat{r^*} - \delta_{IV} \cdot \widehat{\beta f} - \eta_{IV}$, which we regress on the matrix of instruments Z. The residual regressions are reported in Table A1.

5. The test statistic, TR^2, is the number of observations (759 and 739 for the OLS and CLAD estimations) multiplied by the non-centered R^2 of the residual regression. This statistic is distributed as a χ^2 with $\dim(Z) - 2$ degrees of freedom, which is the number of overidentifying restrictions under the null hypothesis that the instruments are valid.

6. In the regressions using the risk-level based measure of heterogeneity h_i, the p-values for the OID test statistics (reported in Table 4) are 0.91 and 0.42 for the OLS and CLAD approaches to estimating r^* and f, respectively. In the regressions using the risk-decile based measure of heterogeneity h_i^d, the p-values are 0.52 and 0.48 for the OLS and CLAD approaches to estimating r^* and f, respectively (reported in Table 5).

7. Based on these p-values, we did not reject the null hypothesis of the validity of the instruments.

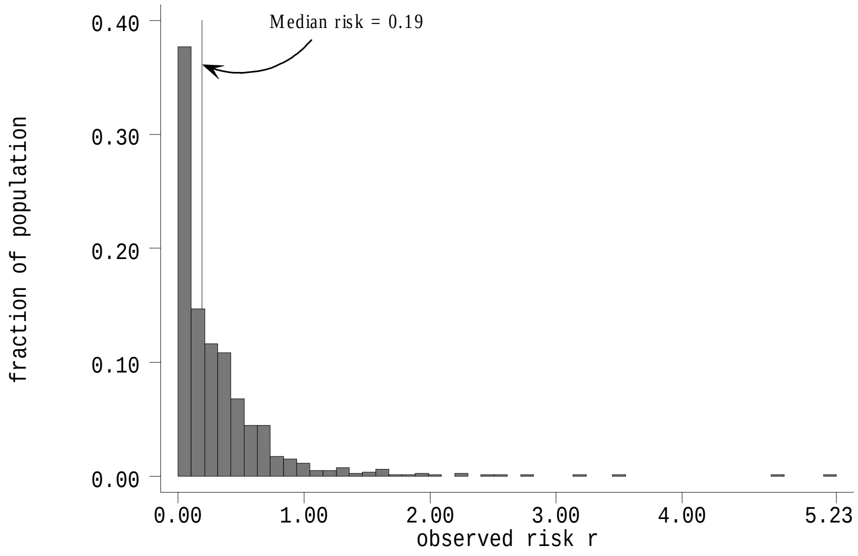

Figure 1: **Distribution of Observed Risk** r_i.

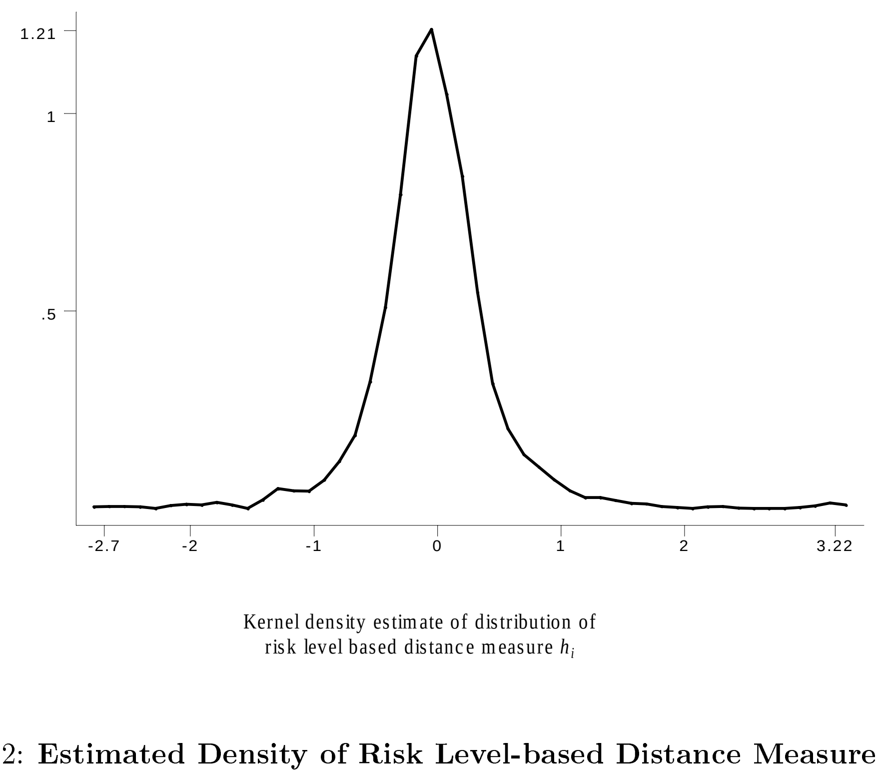

Figure 2: **Estimated Density of Risk Level-based Distance Measure** h_i.

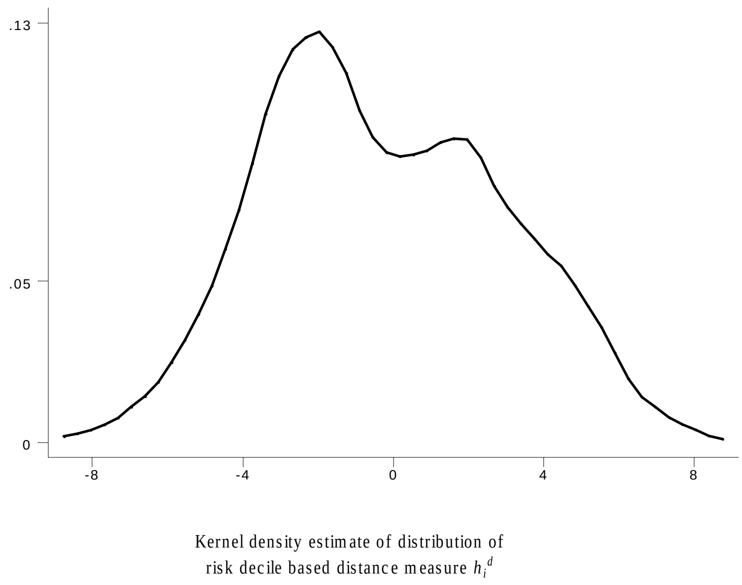

Figure 3: **Estimated Density of Decile-based Distance Measure** h_i^d

Table 1. Descriptive Statistics (continued next page)

Variables	Number observ.	Mean	Std. Dev.*	Min*	Max*	Median*	5%*	95%*
Personal characteristics								
Sex (1=male)	782	0.56	---	---	---	---	---	---
Age	764	37.7	11	18	75	36	23	58
Marital status (0=married)	756	0.22	---	---	---	---	---	---
Years of education	762	3.5	3.20	0	12	3	0	9
No education	762	0.28	---	---	---	---	---	---
Own house	740	0.39	---	---	---	---	---	---
Rural	782	0.18	---	---	---	---	---	---
Years in city	782	22.1	13.25	0	70	20	5	47
Years in market	781	12.9	8.42	0	62	11	3	28
Not born in city	782	0.85	---	---	---	---	---	---
Arrived less than 10 years ago	782	0.27	---	---	---	---	---	---
Sole income provider to household	782	0.40	---	---	---	---	---	---
Credit access								
Access to money-lender	780	0.47	---	---	---	---	---	---
Access to credit from friends/family	780	0.51	---	---	---	---	---	---
Access to formal banks	780	0.04	---	---	---	---	---	---
No other credit source	780	0.10	---	---	---	---	---	---
Only other source is friends/family	780	0.19	---	---	---	---	---	---
Business characteristics								
Non-perishable products	777	0.30	---	---	---	---	---	---
Perishable products	777	0.23	---	---	---	---	---	---
Clothes	777	0.38	---	---	---	---	---	---
Weavers and other non-market sales	777	0.09	---	---	---	---	---	---
Only one business	772	0.79	---	---	---	---	---	---
Average weekly sales ($US)	782	531	770	14	13333	381	112	1203
Bad week sales, as a fraction of good week	782	0.58	0.17	0.03	1.00	0.58	0.29	0.84
Buying merchandise daily	782	0.19	---	---	---	---	---	---
Buying 2-3 times per week	782	0.29	---	---	---	---	---	---
Buying once a week	782	0.35	---	---	---	---	---	---
Years of experience at beginning of group	720	11.5	8.39	1	55	10	2	30
Age of business at beginning of group	725	8.4	6.95	1	50	6	2	22

* Standard deviation, minimum, maximum, median, and 5% and 95% points are not reported for dummy variables.

Table 1. Descriptive Statistics (cont.)

Variables	Number observ.	Mean	Std. Dev.*	Min*	Max*	Median*	5%*	95%*
Group characteristics								
Group size	210	3.7	0.94	3	8	3	3	6
groups of 3	105	0.50 †						
groups of 4	72	0.34 †						
groups of 5	26	0.12 †						
groups of 6	3	0.01 †						
groups of 7	1	0.00 †						
groups of 8	3	0.01 †						
Single sex group	210	0.51	---	---	---	---	---	---
Group composed only of men	210	0.31	---	---	---	---	---	---
Group composed only of women	210	0.20	---	---	---	---	---	---
Loan characteristics								
Loan size (in US$)	782	740	555	56	5000	650	167	1700
Loan size if no fire (in US$)	712	644	370	67	2000	567	167	1333
Loan size/average daily purchases	647	17	20	1	93	12	2	47
Loan with weekly payments	782	0.30	---	---	---	---	---	---
Loan with fortnightly payments	782	0.32	---	---	---	---	---	---
Loans with monthly payments	782	0.38	---	---	---	---	---	---
Payment size (in US$)	782	124	97	5	718	93	29	276
Payment size if no fire (in US$)	712	124	98	7	718	92	30	276
Payment/average daily purchase	647	2.98	3.40	0.02	27.69	1.92	0.29	8.73
Risk characteristics								
Risk measure°	782	0.32	0.47	0	5.23	0.19	0	1.04
Risk=0	782	0.12	---	---	---	---	---	---
Days of advanced saving (good weeks)	781	4.57	3.85	0	30	4	0	10
Days of advanced saving (bad weeks)	781	5.58	4.04	0	30	5	1	15
Number of times needed help in past year	782	0.60	1.22	0	12	0	0	2
Never needed help	782	0.68	---	---	---	---	---	---
Number of times provided help in past year	782	0.72	1.40	0	20	0	0	3
Never provided help	782	0.64	---	---	---	---	---	---
Risk heterogeneity §								
Risk heterogeneity (levels)	782	-0.03	0.51	-2.70	3.23	-0.05	-0.73	0.72
Absolute value	782	0.34	0.38	0.00	3.23	0.22	0.02	1.02
Group completely homogeneous	210	0.03	---	---	---	---	---	---
Risk heterogeneity (deciles)	782	-0.26	3.16	-8.00	8.00	-0.64	-4.97	4.97
Absolute value	782	2.68	1.70	0.00	8.00	2.38	0.50	5.70
Group completely homogeneous	210	0.05	---	---	---	---	---	---

* Standard deviation, minimum, maximum, median, and 5% and 95% points are not reported for dummy variables.
† As a percentage of total number of groups.
° Fraction of expected sales on Monday to Wednesday to cover amount of loan left to save on Sunday night (see text).
§ Average distance between a borrower's risk and the risks of his partners, using either risk levels or risk decile.

Table 2. Estimation of First-Best Risk r* and Friction f.

estimated equation: $r_i = X_i\ \alpha + W_i\ \phi\beta + \varepsilon_i^f$

Variables	OLS Parameter	OLS Std. Err.*	CLAD† Parameter	CLAD† Std. Err.**
	Dependent variable: observed risk level r_i			
Determinants of first-best risk (X)				
No education	0.151	(0.062)	0.070	(0.043)
Primary education	0.118	(0.050)	0.058	(0.038)
Secondary education	0.179	(0.077)	0.112	(0.054)
No work on weekend	-0.938	(0.221)	-0.483	(0.205)
Number of houses	0.069	(0.033)	0.009	(0.025)
Woman	-0.013	(0.038)	-0.064	(0.033)
Log of weekend sales	-0.150	(0.027)	-0.095	(0.019)
Determinants of friction (W)				
Escuintla	0.099	(0.041)	0.127	(0.047)
Antigua	0.186	(0.179)	-0.055	(0.079)
Chimaltenango	-0.043	(0.071)	-0.069	(0.049)
Tecpan	0.246	(0.212)	-0.100	(0.244)
La Gomera	0.178	(0.086)	0.151	(0.087)
Santa Lucia	0.327	(0.051)	0.407	(0.054)
Years in market	-0.007	(0.002)	-0.002	(0.001)
Access to alt. Credit.	0.233	(0.103)	0.038	(0.052)
Alt. Credit x Woman	-0.176	(0.090)	0.045	(0.059)
Intercept	1.161	(0.190)	0.786	(0.142)
F-statistics for determinants of risk	4.63		4.80 ‡	
F-statistics for determinants of frictions	8.64		9.80 ‡	
Number of observations	759		739	
R-squared/Pseudo R-squared	0.14		0.12	

Notes:

* OLS: Robust standard errors reported, correcting for clustering effects by group. F-statistics based on robust variance-covariance matrix.

** CLAD: Standard errors are bootstrapped with 1000 replications, correcting for clustering. The bootstrap procedure only sampled one individual from each group to correct for the clustering effects within groups. The coefficients from each round of the bootstrap were kept. They were then used to calculate an empirical variance-covariance matrix. The F-statistic was calculated as $(b*\Sigma^{-1}*b)$ /dim (b), where b is the vector of coefficients, Σ^{-1} the inverse of the variance-covariance matrix of b, and dim (b) is the size of b.

† CLAD: Convergence of the estimation procedure is defined to be when two consecutive iterations give parameter estimates which differ by less than .1 standard deviations, i.e., $D'(V^{-1})D<.1$ where D is the change in parameter estimates, and V^{-1} is the inverse of the variance-covariance matrix of the parameters. The number of observations reported are from the last iteration.

‡ CLAD: The F-statistics were constructed from the bootstrapped Variance-Covariance matrix of the estimates.

Table 3. IV First stages for First-Best Risk and Friction Measures

Variables	Risk and friction measures derived from OLS estimation Parameter	Std. Err.*	Risk and friction measures derived from CLAD estimation Parameter	Std. Err.*
First stage for risk measure.	**Dependent variable: estimated r***			
Migrant status†	0.812	(0.238)	0.588	(0.153)
Educational Level†	1.226	(0.119)	0.630	(0.075)
Years of Experience†	-0.095	(0.055)	-0.057	(0.036)
Discount rate†	0.068	(0.143)	0.175	(0.092)
No work on Sunday†	-2.194	(0.291)	-1.412	(0.199)
No access to money-lenders	-0.046	(0.033)	0.001	(0.017)
Money lender interest rate	0.054	(0.025)	0.020	(0.013)
Daily interest rate reported	-0.128	(0.049)	-0.062	(0.028)
Number of family members†	0.125	(0.141)	0.049	(0.082)
Times asked for help†	-0.019	(0.086)	-0.027	(0.051)
Intercept	-0.915	(0.010)	-0.646	(0.006)
F-Statistics***	22.50		19.52	
First stage for friction measure	**Dependent variable: estimated βf**			
Migrant status†	0.365	(0.269)	0.112	(0.243)
Educational Level†	-0.184	(0.115)	-0.092	(0.085)
Years of Experience†	0.011	(0.069)	0.036	(0.062)
Discount rate†	0.169	(0.164)	-0.867	(0.173)
No work on Sunday†	0.205	(0.349)	1.233	(0.330)
No access to money-lenders	0.088	(0.044)	0.065	(0.022)
Money lender interest rate	-0.082	(0.033)	-0.047	(0.016)
Daily interest rate reported	0.077	(0.057)	0.050	(0.022)
Number of family members†	-0.443	(0.196)	-0.076	(0.208)
Times asked for help†	-0.231	(0.107)	-0.407	(0.099)
Intercept	0.423	(0.012)	0.233	(0.009)
F-statistics***	5.07		6.37	

Notes:

† the explanatory variable is the average risk over all observations of the same category (same migrant status, same educational level, etc.), excluding the borrower himself and the members of his group.

* Robust standard errors reported, correcting for clustering effects by group.

*** The F-tests report the predictive power of the instruments in the first stage of the IV estimation, testing the hypothesis that these ten instruments are jointly insignificant against an alternative that at least one has some predictive power. Had there been exogenous variables in the second stage, they would have been included in the first stage regression (though omitted from the F-test). However, there are no exogenous variables in the second stage.

Table 4. LEVEL BASED HETEROGENEITY MEASURE h_i :
TSLS Estimates of Heterogeneity instrumenting Risk and Friction in First Stage

	Risk and friction measures derived from OLS estimation			Risk and friction measures derived from CLAD estimation		
Variables	Parameter	Std. Err.*	t	Parameter	Std. Err.*	t
			Dependent variable: risk difference h			
Risk heterogeneity (levels) without instrumenting						
Risk	0.516	(0.119)	4.3	0.715	(0.196)	3.7
Friction	0.122	(0.103)	1.2	0.047	(0.053)	0.9
Intercept	0.382	(0.108)	3.5	0.408	(0.125)	3.3
			Dependent variable: risk difference h			
Risk heterogeneity (levels)						
Risk	0.386	(0.196)	2.0	0.439	(0.318)	1.4
Friction	-0.825	(0.307)	-2.7	-0.308	(0.423)	-0.7
Intercept	0.638	(0.238)	2.7	0.314	(0.233)	1.3
OID test:‡						
TR2	3.31			8.18		
P-value	0.91			0.42		

* Robust standard errors reported, correcting for clustering effects by group.
‡ the OID (overidentification) test-statistics are distributed as a Chi-squared with 8 degrees of freedom (10 instruments, minus 2 endogenous variables) under the hypothesis that the instruments are valid.

Table 5. DECILE BASED HETEROGENEITY MEASURE h_i^d :
TSLS Estimates of Heterogeneity instrumenting Risk and Friction in First Stage

	Risk and friction measures derived from OLS estimation			Risk and friction measures derived from CLAD estimation		
Variables	Parameter	Std. Err.*	t	Parameter	Std. Err.*	t
			Dependent variable: risk difference h_i^d			
Risk heterogeneity (deciles) without instrumenting						
Risk	3.82	(0.829)	4.6	5.34	(1.364)	3.9
Friction	0.48	(0.597)	0.8	0.38	(0.373)	1.0
Intercept	2.97	(0.756)	3.9	3.02	(0.860)	3.5
			Dependent variable: risk difference h_i^d			
Risk heterogeneity (deciles)						
Risk	3.07	(1.311)	2.3	4.62	(2.130)	2.2
Friction	-1.49	(1.060)	-1.4	1.29	(2.193)	0.6
Intercept	3.07	(1.421)	2.2	2.35	(1.553)	1.5
OID test:‡						
TR2	7.13			7.54		
P-value	0.52			0.48		

* Robust standard errors reported, correcting for clustering effects by group.
‡ the OID (overidentification) test-statistics are distributed as a Chi-squared with 8 degrees of freedom (10 instruments, minus 2 endogenous variables) under the hypothesis that the instruments are valid,

Table 6. Two-step procedure for consistent γ

TSLS Estimates of Heterogeneity instrumenting Risk and Friction in First Stage

Variables	Parameter	Std. Err.*	t	Parameter	Std. Err.*	t
	Dependent variable: h_i			**Dependent variable: h_i^d**		
Risk heterogeneity						
Risk	0.578	(0.221)	2.6	3.567	(1.610)	2.2
Friction	-0.257	(0.246)	-1.0	0.920	(0.854)	1.1
Woman (V)	-0.027	(0.038)	-0.7	0.129	(0.205)	0.6
Intercept	-0.370	(0.246)	-1.5	0.034	(0.837)	0.0
OID test:‡						
TR2	5.26			8.14		
P-value	0.73			0.42		

* Robust standard errors reported, correcting for clustering effects by group.

‡ the OID (overidentification) test-statistics are distributed as a Chi-squared with 8 degrees of freedom (10 instruments, minus 2 endogenous variables) under the hypothesis that the instruments are valid, and are detailed in the Appendix.

Table 7. Insurance Occurances in Credit Groups

Insurance in past year. §	**Groups providing insurance**		%	
	none	82	0.39	
	1-4 times	70	0.33	
	more than 4 times	58	0.28	
		210		
	Groups receiving insurance		%	
	none	79	0.38	
	1-4 times	85	0.40	
	more than 4 times	46	0.22	
		210		
How much?	**Amount (in $):**			
	mean	28		
	median	17		
	5%	5		
	95%	88		
	min	2		
	max	167		
	As fraction of payment:*			
	mean	0.24		
	median	0.17		
	5%	0.04		
	95%	0.67		
	min	0.01		
	max	1.25		
Reason insurance was needed.		N	%	
	low sales	155	0.63	
	bad planning	30	0.12	
	robbery	7	0.03	
	family illness	49	0.20	
	other	7	0.03	
		248		
Who provides insurance?		N	%	
	a member of the group	128	0.52	
	the whole group	42	0.17	
	someone from outside	49	0.20	
	self insurance	8	0.03	(savings, money-lender)
	resulted in late payment	20	0.08	(insurance failed)
		247		

§ The difference between the number of groups reporting providing insurance to their members and the number of groups with members reporting receiving insurance is due to membership turnover.

* As we have only current payment information, and not the payment information at the time insurance was given, these are only approximations.

Table 8. Insurance Flows

Net intragroup insurance flows.

	TOTAL		Conditional on ins. activity
	N	%	%
insurance net provider	172	0.22	0.48
insurance net recipient	117	0.15	0.33
balanced insurance	67	0.09	0.19
no insurance activity	426	0.54	
	782	1.00	0.00

Intragroup insurance.

By risk status in group.†

	relatively safe †		relatively risky		TOTAL
	N	%	N	%	%
insurance net provider	98	0.44	74	0.56	0.48
insurance net recipient	83	0.37	34	0.26	0.33
balanced insurance	42	0.19	25	0.19	0.19
	223	1.00	133	1.00	1.00

Pearson Chi2 ‡ = 5.8 p-value = 0.055

By risk status in population.§

	net provider		net recipients	
Risk quantile	N	%	N	%
1	47	0.27	19	0.16
2	31	0.18	24	0.21
3	29	0.17	20	0.17
4	34	0.20	19	0.16
5	31	0.18	35	0.30
	172	1.00	117	1.00

Pearson chi2 ‡ = 8.8 p-value = 0.067

† Relatively safe borrowers are borrowers with predicted r* small than the median predicted r* in their group.
‡ The Pearson Chi2 tests the null hypothesis that the " relatively safe" and "relatively risky" borrowers are not significantly different in their net insurance provision.
§ Borrowers are divided in risk quantiles according to their estimated r^*. Net-providers provided insurance more times than they received; net-recipients received insurance more times than they provided.

Table 9. Homogeneous groups of safe borrowers

	Borrowers in safe homogenous groups		Others				
	Mean	Std. Dev.	Mean	Std. Dev.	Difference	t-statistic	One-sided p-value
Insurance activity	**N = 23**		**N = 759**				
Num. times provided	1.65	2.02	0.85	1.75	0.80	1.88	3.6%
Num. times received	0.83	1.40	0.68	1.42	0.14	0.48	31.8%
Num. times prov./member	0.44	0.61	0.24	0.53	0.20	1.54	6.9%
Num. late payments*	0.35	0.17	0.80	0.04	-0.45	-2.52	0.9%
Years in group	1.33	0.24	2.44	0.07	-1.11	-4.51	0.0%
Group characteristics	**N = 6**		**N = 204**				
Group size	3.83	0.75	3.72	0.94	0.11	0.36	36.7%
Average risk	0.04	0.04	0.33	0.39	-0.30	-9.07	100.0%
Late payments (yes=1)	0.17	0.41	0.50	0.50	-0.34	-1.99	5.0%
Num. late payments	0.50	1.22	1.15	1.55	-0.65	-1.27	12.7%
Age of group	1.35	1.19	3.16	1.98	-1.81	-3.60	0.6%
Groups < 1 yr old			**N = 51**				
Group size			3.75	0.89	0.09	0.27	39.9%
Average risk			0.25	0.22	-0.22	-6.16	0.0%
Late payments (yes=1)			0.51	0.50	-0.34	-1.90	5.0%
Num. late payments			0.84	1.01	-0.34	-0.66	26.7%
Business characteristics	**N = 23**		**N = 759**				
Saves # days before pyt †							
Good weeks	6.22	2.30	4.52	5.03	1.70	3.41	0.1%
Bad weeks	6.78	2.11	5.54	4.08	1.24	2.67	0.6%
Purchases every # days	8.54	5.26	6.08	5.78	2.46	2.21	1.9%
Weekly sales (in US$)	970	504	675	1154	295	2.61	0.7%
Prop. selling on Sundays	1.00	0.00	0.88	0.33	0.12	10.26	0.0%
Sunday sales	229	156	118	116	112	3.41	0.1%
Ratio weekly sales (Bad/Good)	0.58	0.16	0.58	0.17	0.00	0.09	46.6%
Age of business	6.0	5.0	8.5	7.0	-2.48	-2.32	1.4%
Individual charactistics	**N = 23**		**N = 759**				
Age	37.9	12.9	37.7	10.8	0.18	0.07	47.4%
Gender (Male=1)	0.65	0.49	0.56	0.50	0.09	0.92	18.4%
Marital status (Married=1)	0.30	0.47	0.82	0.41	-0.51	0.92	18.3%
House ownership (Yes=1)	0.83	0.39	0.75	0.43	0.07	0.86	19.8%
In city	0.30	0.47	0.43	0.50	-0.13	-1.29	10.4%
In own village	0.57	0.51	0.34	0.48	0.22	2.07	2.5%
Years of education	4.4	3.1	3.5	3.2	0.95	1.45	8.0%
No education	0.17	0.39	0.29	0.45	-0.11	-1.37	9.2%
Not born in workplace	0.83	0.39	0.86	0.35	-0.03	-0.35	36.3%
Less than 5 years in workplace	0.13	0.34	0.10	0.31	0.03	0.36	36.0%
Less than 2 years in workplace	0.00	0.00	0.04	0.01	-0.04	-5.68	0.0%
Age start working	29.04	10.86	24.66	10.20	4.38	1.91	3.4%
Long horizon ‡	0.48	0.51	0.32	0.47	0.16	1.49	7.5%

* The same number of late payments in certain groups is not the same for all members.

† The definition of "good weeks" and "bad weeks" was left to the discretion of the borrowers.

‡ Long horizon borrowers had explicit plans at a 5 year horizon. See text.

Table 9. Homogeneous groups of safe borrowers (continued.)

	Borrowers in safe homogenous groups		Others				
	Mean	Std. Dev.	Mean	Std. Dev.	Difference	t-statistic	One-sided p-value
Loan characteristics							
Loan size	630	238	745	564	-115	-2.14	2.0%
Loan, as fraction of average*	0.87	0.35	1.00	0.69	-0.13	-1.70	5.0%
Loan, as frac. of avg, grp <1 yr. old†	0.87	0.35	0.81	0.51	0.06	0.75	22.9%
Monitoring							
Prop. partners in same seccion of mkt	0.68	0.34	0.58	0.37	0.11	1.46	7.9%
Prop. partners borrowers can see	0.68	0.34	0.45	0.36	0.23	3.15	0.2%
Visiting partners is important (Yes=1)	0.96	0.10	0.87	0.25	0.09	3.94	0.0%
Screening							
Prop. partners borrowers knew before	0.97	0.06	0.90	0.18	0.07	4.90	0.0%
Prop. partners who are friends outside	0.88	0.23	0.84	0.26	0.05	0.96	17.4%
Prop. partners who are *paisanos*	0.41	0.33	0.46	0.37	-0.04	-0.59	27.9%
Prop. partners from same family	0.30	0.33	0.41	0.34	-0.11	-1.51	7.3%
Outside options							
Credit from friends/family	0.17	0.39	0.52	0.50	-0.35	-4.23	0.0%
Credit from friends and moneylenders	0.04	0.21	0.33	0.47	-0.28	-6.06	0.0%
Wholesale credit	0.48	0.51	0.27	0.45	0.20	1.89	3.6%

* Loan size compared to the average loan size, by activity

† Loan size compared to the average loan size of groups less than one year old, by activity