Bogus Joint Liability Groups in Microfinance*

Alexander Karaivanov† Simon Fraser University

Xiaochuan Xing Yale University

Yi Xue UIBE

November 2019

Abstract

In a random sample of clients of CFPAM, the largest microlender in China, 73% of all joint-liability groups practice *Lei Da Hu*. That is, one person uses all group members' loans in a single project. We call such borrower groups 'bogus groups'. The Lei Da Hu practice violates a key premise of group lending, that each borrower must use their loan in a separate project (what we call 'standard group'). We extend the theory of group lending by analyzing the endogenous formation and coexistence of standard and bogus groups and characterize the efficient lending terms. The chosen group form depends on the borrower productivities and probability of success. Bogus groups are formed by heterogeneous borrowers, when the gains from larger expected output exceed the foregone default risk diversification. Accounting for bogus groups in their lending strategy can help MFIs raise productive efficiency and borrower welfare.

JEL classification: C72; D82; G20; O12

Keywords: microfinance; group lending; joint liability; Lei Da Hu; phantom borrowers; strategic default

*We thank audiences at the ThReD, SED and Econometric Society meetings for many useful comments and suggestions. Karaivanov acknowledges financial support from the Social Sciences and Humanities Research Council grant 435-2013-0698. Xue acknowledges financial support from the National Natural Science Foundation of China project 71471040.

†Corresponding author: Department of Economics, Simon Fraser University, akaraiva@sfu.ca

1 Introduction

Data from a random sample of clients of the largest Chinese microlender, the China Foundation for Poverty Alleviation – Microfinance (CFPAM) show that 73% of the surveyed joint-liability groups report practicing *Lei Da Hu*. That is, a single borrower uses the total of all loans given to the group members while the rest of the group only act as cosigners. We call such borrower groups *bogus groups*.

Bogus groups violate a basic premise of group lending in microfinance, both as implemented in practice and as modeled in the economics literature – the requirement that each borrower invest their loan in their own business project (what we call *standard group*). Many authors have emphasized the benefits of joint-liability group loans, stemming from leveraging social capital and reducing default rates, as the borrowers cover for each other and share the risk of inability to repay. In contrast, in a bogus group all funds are pooled in a single business or investment and, if the cosigners are poor, because of limited liability, their nominal role as guarantors offers no protection to the lender.[1] It is therefore surprising that such a large fraction of CFPAM groups report to be bogus since, according to the theory, group lending, in its standard form, can reduce agency costs and increase the probability of repayment and obtaining future credit. In addition, the Lei Da Hu practice directly violates the CFPAM loan covenant terms, although such infractions may be hard to verify.

The existing theoretical literature on microcredit has focused almost exclusively on comparing group lending with individual lending.[2] While bogus groups share a common characteristic with individual loans in the fact that the probability of repayment depends on the success of a single investment, the two differ in several important aspects. First, bogus groups allow the borrowed funds to be invested in a project that is selected among all group members' projects (e.g., the project with the highest expected return). Second, by pooling the members' loans, bogus groups allow a larger scale of investment.

To the best of our knowledge we are the first to formally model and analyze Lei Da Hu microcredit groups and compare them to standard joint-liability groups (and individual loans, in an extension). More generally, we explore how capital is allocated within borrower groups and the basic economic incentives to direct capital toward high marginal product investments vs. risk diversification – a trade-off that has not been studied in the literature. Our paper also relates to a growing literature arguing that the canonical microcredit model may be too rigid and constraining for entrepreneurship or growth. For example, Field et al. (2013) and Aragon et al. (2019) show that more flexible repayment terms (grace period or credit line, respectively) can improve the borrowers' economic outcomes. Similarly, we argue that allowing flexibility in pooling funds can be beneficial.

We characterize the borrowers' choice of standard vs. bogus group form, the possible coexistence of both group types, and the lenders' choice of loan terms when the group form (standard or bogus) is unobserved to the lender.[3] The key trade-off which determines the group form choice and the offered loan terms is as follows. Borrowers can pool all funds into a single project, with the highest return, in a bogus group, or

[1] The cosigning members may agree to join a bogus group because the person who uses the funds may offer monetary or other compensation or because they may wish to ask this person to cosign a future loan.

[2] See the literature review at the end of this section.

[3] This assumption relates more generally to the literature on moral hazard in credit markets, e.g., unobserved effort, unobserved loan use, or unobserved borrowing from other lenders.

diversify across multiple projects with uncorrelated returns, in a standard group. The former choice yields higher expected output, the latter lowers the risk of default. The lower default risk has three benefits: greater probability of continued access to credit, lower interest rate and a larger feasible loan size. The larger loan size in a standard group results because there is a maximum incentive-compatible repayment amount and hence the lower is the risk premium (interest rate), the more room there is for the loan principal to be larger. Essentially, the borrowers and lenders face a risk-return trade-off and the lending terms must be designed accordingly.

Specifically, we model borrowers, each of whom is endowed with a single investment project. The projects are heterogeneous in their productivity (return) which may or may not be observable to the lender. Each project either succeeds (yields positive output), with some probability, or fails (yields zero). The borrowers have no other funds and must obtain credit from a microfinance lender. Consistent with our motivating example of CFPAM, we assume a non-profit / NGO lender that makes zero expected profits per each offered loan.[4] Only group loans are offered (as in the CFPAM data) and, for simplicity, we assume that the loans are made to groups of two borrowers (in Section 5 and Appendix C we allow individual loans and more than two borrowers per group). The project outcomes are assumed i.i.d. across the borrowers.[5] Project output is non-verifiable by the lender. Thus, a borrower can strategically default (declare project failure) if she finds this optimal. In case of project failure we assume that the lender cannot collect anything from that borrower – there is limited liability.

After receiving their loans, the borrowers choose to operate either as a standard group (each invests in her own project) or as a bogus group (one person uses all loaned funds). The group form is unobserved by the lender. There is joint liability – all borrowers bear full responsibility for the group's total debt. If full repayment is not received by the lender, all group members are excluded from access to future credit. In our baseline model the group members make a joint decision whether to repay or not. We consider non-cooperative repayment decisions in Appendix B.

We first model standard joint-liability groups, in which each borrower invest in their own project, and characterize the loan terms that maximize the group payoff subject to no strategic default. We then study bogus groups, in which all loans are pooled into a single project, and compare them to the standard groups. We show that a competitive lender who offers the standard-group loan terms to everyone would incur a loss when some borrowers (those with heterogeneous projects) run bogus groups.

We then analyze the contract design problem of a lender who takes into account that bogus and standard groups form endogenously and can coexist, but who cannot observe the group form choice. In our baseline setting with observable project productivities (Section 4.1), the loan terms (size and repayment amount) and the chosen group form (bogus or standard) depend on the group composition. Specifically, for borrowers with homogeneous (equally productive) projects there is no expected output gain from forming a bogus group and, hence, the efficient, subject to no default, loan size and interest rate are offered. In contrast, for borrower groups with heterogeneous projects, the offered loan terms depend on the productivity differential between the projects. If the differential is relatively small, the lender offers loan terms that induce a standard

[4] We explore allowing for cross-subsidization in Section 4.4.

[5] We assume that the borrower groups are formed before the project productivities are known (exogenous matching). In Section 5 we analyze the implications of allowing endogenous matching.

group, taking advantage of risk diversification. The loan size is, however, reduced relative to that offered to homogeneous groups, to prevent the borrowers to operate as bogus group. The reduced loan size reflects the welfare loss from asymmetric information. In contrast, for a large productivity differential, it is optimal to forego diversification and offer loan terms that induce a bogus group, that is, the lender encourages all funds to be invested in the highest-yield project.

We show that our main results remain qualitatively unchanged when the project productivities are unobserved by the lender (Section 4.2). In that setting, in addition to the moral hazard problem of unobserved group form choice, the lender must also address the adverse selection problem of groups possibly misreporting their project composition type. There are two possible outcomes, depending on the model parameters. Bogus groups optimally result when the productivity differential across the members' projects is sufficiently large. Otherwise, for small or zero productivity differential, the borrower payoff maximizing loan terms induce standard group form for all borrowers, however, with reduced loan size to ensure incentive compatibility. In Section 4.4 we also explore the possibility of a lender who can cross-subsidize across offered loans but is still subject to asymmetric information about the project productivities and group form. We show that such lender would offer the same loan terms to all borrowers and that bogus and standard groups would co-exist.

In Section 5 we consider several extensions of the baseline model. We first show how the analysis can be generalized to endogenize the continuation value of future credit in an infinite horizon setting (Section 5.1). In Section 5.2 we show an isomorphic version of our model in which repayment is supported by penalties for default. In Section 5.3 we study endogenous sorting by the borrowers into groups. We find that positive assortative matching by project type results and all borrower groups are standard if the project productivity differential is small, while negative assortative matching and bogus groups results if the productivity differential is sufficiently large. In Section 5.4 we consider the possibility that, for exogenous reasons, the lender is restricted to offer the same standardized loan to all borrowers and show that this can be efficient in certain cases, namely with unobservable productivities and a small fraction of heterogeneous groups, but is inefficient otherwise, when offering a menu with different loan terms can yield higher expected payoff. In Section 5.5 we discuss individual loans and show that, in our setting, they are dominated by group lending – that is, no group of borrowers would be better off by switching to individual loans.

Related literature

There is a large literature on joint liability microfinance and the comparison between group lending and individual lending (Ghatak, 2000; Chowdhury, 2005; Gangopadhyay et al., 2005; Banerjee et al., 1994; Rai and Sjostrom, 2004 among many others). In both theory and practice, the main advantage of joint liability group lending is that it can create a substitute for asset collateral by using the social capital embedded in the borrowers' networks and relationships to mitigate moral hazard, adverse selection, costly state verification or debt enforcement problems (Mosley, 1986; Udry, 1990; Besley and Coate, 1995; Morduch, 1999; Ghatak, 1999; Ahlin and Waters, 2016). In comparison to individual loans, the joint liability design also allows lending at lower interest rates, due to the higher repayment rate enabled by risk diversification, peer selection, peer monitoring, and peer enforcement within the group (Ghatak and Guinnane, 1999; Karlan, 2007; Besley and Coate, 1995; Armendariz de Aghion, 1999; Ghatak, 2000; Stiglitz, 1990). With limited enforcement, however, joint liability may lower the repayment rate relative to individual lending when a borrower is unable to pay for another member (Besley and Coate, 1995). De Quidt et al. (2016) further show that individual lending can be welfare improving compared to joint liability when borrowers have sufficient social capital to sustain mutual insurance.

In the empirical literature, Gine and Karlan (2014) find no significant difference in repayment rates between group and individual loans in the Philippines. Ahlin and Townsend (2007a) use data from Thai borrowing groups and find a U-shaped relationship between borrower's wealth and the prevalence of joint-liability relative to individual loans. Ahlin and Townsend (2007b) further document that repayment rates are negatively affected by the joint liability rate and social ties and positively affected by the strength of local sanctions or correlated returns. De Quidt et al. (2018) explore commercialization in microfinance and the perceived decline in joint liability lending. Using MIX Market data they find no evidence for a decline but document an increase in competition and shift from non-profit to for-profit lending which, they argue theoretically, can cause lenders to reduce using joint liability loans.

The above papers assume that each borrower invests in a separate project, most often with uncorrelated returns across the borrowers (what we call standard groups).[6] The various authors mostly focus on the comparison and the possible (dis-)advantages of joint liability loans relative to individual loans. In contrast, motivated by the evidence from China, we model and analyze the endogenous formation of bogus joint liability groups in which all loans are pooled into a single project and their coexistence alongside standard joint liability groups.

Our paper is related to Bernhardt et al. (2019) who document, using data from Ghana, India and Sri Lanka, that wives often invest the grants and loans they receive into their husbands' enterprises which may be more profitable compared their own enterprises. Their results offer a possible explanation for the findings of de Mel et al. (2008) and others, that male but not female-operated enterprises benefit from access to cash grants. The authors emphasize the importance of measuring returns at the household level, as opposed to at the enterprise level, as we do here by focusing on the group incentives in allocating the borrowed funds.

Further evidence on the importance of Lei Da Hu and similar practices comes from Pakistan where, in a policy report, Burki (2009) observes that: "...often, the group leader had accessed more loans from an MFI than the MFI had record of by borrowing through a dummy or ghost borrower". Also related is the qualitative study by Cieslik et al. (2015) on 'unruly' entrepreneurs in rural Burundi.[7] The authors describe strategies used by poor entrepreneurs to bypass institutional microcredit rules. One strategy is "loan arrogation", in which an entrepreneur asks another community member to obtain additional credit to be invested in the entrepreneur's business. The authors argue that this allows for "...larger-scale investment, cementing social bonds and empowerment", and more generally, that such illicit practices can be interpreted as value creating.

2 Microfinance and Lei Da Hu in China

Microfinance was introduced in China in the early 1990s with the primary goal of alleviating rural poverty. Since 1996 the government has regarded microfinance as an effective channel of credit provision to the poor

[6] An exception is Banerjee et al. (1994)'s work on German cooperatives in which only one member is assumed to have an investment opportunity while the other member serves as a guarantor and monitor. In contrast, our focus is the choice of group form itself.

[7] We thank J-M. Baland for this reference.

(Zhang et al., 2010). Non-profit and NGO microfinance institutions (MFIs) play a major role in this process. The main non-profit MFI in China, CFPAM (China Foundation for Poverty Alleviation – Microfinance) is the largest microlender by total issued loans and active members.[8] CFPAM has been using joint-liability lending since its founding, offering only group loans until 2014. Recently most Chinese MFIs have introduced individual or other loan forms alongside group lending.[9] This could be partly because of the Lei Da Hu phenomenon that we study and also reflects general trends worldwide, as microlenders expand their outreach beyond the extremely poor, who were traditionally targeted.

Microcredit in China is commonly regarded as constrained by financial regulation, lack of supply of wholesale funds, and credit risk. The latter is manifested in part as Lei Da Hu – bogus microfinance groups or ‘phantom’ borrowers. According to the Bank for International Settlements (BIS) (2010), “...the complexity and distinctive features of microlending, especially the decentralized lending process, raise important risk management issues for microfinance activities and institutions”. In particular, the BIS report mentions the problem of phantom borrowers, in whose presence the risk diversification and peer monitoring benefits of group lending are weakened or absent and the resulting increase in credit risk can cause MFIs to incur losses. Another policy report, by Planet Rating, a microfinance rating agency, concludes that CFPAM’s control functions regarding Lei Da Hu practices were “...not sufficiently formalized and existing control forms are not utilized” (Planet Rating, 2005).[10] These reports, together with the additional evidence from Pakistan and Burundi discussed at the end of Section 1, suggest that Lei Da Hu or phantom borrowers can be an important factor affecting MFI’s credit risk and performance.

As further motivating evidence for the importance of Lei Da Hu / bogus groups, we use a confidential random sample of 81 CFPAM group loans issued in 2011 to 353 female borrowers in three of the poorest counties in Liaoning province in China – Beizhen, Xiuyan and Xingcheng. CFPAM advertises regularly in rural China and most people are aware of its microcredit program. Interested borrowers first form a group and then approach CFPAM. The CFPAM lending rules stipulate that each group must consist of 2 to 5 self-chosen members from the same village. There should be no more than one member from the same household per group and it is also not desirable for close relatives to be in the same group. After a group is created it elects a group leader from its members. If the group meets the basic requirements (each member has an existing business or business plan, understands the credit rules and needs a loan), CFPAM holds a training session explaining joint liability, group operations, the importance of group solidarity, monitoring, and meeting attendance by all members. Then each group member receives their first loan. The whole process typically takes a week.

Our random sample of CFPAM loans includes loan identification number, starting date, size, duration, required monthly payment, interest rate and proposed loan use (by the applicant). There is also basic demographic information about the borrowers, including age, gender, ethnicity, education and marital status. According to CFPAM loan officers, some of the self-reported information is unverifiable, e.g., the proposed

[8] In 2015 CFPAM issued 324,228 loans with total value 4.13 billion RMB and served 306,101 borrowers.

[9] See for example: http://www.rong360.com/gl/2015/11/24/82.097.html. Grameen Bank – China has kept group lending as its main lending practice but borrower groups are now formed using an online platform, instead of the traditional way of grouping people closely familiar with each other.

[10] A possible interpretation is that the lender was aware of Lei Da Hu practices but chose not to act. This is consistent with our theoretical results when the efficient outcome has co-existing standard and bogus groups.

loan use or a borrower's education. In addition to the administrative CFPAM data, each borrower in the sample was called by phone at randomly chosen time during working hours by the survey team and asked:[11] "Does your borrower group practice Lei Da Hu?" The survey team explained the meaning of Lei Da Hu if a borrower was unaware. In 59 of the 81 groups (73%) at least one member reports that their group practices Lei Da Hu; in 54 groups *all* borrowers report so; and in 56 groups the majority (50% of more) of the members report so. The possibility of phantom group members is ruled out by the strict documentation and attendance requirements and the phone interviews. We thus do not model fabricating borrowers to enlarge one's available credit (as mentioned in some of the references above) but the different issue of pooling credit within an existing legitimate group.

Unfortunately, the CFPAM data have serious limitations – the small sample size and available variables do not allow formal statistical analysis.[12] There is almost no variation in the interest rate and repayment schedule: 79 of the 81 loans have the same interest rate (13.5%) and only 2 loans have a different interest rate (12% and 16%). Only 1 of the 81 loans (the outlier with 16% interest) has a different total required number of repayments (4, versus 10 for the rest). The loan size and corresponding monthly required repayment vary across the borrowers but do not differ in a statistically significant way with the reported group form (bogus or standard). Sample statistics are reported in Table A1 in Appendix C.

The main patterns observed in the CFPAM data are:

(i) bogus groups (self-reported Lei Da Hu practice) constitute a large fraction of the random sample and coexist with standard groups.

(ii) being in a bogus group is not correlated with the observed borrower characteristics: age, marital status, education (% with college is the only exception), ethnicity and reported intended loan use; although it could be associated with other unobserved by us factors, such as project quality or heterogeneity.

(iii) CFPAM offered similar loan terms to all groups.

Facts (i) and (ii) motivate our theoretical model of bogus groups and their coexistence with standard groups. The observed homogeneity of loan terms in the sample (fact iii) is consistent with our results in Section 4.4 or with the hypothesis that CFPAM was unaware or ignored bogus groups and offered standardized loan terms to all borrowers. Our earlier discussion about losses incurred by Chinese MFIs from phantom borrowers / Lei Da Hu lends indirect support for this. Unfortunately, we do not have direct data on default rates or subsequent survey rounds to test this formally.

3 Model

3.1 Setting

Consider an economy populated by lenders and borrowers. Each borrower has a single investment project requiring initial investment $L > 0$. The borrowers have no wealth. Hence, the entire initial investment $L > 0$ needed to implement the investment project must be financed at time $t = 0$ by taking a loan from a lender (microfinance institution). There are two types of projects: a 'conventional' project with productivity k_l and

[11] All borrowers were also asked and confirmed their awareness of the joint liability clause and knowing the other group members personally.

[12] We did try multiple econometric specifications.

a 'high-return' project with productivity k_h, where $k_h \geq k_l > 0$. The project output, Y_i for $i = l, h$ is generated at $t = 1$ and is stochastic:

$$Y_i = \begin{cases} k_i L & \text{with probability } p, \\ 0 & \text{with probability } 1 - p. \end{cases}$$

The parameter $p \in (0, 1)$ is interpreted as the probability of a project being successful. The project output realizations are assumed i.i.d. across borrowers. The borrowers' outside option (if they do not invest) is normalized to zero.

Both the lenders and the borrowers are risk-neutral. The lenders have an opportunity cost of funds normalized to 1. Because of information or enforcement frictions, the project return Y_i is non-verifiable by the lender. This allows the possibility of strategic default – a borrower can report failure while in fact her investment project has succeeded. In addition, the borrowers have limited liability: if the investment fails (yields zero), the borrower defaults involuntarily and the lender does receive anything. The loan terms must therefore be chosen so that borrowers are given incentive to pay back when their projects succeed.

As in our motivating example of CFPAM, we assume a non-profit / NGO microlender that earns zero profits. For most of the analysis we assume that for institutional, accounting, political or other reasons the lender cannot cross-subsidize across borrowers, that is, each loan offered by the lender breaks even. For example, in our data each loan is the responsibility of a loan officer who is evaluated based on the loan's performance. In Section 4.4 we discuss allowing cross-subsidization. We also assume that only group loans are provided. Allowing for individual loans is discussed as an extension in Section 5. As in most of the theoretical literature on microcredit, suppose groups consist of two members (we relax this in Appendix C). The group lending contract thus consists of two loans, each with size L and required repayment R, where $R \geq L$ (the gross interest rate is R/L). Since the borrowers have no wealth, requiring collateral is infeasible. Instead, the group loan has a joint liability clause – each member is fully responsible for the entire group obligation $2R$. If the lender does not receive $2R$ at $t = 1$, from one or both borrowers combined, then both borrowers are cut off from access to future credit.[13]

To keep the analysis tractable, suppose that the borrowers' project returns k_i are i.i.d. over time, e.g., each borrower draws a new project per credit cycle. That is, we can think of the borrower groups as being randomly 'reset' in any future period $t > 1$, that we do not model explicitly. The i.i.d. assumption allows us to focus on a single credit cycle (two periods, $t = 0$ and $t = 1$ only), since it implies that the *ex-post* continuation value of future credit access is the same for all borrowers who are not cut off from credit because of default.[14] We call this ex-post continuation (future-credit access) value V, where $V > 0$, and treat it as given hereafter. Note, however, that we do allow the *ex-ante* (expected) continuation value of future credit access to differ across borrower groups when their endogenous probability of repayment is different.

Each borrower group can operate as either a *standard group* or as a *bogus group*. In a standard group each member invests L into her *own* business project, as assumed in the literature and as required by the MFI

[13]This is standard assumption in the literature (e.g., Ghatak and Guinnane, 1999 or De Quidt et al. 2016) and corresponds to the maximum penalty for strategic default.

[14]Our emphasis is on the heterogeneity in project quality within borrower groups. In reality, the continuation value may also depend on other factors that we abstract from such as age, education, family status, business experience, etc.

loan terms in practice. In contrast, in a bogus group the members invest the total loan amount $2L$ into *one* of the two projects. The standard vs. bogus group decision is made jointly by the members to maximize the group payoff – there is transferable utility and (uncompensated) coercion by a powerful member is ruled out.

The project productivities k_i and k_j are known to both borrowers in a group but may be observed or unobserved by the lender (we study both cases). A possible interpretation of the observed productivity case is that most microlenders require detailed information about the investments that borrowers intend to implement before providing a loan. To simplify the notation, we will say that a group 'has type ij' if the productivities of its members' projects are k_i and k_j where $i, j \in \{h, l\}$. There are three possible group types, $ij \in \{hh, ll, hl\}$. Without loss of generality assume that $k_i \geq k_j$.

In our baseline setting we assume that the loan repayment decisions are made *jointly* by the group members, similar to Stiglitz (1990). The members repay if their expected joint payoff from keeping access to future credit exceeds the payoff from defaulting. This assumption can be motivated by the members being able to observe each other's project outputs or sharing sufficient social capital to enforce social penalties in case of uncoordinated strategic default. In Appendix B we also consider the alternative scenario of individual (non-cooperative) default decisions, similar to Besley and Coate (1995), which introduces strategic interaction and free riding on a partner's joint-liability obligation.

Timing

- Stage 0: a group of two borrowers is formed; then each borrower draws an investment project with productivity k_i, $i \in \{h, l\}$;[15]
- Stage 1: the lender offers a group loan with terms (L, R);
- Stage 2: the borrowers choose to operate as a standard or bogus group (unobserved to the lender);
- Stage 3: each borrower's investment is launched and the project output Y_i is realized one period later; Y_i is non-verifiable by the lender;
- Stage 4: repayment / default decision is made jointly by the group members;
- Stage 5: all payoffs are realized (see below for details).

3.2 Standard groups

We start the analysis with the basic setting from the literature in which bogus joint liability groups are exogenously ruled out and only standard groups exist. Assume for now that the group member's project productivities k_i and k_j with $k_i \geq k_j$ are observed/known by the lender (the case of unobservable productivities will be discussed later on).

We say that a group loan is feasible if each member's project generates sufficient output upon success to be able to cover $2R$, that is, paying for oneself and one's partner. If this condition did not hold, that is if output upon success were only sufficient to repay for oneself ($k_iL \in (R, 2R)$), then the probability of repayment

[15] This timing rules out endogenous matching between borrowers (see Section 5 for more discussion).

and access to future credit in a joint liability standard group would equal p, which is the same as that of an individual loan and that in a bogus group (see Section 3.3). Hence, there would be no risk diversification advantage from group lending and no meaningful trade-off between standard and bogus groups.

Feasibility requires $\min\{k_i L, k_j L\} \geq 2R$, or

$$R \leq \tfrac{1}{2} k_j L. \tag{1}$$

We make the following parameter assumption which, as shown below, ensures that the project returns are sufficiently large so that repayment is always feasible.[16]

Assumption 1 (feasibility)

$$k_h \geq k_l \geq \tfrac{2}{p(2-p)}$$

Assumption 1 implies

$$pk_i > 1 \text{ for } i = l, h \tag{SE}$$

that is, both projects are socially efficient – the expected payoff per dollar invested strictly exceeds the opportunity cost of funds.

In the repayment stage (Stage 4 in the timing) the borrowers choose between repaying the entire group liability $2R$ or repaying zero (default). Note that it is never optimal to make a partial repayment (an amount between 0 or $2R$) since, because of the joint liability clause, either defaulting (repaying zero) and forfeiting the continuation value $2V$, or repaying in full and securing $2V$ is dominant strategy. Thus, to maximize the joint group payoff, the borrowers repay if and only if[17]

$$R \leq V \tag{2}$$

Lending to a standard group increases the loan repayment probability from p (the project success probability) to $1-(1-p)^2 = p(2-p)$. This reflects the classic argument for joint-liability lending as compared to individual lending. For loan terms (L, R) satisfying (1) and (2), the expected payoff of a standard group of type ij equals

$$W_{ij}^S(L,R) = p(k_i + k_j)L - 2p(2-p)R + 2p(2-p)V. \tag{SEP}$$

The first term in (SEP) is the expected output in a standard group, the second term is the expected repayment and the third term, $2p(2-p)V$ is the expected continuation value of access to future credit.

The lender receives $2R$ with probability $p(2-p)$ and zero otherwise (limited liability). Hence, the lender's break-even condition is $2p(2-p)R - 2L \geq 0$, or

$$R \geq \tfrac{L}{p(2-p)}. \tag{3}$$

[16] Contrast with Besley and Coate (1995) or De Quidt et al. (2016) where a borrower may be unable to repay both loans.

[17] If both projects succeed, the borrowers repay if $(k_i + k_j)L - 2R + 2V \geq (k_i + k_j)L$. Similarly, if only project i succeeds, the borrowers repay if $k_i L - 2R + 2V \geq k_i L$.

The optimal loan terms for standard group $ij \in \{ll, hl, hh\}$ solve

$$\max_{L,R} W_{ij}^{S}(L, R) \tag{SP}$$
$$\text{s.t. (2) and (3)}$$

The break-even condition (3) must bind at optimum, that is $R = \frac{L}{p(2-p)}$. If not, the lender can offer a loan with lower repayment R and still break even while increasing the borrowers' payoff. The feasibility condition (1) is then equivalent to $\frac{L}{p(2-p)} = R \leq \frac{1}{2}k_j L$, which is satisfied given Assumption 1.

Proposition 1

(a) The loan size and repayment terms solving problem (SP) are (L_S, R_S) *with*

$$L_S = p(2-p)V \text{ and } R_S = V.$$

(b) The same loan terms (L_S, R_S) *would be offered when the project productivities are unobservable to the lender.*

Proof: see Appendix A.

The reason for the result in (b) is that the loan terms (L_S, R_S) are determined by the break-even and no-default conditions, neither of which depends on the project productivities k_i and k_j. This also implies that our assumption of both borrowers receiving the same loan size even though they may have different k's is not restrictive.

3.3 Bogus groups

A standard assumption in the theoretical literature on joint liability lending and also standard practice in microfinance is that each borrower is expected to invest in her own business project. However, motivated by the evidence reviewed in the introduction, suppose that lenders are unable to enforce or verify the requirement that each group member invest in her own project, hence bogus groups may exist..

In a bogus group all loaned funds $2L$ are invested into a *single* project run by one of the members. The other borrower is a cosigner in the eyes of the lender (joint liability formally applies) but, because of limited liability, this co-signing borrower has no income or wealth for the lender to go after in case of declared default. As with the repayment decision, we assume that the borrowers form a bogus group if and only if this is jointly beneficial for them. A possible interpretation is the presence of social capital that allows within-group enforcement.

For any given loan terms (L, R) a bogus group invests all funds $2L$ into the higher-productivity (k_i) project. As in Section 3.2, it is not optimal to repay partially. Conditional on project success, the group's joint payoff from repaying is larger than the payoff from default if $k_i(2L) - 2R + 2V \geq k_i(2L)$ or, equivalently

$$R \leq V$$

which is the same no strategic default condition as (2) for standard groups.

The above implies that, for given loan terms (L, R) satisfying the no-default condition $R \leq V$, the expected payoff of a bogus group with project productivities k_i, k_j is:

$$W_{ij}^{B}(L, R) = 2pk_iL - 2pR + 2pV. \qquad \text{(BEP)}$$

The first term in (BEP) is the expected output in a bogus group, the second term is the expected repayment and the third term, $2pV$ is the expected continuation value of access to future credit. Note that the expected continuation value differs from its counterpart in a standard group, $2p(2-p)V$ in (SEP).

For $R \leq V$ the lender is repaid $2R$ with probability p and zero otherwise. Hence, the lender's break-even condition for lending $2L$ to a bogus group implies $L = pR$. Repayment is feasible upon success since $pk_i \geq 1$ by Assumption 1. Using the no-default condition $R \leq V$ and the fact that $W_{ij}^{B}(pR, R)$ is strictly increasing in R, the payoff-maximizing loan terms for a bogus group (as if known as such by the lender and taken in isolation) are therefore

$$L_B = pV \text{ and } R_B = V. \qquad (4)$$

3.3.1 Standard vs. bogus groups – comparison

There are five main differences between bogus and standard groups. First, for given loan terms (L, R), we have:

(i) *risk diversification*: in a standard group a borrower receives the continuation value V when her own project fails but her partner's project succeeds. The i.i.d. project returns assumption is important for this, as the joint liability clause ensures that in such scenario the group is not cut off from future credit. Thus, the expected continuation value per member is $p(2-p)V$ in a standard group vs. pV in a bogus group.

(ii) *expected repayment*: the borrowers in a standard group repay more in expectation, $p(2-p)R$ vs. pR in a bogus group since standard group members cover for their partners. For given loan terms (L, R) the difference between the increased expected continuation value in item (i) and the larger expected repayment (item ii) is non-negative for $R \leq V$ and zero for $R = V$ (see more on this below).

(iii) *expected output* is larger in a bogus group with heterogeneous borrowers ($2pk_hL$) than in a standard group with the same borrowers, $p(k_h + k_l)L$.

We characterize the interplay of items (i), (ii) and (iii) for given (L, R) in the Lemma below.

Lemma 1: *For given loan terms (L, R) satisfying the no-default condition $R \leq V$, borrowers with projects productivities k_i and k_j with $k_i \geq k_j$ prefer to form a bogus group if and only if*

$$p(k_i - \tfrac{k_i+k_j}{2})L > [1 - (1-p)^2 - p](V - R) \qquad (5)$$

and prefer to form a standard group otherwise.

Proof: see Appendix A

More generally, when the loan terms (L, R) can differ for standard and bogus groups:

(iv) *the interest rate* ($\frac{R}{L}$, gross) at which the lender breaks even is strictly lower in a standard group, $\frac{1}{p(2-p)}$ than in a bogus group, $\frac{1}{p}$. The reason, as in item (i), is that the lender is repaid with probability $1-(1-p)^2$ in a standard group in which two i.i.d. projects are funded vs. repaid with probability p in a bogus group.

(v) *the loan size* can be larger in standard groups due to the more likely repayment. Specifically, the maximum feasible repayment $R = V$ is the same but the higher probability of repayment allows standard groups to be offered larger loans (and lower interest rate) than bogus groups.

Lemma 1 implies that, all else equal, larger loan size L, or larger repayment amount R or larger productivity differential $k_i - k_j$ make bogus groups more preferred. The left hand side of (5) is the gain in expected output per member from forming a bogus group in comparison to a standard group, that is, the expected gain from investing $2L$ at the high return k_i instead of investing L at return k_i and L at return k_j. This corresponds to item (iii). Conversely, the right hand side of (5) is the net gain per member from forming standard group instead of bogus group thereby raising the probability of repaying and obtaining the continuation value V from p to $1-(1-p)^2$. This corresponds to items (i) and (ii) in the list.

The right hand side of (5) is always non-negative since $R \leq V$ by the no-default condition (51) and the left hand side is also non-negative since $k_i \geq k_j$ by assumption. For homogeneous (hh or ll) borrower pairs, the left hand side is zero – forming a bogus group does not offer any benefit in additional project return while it foregoes the diversification benefit of a standard group. Thus, for given (L, R), homogeneous pairs weakly prefer a standard group (strictly if $R < V$). For $R = V$, which is the maximum repayment amount satisfying the no-default condition (2), homogeneous pairs (with $k_i = k_j$) are indifferent between the two group forms. Hereafter we assume that, if indifferent, borrowers choose to form a standard group. This can be justified by adding a small exogenous cost (e.g., detection risk) of operating as bogus group.

To sum up, the diversification effect from investing in two projects favors standard groups due to the diversification effect from investing in two different projects, which is also the reason for the lower standard-group interest rate (item iv). In addition, larger loan size (item v) is feasible in standard groups. Conversely, items (ii) and (iii) favor bogus groups, by allowing the borrowers to benefit from investing a larger amount in the high-return project and from a smaller expected repayment.

3.3.2 Bogus groups and lender's loss

We now use Lemma 1 to show that a lender would lose money if (s)he ignored the possibility of bogus groups.

Proposition 2: *If all borrowers are offered loan terms $(L_S, R_S) = (p(2-p)V, V)$ then,*
(a) all heterogeneous (hl) borrower pairs form bogus groups
(b) all homogeneous (hh or ll) borrower pairs are indifferent between forming bogus or standard group
(c) bogus groups cause a loss to the lender.

Proof: see Appendix A.

The right hand side of condition (5) in Lemma 1 corresponds to the net benefit from forming a standard vs. bogus group for given loan terms (L, R). At the given loan terms (L_S, R_S) with $R_S = V$, this benefit is zero. Thus, only the increased output effect (item iii) remains and hence all heterogeneous pairs prefer to form bogus groups while homogeneous pairs are indifferent. All hl borrower pairs benefit from the larger expected output in a bogus group but repay less often than required for the lender to break even at the gross interest rate $\frac{R_S}{L_S}$. Hence, the lender loses money on any such loans. The implication is that endogenous bogus groups formation must be addressed by the lender by designing appropriate loan terms.

4 Endogenous group form

We now analyze the decision problem of lenders who design loan terms taking into account the borrowers' hidden action to form bogus or standard group. We first study the simpler setting in which the borrowers' project productivities k_i and k_j are observable by the lender. We interpret this as the lender knowing or monitoring the intended loan use: agriculture, retail, etc. In Section 4.2 we relax this assumption and study an alternative setting in which lenders do not observe the productivities k_i and k_j.

4.1 Observed productivity

Facing a group of type ij the non-profit lender would offer the loan terms that maximize the group payoff subject to breaking even. When the project returns k_i and k_j are observable, the lender rationally expects (knows) the optimal group form choice, standard or bogus, for any (L, R), and thus offers loan terms inducing the payoff-maximizing group form. The lender's break-even condition and no-strategic-default constraint depend on the induced group form. Below we suppress the subscripts ij for notational simplicity, e.g., we write $W^S(L, R)$ instead of $W^S_{ij}(L, R)$, etc. Given k_i, k_j and p the lender's problem can be written as:

Problem OP

$$\max_{L,R,\tau\in\{0,1\}} \tau W^S(L, R) + (1-\tau)W^B(L, R)$$

$$\text{s.t. } \tau W^S(L, R) + (1-\tau)W^B(L, R) \geq \tau W^B(L, R) + (1-\tau)W^S(L, R) \text{ (IC)}$$

$$R \leq V \text{ (no default)}$$

$$R = \tau\frac{L}{p(2-p)} + (1-\tau)\frac{L}{p} \text{ (break even)}$$

where the expected payoffs of a standard group, $W^S(L, R)$ and a bogus group, $W^B(L, R)$ are as defined in (SEP) and (BEP) above, omitting the subscript ij.

We use the binary variable $\tau \in \{0, 1\}$ to write in a compact way the choice of the larger of the two payoffs, $W^S(L, R)$ and $W^B(L, R)$, corresponding to the two possible group forms, standard or bogus. Constraint (IC) is an incentive constraint, requiring that if the lender chooses to offer loan terms (L, R) inducing a standard group ($\tau = 1$), then the borrowers must indeed prefer to form a standard group, i.e., $W^S(L, R) \geq W^B(L, R)$. The opposite, $W^B(L, R) > W^S(L, R)$ must hold if inducing a bogus group ($\tau = 0$) is payoff-maximizing. The no default and break even constraints were derived in Section 3. Proposition 3 characterizes the payoff-maximizing loan terms and group form depending on the parameters k_l, k_h and p.

Proposition 3: *Suppose the project productivities k_i and k_j are observable. The loan terms* *solving* ***Problem OP*** *for any group type $ij \in \{hh, ll, hl\}$ and any p, k_l, k_h satisfying Assumption 1 are:*

group type, $ij =$	loan terms	group form	for
1. homogeneous, ll *or* hh $k_i = k_j \in \{k_l, k_h\}$	(L_S, R_S)	*standard*	*any* k_l, k_h, p
2. heterogeneous, hl	(L_E, R_E)	*standard*	$k_h \in [k_l, \bar{k}]$
$k_i = k_h, k_j = k_l$	(L_B, R_B)	*bogus*	$k_h > \bar{k}$

where $L_E \equiv \frac{p(2-p)V}{1+\frac{p(2-p)}{2(1-p)}(k_h - k_l)}$, $R_E = \frac{L_E}{p(2-p)}$ *and* $\bar{k}$ *is the larger root of* $W^S_{hl}(L_E, R_E) = W^B_{hl}(L_B, R_B)$ *written as quadratic equation in* k_h.[18]

Proof: see Appendix A.

Intuitively, homogeneous borrower pairs (ll or hh) always prefer a standard group, since they benefit from the reduced risk of default (diversification) and the associated lower interest rate and for them there is no increase in expected output from forming a bogus group. Hence, homogeneous pairs always form standard groups and are offered the maximum feasible (subject to no default) loan size L_S.

In contrast, heterogeneous borrower pairs (hl) face a trade-off between the larger expected output that can be obtained in a bogus group vs. the larger expected payoff from risk diversification in a standard group. Therefore, when the productivity differential k_h vs. k_l is sufficiently large, namely for $k_h > \bar{k}$ which is equivalent to $W^B(L_B, R_B) > W^S(L_E, R_E)$, it is optimal to induce a bogus group and offer the largest feasible loan size L_B. Note that, since $\bar{k} > k_l$ as shown in the proof of Proposition 3, the condition $k_h > \bar{k}$ is tighter than the condition $k_h > k_l$ for forming a bogus group in Proposition 2 (with loan L_S, R_S) – here the lender optimally responds to the possibility of bogus groups by adjusting the loan size to L_B and the interest rate to $1/p$. Conversely, when the project productivity differential is relatively small, for $k_h \in [k_l, \bar{k}]$ which is equivalent to $W^S(L_E, R_E) \geq W^B(L_B, R_B)$, the risk diversification benefit from inducing a standard group outweighs the loss of larger expected output. However, in that case a reduced loan size, $L_E < L_S$ must be offered, to deter the heterogeneous borrowers from switching to bogus group. Note also that $L_E \to L_S$ when the productivity differential shrinks to zero, $k_h \to k_l$.

Comparing the three possible loan sizes in Proposition 3, it is easy to verify that $L_S > L_B$ and $L_S > L_E$. That is, our model predicts that homogeneous groups (hh and ll) receive larger loans than heterogeneous groups (hl). The reason is the lower probability of default in the standard groups formed by homogeneous borrowers. Comparing standard vs. bogus groups, we have $L_S > L_B$, however, either of L_B and L_E can be larger, depending on the parameter values. Thus, from observing group form alone, without knowing the within-group project quality composition, we cannot draw a definite conclusion about the relationship between loan size and group form. The interest rate, R/L is always lower in standard groups ($\frac{1}{p(2-p)}$) compared to in bogus groups ($\frac{1}{p}$) and weakly lower for homogeneous compared to heterogeneous groups.

The above results show that there are gains in efficiency and lender losses are avoided (see Proposition 2) when the endogenous group form (bogus or standard) is taken into account by the lender and the loan terms

[18] The threshold $\bar{k}$ depends on the values of k_l and p – see the proposition proof for details.

are appropriately designed. The takeaway is that Lei Da Hu should not be viewed by lenders as money-losing illicit practice but instead as an opportunity that MFIs can incorporate in their lending strategy.

4.2 Unobserved productivity

4.2.1 Setup

Suppose now that the borrowers' project productivities $k_i, k_j \in \{k_l, k_h\}$ are unobserved by the lender. The lender knows the values k_l and k_h, that is, the return of a high- and a low-productivity project. Assume also that the lender knows the fraction of agents, q_{ij} belonging to hh, ll and hl borrower pairs, with $q_{hh} + q_{ll} + q_{hl} = 1$.[19] As we show below, our main results and intuition from Section 4.1, in which productivity was assumed observable, generalize with some small modifications.

Note first that the loan terms from Proposition 3 are no longer feasible, since if the lender does not observe the group members' productivities k_i and k_j then the lender cannot offer terms contingent on the group composition ij. Second, when k_i, k_j are unobserved, the borrowers have incentive to report the ij value which maximizes their joint payoff, which may differ from the true group composition. For example, the borrowers in an hl pair that would be offered loan terms (L_E, R_E) in Proposition 3 can misreport their type as hh, receive loan (L_S, R_S) and form a bogus group, which would cause a loss to the lender.

Unlike in Proposition 3 where the lender customizes the loan terms for each observed borrower pair (k_i, k_j), when the k's are unobservable the lender must design the loan terms so that no borrower pair has incentive to mis-report its type (ll, hh or hl) and also, after taking a loan, no borrower pair has incentive to choose group form (bogus or standard) different from the form intended by the lender. This is a mechanism design problem which is more complex than the standard screening or adverse selection problem, since the lender faces both *unobserved types* (the k_i, k_j) and an *unobserved action* (the ex-post moral hazard in group form choice).

4.2.2 Loan terms

Since the borrower productivities are unobserved by the lender all borrowers are treated ex-ante equally. That is, the lender offers a loan menu to all and the borrowers self-select. Depending on the model parameters and the fractions q_{ij} of different borrower pairs, the non-profit / NGO lender decides the loan terms and what group form to induce, subject to no-default and incentive compatibility constraints.

The lender makes zero profits (breaks even) on each offered loan (see Section 4.4 where we relax this assumption). Breaking even implies that the gross interest rate $\frac{R}{L}$ must equal either $\frac{1}{p(2-p)}$ (if a standard group is being induced) or $\frac{1}{p}$ (if a bogus group is being induced). Any lower interest rate would cause a loss to the lender while any higher interest rate can be lowered, by raising the loan size, to achieve higher borrower payoff.

Remember from Section 3 that for any loan terms, either $(L, \frac{L}{p})$ or $(L, \frac{L}{p(2-p)})$, the group payoff is strictly increasing in the loan size L, regardless of the borrowers' productivities. This implies that, within each of the two possible loan types defined by the interest rate, $(L, \frac{L}{p})$ or $(L, \frac{L}{p(2-p)})$, only a unique loan size can be

[19] For example, if the project productivities are i.i.d. and k_l and k_h occur with equal probability after the group is formed, then the fraction of hh and ll groups would be $1/4$ each, while the fraction of hl groups would be $1/2$.

offered, namely the size that maximizes the borrowers' joint payoff subject to the no-default and break-even constraints. If two distinct loan sizes were offered both carrying the same interest rate, then borrowers would be better off with the larger loan.

The above observations imply that the lender offers a loan menu consisting of at most two different loans, $\mathcal{N} \equiv (L_N, R_N)$ and $\mathcal{M} \equiv (L_M, R_M)$, inducing respectively a standard group and a bogus group, and using which the lender screens the unobserved group composition. The loan terms $\mathcal{N}$ and $\mathcal{M}$ solve Problem UP stated below and are such that:

(a) each borrower pair $ij \in \{ll, hh, hl\}$ which selects loan $\mathcal{N}$ chooses to be standard and each pair ij which selects loan $\mathcal{M}$ chooses to be bogus;

(b) no borrower pair defaults strategically;

(c) the lender breaks even for each offered loan, $\mathcal{N}$ and $\mathcal{M}$;

(d) the total borrowers' payoff is maximized.

Problem UP

$$\max_{L_N, R_N, L_M, R_M, \tau_{ij} \in \{0,1\}} \sum_{ij} q_{ij}[\tau_{ij} W_{ij}^S(L_N, R_N) + (1-\tau_{ij}) W_{ij}^B(L_M, R_M)]$$

subject to

$$R_M \leq V \tag{6}$$

$$R_M = \frac{L_M}{p} \tag{7}$$

$$R_N \leq V \tag{8}$$

$$R_N = \frac{L_N}{p(2-p)} \tag{9}$$

$$\tau_{ij} W_{ij}^S(L_N, R_N) + (1-\tau_{ij}) W_{ij}^B(L_M, R_M) \geq \max\{W_{ij}^B(L_N, R_N), W_{ij}^S(L_M, R_M)\}, \ \forall ij \in \{hh, hl, ll\} \tag{IC2}$$

where, as in Section 4.1, $\tau_{ij} \in \{0, 1\}$ is a binary variable indicating the group form (standard or bogus) that the lender wishes to induce for borrower pair ij and where the group payoffs $W_{ij}^S(L, R)$ and $W_{ij}^B(L, R)$ are defined in (SEP) and (BEP).

The first four constraints are the no-default and break-even constraints for each loan. Constraints (IC2) ensure that any pair ij would choose its intended contract ($\mathcal{N}$ or $\mathcal{M}$) and group form (standard or bogus) which maximize its payoff. Selecting an alternative contract or deviating to the alternative group form, or both, is not optimal.

> **Proposition 4:** *Suppose the project productivities k_i and k_j are unobserved by the lender. Let $\mathcal{E} \equiv (L_E, R_E)$, $\mathcal{B} \equiv (L_B, R_B)$ and $\bar{k}$ be as defined in Proposition 3. Define also $L_F \equiv \frac{pk_h - 1}{pk_h - \frac{1}{2-p}} pV$, $R_F = \frac{L_F}{p(2-p)}$ and $\mathcal{F} \equiv (L_F, R_F)$. Then, for any k_l, k_h and p satisfying Assumption 1, the loan menu offered by the lender and the borrowers' chosen loan and group form solving **Problem UP** are:*[20]

[20] As stated in Section 3, these results assume a single lender which makes zero profits on each offered loan (no cross-subsidization across borrowers). Cross-subsidization could alternatively be ruled out by assuming free entry by lenders, however, in that case an equilibrium may not exist as in Rothschild and Stigitz (1976) (details available upon request). We allow for cross-subsidization in

	loan menu	chosen loan terms and group form	
parameter condition	$\mathcal{N}, \mathcal{M} =$	*ll and hh pairs*	*hl pairs*
1. $k_h \in [k_l, \bar{k}]$	$\mathcal{E}$, ($\mathcal{B}$)	$\mathcal{E}$, standard	$\mathcal{E}$, standard
2. $k_h > \bar{k}$	$\mathcal{F}$, $\mathcal{B}$	$\mathcal{F}$, standard	$\mathcal{B}$, bogus

Proof: See Appendix A.

As in Proposition 3, bogus groups are formed by the heterogeneous borrower pairs, when the heterogeneity in project productivity is relatively high ($k_h > \bar{k}$), while all borrowers form standard groups when the project heterogeneity is relatively low ($k_h \in [k_l, \bar{k}]$). Also as in Proposition 3, the loan size L_N optimally chosen to induce a standard group is reduced from the maximum feasible size L_S to either L_E or L_F. In the case $k_h \in [k_l, \bar{k}]$ the reduced loan size L_E ensures that heterogeneous pairs do not take loan $\mathcal{N}$ but form a bogus group instead of standard group. In the case $k_h > \bar{k}$ the reduced loan size L_F prevents heterogeneous (bogus) groups from taking loan $\mathcal{N}$ instead of $\mathcal{M}$. In contrast, loan $\mathcal{M}$ inducing a bogus group has the maximum feasible loan size L_B.[21] The basic intuition and trade-off between the diversification benefit in standard groups vs. the expected output gain in bogus group apply as before.

The asymmetric information about the project productivities does, however, reduce efficiency compared to the observable productivities setting in Section 4.1. Comparing with Proposition 3, in Proposition 4 the reduced loan size for standard groups (L_E or L_F, both smaller than L_S) applies for all borrowers and all k_l, k_h and p (even though constraint (IC2) is slack for homogeneous pairs, see the proposition proof), since the loan terms cannot depend on the unobserved group composition. Also, for any given parameter values, the offered loan sizes are weakly smaller than their counterparts in the observable productivity setting.

Comparing the loan terms, Proposition 4 implies a gross interest rate R/L that is weakly lower for homogeneous groups vs. heterogeneous groups and a strictly lower interest rate for standard vs. bogus groups. Without additional information there is no clear prediction about the loan size, since the relative magnitudes of L_B, L_E and L_F depend on the model parameters. The required repayment amount is strictly larger for bogus groups than for standard groups (since $R_B = R_S = V$ and $R_S > R_E$, $R_S > R_F$).

4.3 Welfare analysis

To further clarify the sources of inefficiency because of asymmetric information, we compare the loan terms derived in Sections 4.1 and 4.2 with the *full information* benchmark, when project productivities are observable and group form is observable and contractible (but there is still limited enforcement of repayments).

Lemma 2. Full information

With observable project productivities and contractible group form:

(i) *homogeneous groups (ll* or *hh) are offered loan terms (L_S, R_S) and are standard;*

Section 4.4.

[21] In the case $k_h \in [k_l, \bar{k}]$ in Proposition 4 bogus groups are not induced at optimum and hence setting $\mathcal{M} = \mathcal{B}$ is without loss of generality (see the proposition proof). Alternatively, the lender could just offer contract $\mathcal{E}$ to all borrowers.

(ii) *heterogeneous groups* (hl) *are offered loan terms* (L_B, R_B) *and are bogus if* $k_h > \frac{2-p}{p}k_l$ *or are offered* (L_S, R_S) *and are standard if* $k_h \in (k_l, \frac{2-p}{p}k_l]$.

Proof: see Appendix A

It is easy to show that $\frac{2-p}{p}k_l > \bar{k}$ where $\bar{k}$ was defined in Proposition 3. This implies that in Proposition 3 (observable k_i) a welfare loss is present in two cases, both involving the heterogeneous groups hl:

1. if $k_h \in (k_l, \bar{k})$, the hl groups would be offered (L_S, R_S) with full information, but are instead offered a reduced loan size (L_E, R_E)

2. if $k_h \in [\bar{k}, \frac{2-p}{p}k_l)$, the hl groups would be offered (L_S, R_S) and be standard with full information, but are instead offered a reduced loan size (L_B, R_B) (since $L_B < L_S$), and an inefficient (bogus) group form, with higher default risk relative to under full information. The inefficiency arises because of the unobserved action by the borrowers (moral hazard), that is, their ability to choose the group form after accepting the loan. Finally, in the case $k_h > \frac{2-p}{p}k_l$ there is no welfare loss.

In the setting with unobservable productivities (Proposition 4) there can be an additional welfare loss, stemming from the asymmetric information about the group composition. Homogeneous groups incur a welfare loss by receiving reduced loan size compared to the full information case ($L_E < L_S$ and $L_F < L_S$). Heterogeneous groups are offered loan terms different from the full-information terms (and hence inefficient) if $k_h \in (k_l, \frac{2-p}{p}k_l)$ but are offered the efficient loan terms (L_B, R_B) otherwise.

What if the lender could costlessly prevent the formation of bogus groups? Our results show that this could yield an efficiency gain if the productivity heterogeneity is not too large (for $k_h \leq \frac{2-p}{p}k_l$), since in that case the efficient loan terms (L_S, R_S) could be offered to all borrowers. However, if the productivity heterogeneity is large (for $k_h > \frac{2-p}{p}k_l$), then preventing bogus groups would be inefficient since a larger total payoff can be achieved by accommodating them with appropriate loan terms.

4.4 Cross-subsidizing lender

In this section we analyze an alternative setting, with a single non-profit / NGO lender that can cross-subsidize across borrowers subject to an overall zero expected profits constraint. We only focus on the unobservable productivities case (unobservable group composition ll, hh or hl) in which all borrowers are treated ex-ante equally, that is, the lender offers the same loan menu to all borrowers and they self-select a loan and group form.[22] As in Section 4.2, suppose the lender offers the menu $\mathcal{N}, \mathcal{M}$ where loan $\mathcal{N}$ induces a standard group and loan $\mathcal{M}$ induces a bogus group.

In Section 4.2 (see Lemma A0 in the proof of Proposition 4) we proved that it is always optimal to induce a standard group for homogeneous pairs ($\tau_{ll} = \tau_{hh} = 1$), i.e., $W^S_{ii}(L, R) \geq W^B_{ii}(L, R)$ for any feasible L, R and $ii \in \{ll, hh\}$. If the parameters are such that it were also optimal to induce standard group for heterogeneous hl pairs ($\tau_{hl} = 1$), then there would be no cross-subsidization – all groups use loan $\mathcal{N}$, breaking even requires $L_N = p(2-p)R_N$ and hence this case was already analyzed in Proposition 4 (the case of loan $\mathcal{E}$).

[22]We find this setting more compelling than the setting in which the lender observes the project productivities and can freely reward or 'tax' the different borrower groups (hh, ll or hl).

Consequently, in this section we focus on the case in which the lender finds it optimal to induce heterogeneous groups to be bogus (that is, $\tau_{hl} = 0$ is optimal, in addition to $\tau_{ll} = \tau_{hh} = 1$) and write the lender's problem as:

Problem UP′

$$\max_{L_N, R_N, L_M, R_M} q_{ll} W_{ll}^S(L_N, R_N) + q_{hh} W_{hh}^S(L_N, R_N) + q_{hl} W_{hl}^B(L_M, R_M)$$

subject to

$$R_M \leq V \text{ and } R_N \leq V \quad \text{(no default)}$$
$$(1 - q_{hl})(p(2-p)R_N - L_N) + q_{hl}(pR_M - L_M) = 0 \quad \text{(break even)}$$
$$W_{hl}^B(L_M, R_M) \geq W_{hl}^S(L_N, R_N) \quad (\tau_{hl} = 0 \text{ is optimal})$$
$$W_{hl}^B(L_M, R_M) \geq \max\{W_{hl}^B(L_N, R_N), W_{hl}^S(L_M, R_M)\}$$
$$W_{ii}^S(L_N, R_N) \geq W_{ii}^S(L_M, R_M) \quad \text{for } ii = \{ll, hh\}$$

Comparing Problem UP′ with Problem UP in Section 4.2, note that the break-even constraint now allows cross-subsidization across loans (L_N, R_N) and (L_M, R_M). The last two constraints are the incentive / self-selection constraints simplified from (IC2) after the observations made above about τ_{ij}, W_{ii}^S and W_{ii}^B.

Lemma 3. *The loan terms* $\mathcal{M} = (L_M, R_M)$ and $\mathcal{N} = (L_N, R_N)$ *solving Problem* UP′ *are:*

$$L_N = L_M = \tilde{L} \equiv ((2-p)(1-q_{hl}) + q_{hl})pV \text{ and } R_N = R_M = \tilde{R} \equiv V$$

Proof: see Appendix A.

Lemma 3 shows that, for model parameters for which $\tau_{hl} = 1$ is optimal,[23] a cross-subsidizing lender maximizes total borrower welfare by offering the same loan terms $(\tilde{L}, \tilde{R})$ to all borrowers, that is, setting $\mathcal{M} = \mathcal{N} = (\tilde{L}, \tilde{R})$ solves Problem UP′. Note that if all pairs are heterogeneous, for $q_{hl} \to 1$, the loan size $\tilde{L}$ converges from above to the maximum feasible bogus group loan size L_B, while if all pairs are homogeneous, for $q_{hl} \to 0$, the loan size $\tilde{L}$ converges from below to the maximum feasible standard group size L_S. Since $p\tilde{R} < \tilde{L}$ the lender loses money from each heterogeneous (bogus) group (loses money on 'loan $\mathcal{M}$'), however, this loss is cross-subsidized by profits from the homogeneous (standard) groups (makes a profit on 'loan $\mathcal{N}$'). This is in contrast to Sections 4.1 and 4.2 where we assumed that the lender breaks even on each offered loan.

Discussion

In Proposition 4 respecting the incentive constraints (IC2) can be costly in terms of welfare if the fraction of heterogeneous groups q_{hl} is small. The reason is that even though the incentive constraint (IC2) is slack for homogeneous pairs, the loan intended for standard groups must have a reduced size (L_E or L_F instead of L_S), to discourage heterogeneous pairs from taking it but instead forming bogus groups. Clearly, in the limit as $q_{hl} \to 0$, the payoff loss from the few heterogeneous pairs using loan $\mathcal{E}$ or $\mathcal{F}$ and forming a bogus

[23] See the discussion at the end of this section.

group would be negligible compared to the payoff gain if all homogeneous pairs were able to receive larger loans. Therefore, if the lender could run a loss on some borrowers (hl) and compensate it by profits from other borrowers (hh and ll), as allowed in Lemma 3, a larger total payoff can be possible.

Compare the payoff maximizing cross-subsidizing loan $(\tilde{L}, \tilde{R})$ derived in Lemma 3 with the loan terms in Proposition 4. Consider first the case of relatively high productivity heterogeneity, $k_h > \bar{k}$ in Proposition 4 in which the homogeneous borrower pairs ll and hh select loan $\mathcal{F}$ and form standard groups while the heterogeneous pairs hl select loan $\mathcal{B}$ and form bogus groups. Note that these are the same group forms ($\tau_{ll} = \tau_{hh} = 1$ and $\tau_{hl} = 0$) as the forms chosen by the cross-subsidizing lender. Clearly loans $\mathcal{F}$ and $\mathcal{B}$ are feasible in Problem UP′ – that is, a cross-subsidizing lender could have chosen the loan menu $(\mathcal{F}, \mathcal{B})$ but instead the lender optimally chose different loan terms $(\tilde{L}, \tilde{R})$ for all borrowers, as proved in Lemma 3. This implies that for $k_h > \bar{k}$, if the lender could cross-subsidize, total borrower payoff would be larger than in Proposition 4.

Now consider the case of relatively low productivity heterogeneity, $k_h \in [k_l, \bar{k}]$ in which all borrowers optimally use loan $\mathcal{E}$ in Proposition 4 and form standard groups. Using (SEP) evaluated at (L_E, R_E), the borrowers' total payoff equals

$$W^{total}(L_E) = p(2q_{ll}k_l + 2q_{hh}k_h + q_{hl}(k_h + k_l) - 2)L_E + 2p(2-p)V$$

In comparison, in Lemma 3 all borrowers receive loan terms $(\tilde{L}, \tilde{R}) = (\tilde{L}, V)$, homogeneous groups are standard and heterogeneous groups are bogus, yielding a total payoff of

$$\tilde{W}^{total} = p(2q_{ll}k_l + 2q_{hh}k_h + 2q_{hl}k_h)\tilde{L}$$

Depending on the parameter values, we may have $\tilde{W}^{total} > W^{total}(L_E)$, that is cross-subsidization yields larger total payoff, or vice versa.[24] For example, for q_{hl} sufficiently close to zero and $k_h > k_l$, we have $\tilde{W}^{total} > W^{total}(L_E)$, since $\tilde{L} \to L_S$ and $\tilde{W}^{total} \to W^{total}(L_S) > W^{total}(L_E)$. Alternatively, for $k_h = k_l$ and $q_{hl} > 0$, we have $L_E = L_S > \tilde{L}$ and so $W^{total}(L_E) > \tilde{W}^{total}$ (by continuity the latter also holds for k_h close to k_l).

Note that if the lender could cross-subsidize and loan $(\tilde{L}, \tilde{R})$ were payoff-maximizing (e.g., the case $k_h > \bar{k}$), then all borrowers receive the same loan terms but some borrower groups (heterogeneous pairs) are bogus while other groups (homogeneous pairs) are standard. The lender is losing money on the bogus groups but this is compensated by the standard groups. This outcome, coexisting bogus and standard groups and common loan terms, is consistent with what the broad patterns in our Chinese data although, as mentioned in Section 3.1, there is no evidence that cross-subsidization was possible in CFPAM loans. We thus prefer to interpret this analysis as purely theoretical.

Finally, note that the cross-subsidization setting analyzed here relies strongly on the assumed lender exclusivity and cannot be supported with entry, since then another lender could offer loan $(\tilde{L} - \delta, \tilde{R} - \varepsilon)$ with $\varepsilon < \frac{q_{hl}(1-p)k_h pV}{(2-p)(pk_h - 1)}$ and $\delta \in (\frac{\varepsilon}{k_h}, \frac{(2-p)\varepsilon}{k_h})$ and attract all hh and ll groups while earning positive profit (the proof of this claim is available upon request).

[24] The exact parametric conditions are easy to derive but not very informative.

5 Extensions

5.1 Endogenous continuation value V

We show how our model can be extended to an infinite horizon setting endogenizing the continuation value of access to future credit, V. Suppose for simplicity at first that the borrower project types are fixed over time (we will relax this below) and future payoffs are discounted with factor $\delta \in (0,1)$. Since the loan size would be indeterminate with linear technology, assume that the project output, given loan size L and productivity k_i, is

$$Y_i = \begin{cases} k_i L^{\alpha} & \text{with probability } p, \\ 0 & \text{with probability } 1-p. \end{cases}$$

where $\alpha \in (0,1)$. Note that the concavity assumption reduces the gains from forming bogus groups since there are diminishing marginal returns from pooling the loans. We only analyze the observable productivity case and show results that parallel Propositions 1 and 3.

5.1.1 Standard groups

Consider a standard group with $k_i \geq k_j$. Following the analysis in Section 3, the per-borrower continuation value V_{ij} solves

$$2V_{ij} = v_{ij} + 2\delta p\,(2-p)\,V_{ij} \tag{10}$$

where v_{ij} is the group's current-period expected payoff

$$v_{ij} = p\,(k_i + k_j)\,L^{\alpha} - 2p\,(2-p)\,R \tag{11}$$

Using the lender's break-even condition $L = p\,(2-p)\,R$ we derive the continuation value V_{ij} as function of loan size:

$$V_{ij} = \frac{p\frac{k_i+k_j}{2}L^{\alpha} - L}{1-\delta p(2-p)} \tag{12}$$

The no strategic default constraint is $\delta V_{ij} \geq R$ which, using (12), is equivalent to:

$$L \leq \left(\delta p^2\,(2-p)\,\frac{k_i+k_j}{2}\right)^{\frac{1}{1-\alpha}} \tag{13}$$

The lender would choose L that maximizes the group payoff $2V_{ij}$ subject to no default and breaking-even. Because of the assumed concavity of project output, there are two possibilities: either the no-default constraint (13) binds, which determines L and R (as in Sections 3 and 4) or there is an interior value for L that maximizes V_{ij} in (12). To keep the analysis as close to the baseline model as possible, assume that

$$\text{E1:}\;\; \delta p\,(2-p) \leq \alpha \tag{14}$$

which ensures that the no-default constraint binds.[25]

[25] For example, E1 is always satisfied for $\alpha \geq \delta$.

Therefore, in parallel with Proposition 1, the payoff-maximizing standard group loan is $C_s \equiv (L_s, R_s)$ with

$$L_s \equiv \left(\delta p^2(2-p)\frac{k_i+k_j}{2}\right)^{\frac{1}{1-\alpha}} \text{ and } R_s \equiv \frac{L_s}{p(2-p)}$$

where we also assume that the repayment feasibility condition $k_j L^\alpha \geq 2R$ is satisfied at C_s.[26] Unlike in Proposition 1, since V_{ij} is endogenous, the loan terms C_s depend on the group's composition (its average productivity).

5.1.2 Bogus groups

We derive the payoff-maximizing loan for a bogus group in a similar way. Calling the continuation value U_{ij} and the current value u_{ij},

$$2U_{ij} = u_{ij} + 2\delta p U_{ij}$$

where

$$u_{ij} = pk_i\,(2L)^\alpha - 2pR.$$

Using the lender's break-even condition $L = pR$, the endogenous continuation value is

$$U_{ij} = \frac{2^{\alpha-1}pk_iL^\alpha - L}{1-\delta p}$$

The no-default constraint is

$$\delta U_{ij} \geq R \Longleftrightarrow L \leq \frac{1}{2}\left(\delta p^2 k_i\right)^{\frac{1}{1-\alpha}} \tag{15}$$

It is easy to verify that constraint (15) must bind. The payoff-maximizing bogus group loan is thus $C_b \equiv (L_b, R_b)$ where

$$L_b \equiv \frac{1}{2}\left(\delta p^2 k_i\right)^{\frac{1}{1-\alpha}} \text{ and } R_b \equiv \frac{L_b}{p}$$

Repayment feasibility, $k_i(2L_b)^\alpha \geq 2R_b$ is equivalent to $\delta p \leq 1$ which holds since $\delta, p \in (0,1)$. Note that, as in our baseline model, for homogeneous groups the standard-group loan is strictly larger: $L_s > L_b$.

5.1.3 Loan terms

We first argue that it is never optimal for the lender to induce a homogeneous group ii to be bogus. Indeed, the maximum feasible expected payoff from a standard group is $W_{ii}^S(L_s, R_s) = 2pk_iL_s^\alpha$ while the maximum feasible payoff from the same group being bogus is $W_{ii}^B(L_b, R_b) = pk_i(2L_b)^\alpha$. It is always the case that $W_{ii}^S(L_s, R_s) > W_{ii}^B(L_b, R_b)$ since $\alpha < 1$ and since $L_s > L_b$ for homogeneous groups.

Therefore, inducing a bogus groups could be optimal only for heterogeneous (hl) groups and only if the gain from extra output is sufficiently large.[27] Using these results, it is possible to derive the loan terms as

[26] This requires the parametric restrictions $k_i \leq (\frac{1}{\delta p} - 1)k_j$ and $2p\delta < 1$ in addition to E1.

[27] A sufficient condition is $W_{hl}^B(L_b, R_b) > W_{hl}^S(L_s, R_s)$ which, by direct calculation, is equivalent to

$$k_h > \frac{(\frac{2-p}{2})^\alpha}{1-(\frac{2-p}{2})^\alpha}k_l$$

in Proposition 3, although there is no closed-form solution in some cases because of the power function.[28] Our main results thus remain robust to endogenizing the continuation value of access to future credit – homogeneous groups are always standard and offered the efficient loan size while heterogeneous groups may be bogus or standard with possibly reduced loan size.

5.1.4 Switching project types

The above analysis could be further extended for borrower project types that switch over time. For example, suppose there is equal mass of each type and the switching probability from any productivity type $k_t \in \{k_l, k_h\}$ to any $k_{t+1} \in \{k_l, k_h\}$ is $1/2$. Focus, for example, on the case with loan C_s for hh and ll groups and loan C_b for hl groups. The (ex-ante) continuation value V per borrower would then solve

$$2V = \frac{1}{4}v_{ll} + \frac{1}{4}v_{hh} + \frac{1}{2}u_{hl} + 2\delta V$$

where v_{ll}, v_{hh} and u_{hl} are defined as above.

5.2 Default penalties

Our baseline model is isomorphic to a one-period model with default penalties (see also Dubey et al., 2005 or Besley and Coate, 1995). Suppose there is an exogenous penalty f imposed on each group member upon default, either strategic or involuntary. There is no continuation value and the default penalty is the only way to prevent reneging. A borrower would then repay as long as $R \leq f$.

A standard group repays with probability $p(2-p)$ ($= (1-(1-p)^2$) and its expected payoff is:

$$\begin{aligned} W^S(L,R) &= p(k_i+k_j)L - 2p(2-p)R - 2(1-p(2-p))f = \\ &= p(k_i+k_j)L - 2p(2-p)R + 2p(2-p)f - 2f \end{aligned}$$

which, setting $f = V$, equals the payoff $W^S(L,R)$ from our baseline model shifted by the constant $-2f$.

Similarly, the expected payoff of a bogus group is

$$W^B(L,R) = 2pk_iL - 2pR + 2pf - 2f.$$

Assuming that the productivities k_i, k_j are large enough so that the above payoffs are non-negative (modifying Assumption 1), the previous analysis goes through, since all comparisons between standard and bogus group payoffs remain unchanged. One advantage of the version of the model with default penalties, is that it avoids the possible issue with the endogeneity of the continuation value V.

5.3 Endogenous sorting

An important issue that we have not addressed yet is endogenous sorting when borrower groups are formed. We modify the model timing so that the project productivities are drawn first and then, knowing these produc-

[28] To induce a heterogeneous group to be standard, the following condition must hold for some $L \leq L_s$, $W^S_{hl}(L, \frac{L}{p(2-p)}) \geq W^B_{hl}(L, \frac{L}{p(2-p)})$.

tivities, the borrowers sort into groups rationally expecting the loan terms which will be offered. Consistent with our baseline model, we focus only on sorting based on the borrowers' project types (productivities). In reality additional factors, such as borrowers' initial wealth or outside options, may also matter for their choice of what group to join.

For simplicity assume that there is equal mass of borrowers with projects of each type. In that case, it is easy to show that, if only standard groups could be formed (Section 3.2), any equilibrium matching pattern is optimal since two hl groups achieve the same expected surplus, $2p(k_l + k_h)L_S$ as one hh and one ll group.

With endogenous unobserved group form and observable productivities k_i, k_j (Section 4.1), the lenders would offer the loan terms described in Proposition 3 to whichever groups they face after the matching stage. Hence, we simply need to check whether positive (PAM) or negative (NAM) assortative matching is ex-ante optimal. If the parameters are such that $k_h \in (k_l, \bar{k}]$ (the case with $\mathcal{S}$ and $\mathcal{E}$ loans in Proposition 3) then it is easy to verify that

$$2W_{hl}^S(L_E, R_E) = 2p(k_h + k_l)L_E < 2p(k_h + k_l)L_S = W_{hh}^S(L_S, R_S) + W_{ll}^S(L_S, R_S)$$

and so PAM sorting is optimal (all groups are homogeneous and standard).

In contrast, if the productivity differential is larger, that is $k_h > \bar{k}$ (the case with $\mathcal{S}$ and $\mathcal{B}$ loans in Proposition 3) then

$$2W_{hl}^B(L_B, R_B) > W_{hh}^S(L_S, R_S) + W_{ll}^S(L_S, R_S)$$

is equivalent to $k_h > \frac{2-p}{p}k_l$. Hence, NAM sorting is optimal (all groups are heterogeneous and bogus) if $k_h > \frac{2-p}{p}k_l$ while PAM sorting is optimal (all groups are homogeneous and standard) for $k_h \in (\bar{k}, \frac{2-p}{p}k_l]$.

Allowing for endogenous matching is more complex if the lenders cannot observe the project productivities (Section 4.2). The reason is that the loan menus in Proposition 4 cannot be taken as given by the borrowers, since a lender may wish to offer a different menu if he knew, for example, that all groups that he would face would be hl (the IC constraint is affected). On the other hand, the equilibrium group composition and matching pattern would depend on the loan menu that the borrowers expect. Thus is a hard problem, potentially with multiple equilibria, that we leave for future research.

5.4 Standardized loan

Suppose that for exogenous reasons (e.g., complexity, given the personnel MFIs rely on; the microcredit clientele, the political environment, or others) lenders were restricted to offer the same standardized loan terms to all borrowers. What would these terms be?

As in Section 4.4 (see the proof of Lemma 3, after proving that $L_M = L_N$ and $R_M = R_N$), it is easy to show that the borrower payoff maximizing standardized loan terms are

$$\tilde{L} = [(1 - q_{hl})(2 - p) + q_{hl}]pV \text{ and } \tilde{R} = V.$$

At $(\tilde{L}, \tilde{R})$ the lender loses money from each (bogus) heterogeneous group since $p\tilde{R} < \tilde{L}$, but this loss is cross-subsidized by profits from the (standard) homogeneous groups. Our discussion from the second half of Section 4.4 applies.

5.5 Individual loans

So far, motivated by the CFPAM data, we only considered joint-liability group loans and the endogenous choice of borrowers to form bogus or standard groups. To further clarify our contribution to the literature we now explore allowing individual loans in our model. We show that when payoffs are transferable within the group then no borrower pair has incentive to switch to individual loans.

Suppose a borrower with productivity $k_i \in \{k_l, k_h\}$ is given an individual loan with size L and required repayment R. The borrower's payoff upon strategic default is $k_i L$, the payoff upon repayment is $k_i L - R + V$. Hence, the no strategic default constraint is $R \leq V$ which is the same as that for group loans derived in Section 3.

The lender's break-even condition is $L = pR$, the same as that in a bogus group. The borrower's expected payoff, given $R \leq V$, is

$$W^I(L, R) = pk_i L - pR + pV$$

These results imply that the borrower payoff-maximizing individual loan has terms

$$L_I = pV \text{ and } R_I = V,$$

which are identical to the bogus group terms (L_B, R_B) in Section 3. Feasibility of repayment requires $pk_i \geq 1$ which is satisfied by Assumption 1. Consistent with the previous analysis, assume transferable utility within the group, i.e., the borrowers maximize their joint expected payoff.

> **Proposition 5**: *Suppose, starting from the loan terms in Proposition 3, borrowers are offered individual loans (L_I, R_I) defined above. Then, for any group composition (k_i, k_j), the borrowers have no incentive to switch to individual loans.*

Proof: see Appendix A.

Intuitively, group lending dominates individual lending in our model for two reasons. First, in a standard group, group loans reduce the risk of default and allow larger loan size compared to individual loans. Second, in a bogus group, group lending enables the borrowers to benefit from investing a larger amount into the higher return project.

6 Conclusions

We study group lending by explicitly modeling 'bogus' microfinance groups, that is, groups in which one borrower invests all members' loans into a single project, a practice called *Lei Da Hu* in China. We model the endogenous formation of bogus groups and their coexistence with 'standard' borrower groups, in which each member invests their loan in a separate project.

We highlight two main factors which determine the offered loan terms and the endogenously chosen group form. The first factor is the risk diversification benefit of a standard group – the probability that the borrowers can repay their loans and obtain the continuation value of future credit is strictly higher compared to in a bogus group. This allows lenders to offer lower interest rate and larger loan size to standard groups.

The second factor is the larger expected output (return) in a bogus group with heterogeneous investment projects, since a bogus group always invests all loaned funds into the project with the highest productivity among all member's projects. The trade-off between these two factors underpins our results. The main takeaway is that bogus groups arise when the productivity differential between the group members' projects is sufficiently large, to benefit from the larger expected output. In contrast, the gain from risk diversification prevails among homogeneous borrowers who form standard groups.

An important conclusion is that Lei Da Hu should not be viewed as an undesirable phenomenon which microlenders must eradicate but, instead, as an optimal response by the borrowers which can increase credit allocation efficiency, provided the lenders design appropriate loan terms or menus. This result is predicated on our assumption that group members choose the group form efficiently, to maximize the group payoff. The theoretical and policy implications could differ, for example, if there was coercion, non-transferable utility or fraud, as in fabricating phantom borrowers.

While the available CFPAM data is insufficient for a formal empirical test of our model, our results do provide guidance on how the theory could be tested. The key theoretical prediction (see Lemma 1 and Propositions 3 and 4) is that, all else equal, the larger is the differential between the maximum project productivity (return) and the average productivity within a group, the more likely is to observe a bogus group. Hence, if we could measure the productivity (return) of the investment or business project of each borrower, we would be able to test this prediction. More generally, Lei Da Hu is more likely when one borrower has a high investment return relative to the others in her group, while standard groups should be expected otherwise, when any potential gains from pooling loans into a single project are insufficient to offset the loss in default risk diversification and continued access to credit.

A main ingredient in our model is the asymmetric information friction between the borrowers and the lender regarding the group form, that is, how the members' loans are used after disbursal. If this moral hazard friction were relaxed, then the inefficiency in loan size (e.g., case (L_E, R_E) in Proposition 3) would be removed. However, our conclusion that bogus groups can be efficiency-improving for heterogeneous groups stands.

Our setting parallels, to an extent, the decision problem of an investor who chooses whether to invest a large amount into a single (riskier) asset vs. smaller amounts into multiple assets. However, this parallel is incomplete, since we go beyond individual portfolio choice and in addition model the strategic interaction between a lender and group of borrowers who are jointly liable, and also the strategic interaction within the group of borrowers in one of the extensions (see Appendix B).

A common issue with group lending in its standard form is that it may be difficult for entrepreneurs to find partners with ready-to-go investment projects or business ideas with whom to form a borrowing group. We show that bogus groups offer a possible solution to this problem, by allowing all funds to be invested in a single project while preserving or enabling future access to credit to all group members.

MFIs are unlikely to avoid Lei Da Hu by using sequential lending within groups, e.g., by waiting for a repayment to be received before the next loan is disbursed. The borrowers could simply funnel the funds to the most productive use (in a bogus group) and/or productive borrowers would have to wait and available funds would not be utilized (in a standard group). In addition, the benefit from default risk diversification through joint liability is lost in sequential lending when the cosigners have no other source of wealth.

Finally, the assumption of risk neutrality matters for our results regarding the trade-off between standard vs. bogus groups. If the borrowers and/or lenders were risk averse instead, then there would be an additional insurance benefit from standard groups and diversifying the risk of project failure.

References

[1] Ahlin, C. and B. Waters, 2016, “Dynamic microlending under adverse selection: Can it rival group lending?”, *Journal of Development Economics* 121(C), 237-257.

[2] Ahlin, C. and R. Townsend, 2007a, “Selection into and across credit contracts: Theory and field research”, *Journal of Econometrics* 136, 665–698.

[3] Ahlin, C. and R. Townsend, 2007b, “Using repayment data to test across models of joint liability lending”, *Economic Journal* 117, F11-F54.

[4] Aragon, F., A. Karaivanov and K. Krishnaswamy, 2019, “Credit lines in microcredit: Short-term evidence from a field experiment in India”, working paper, Simon Fraser University.

[5] Armendariz de Aghion, B., 1999, “On the design of a credit agreement with peer monitoring”, *Journal of Development Economics* 60, 79–104.

[6] Banerjee, A., T. Besley and T. Guinnane, 1994, “Thy neighbor’s keeper: the design of a credit cooperative with theory and a test“, *Quarterly Journal of Economics* 109, 491–515.

[7] Bank for International Settlements, 2010, “Microfinance activities and the core principles for effective banking supervision”, Basel Committee on Banking Supervision, https://www.bis.org/publ/bcbs175.pdf

[8] Bernhardt, A., E. Field, R. Pande and N. Rigol, 2019, “Household matters: Revisiting the returns to capital among female micro-entrepreneurs”, *American Economic Review: Insights* (forthcoming)

[9] Besley, T., and S. Coate, 1995, “Group lending, repayment incentives and social collateral”, *Journal of Development Economics* 46, 1–18.

[10] Burki, H-B., 2009, “Unraveling the Delinquency Problem (2008/2009) in Punjab – Pakistan”, *MicroNOTE* No.10, Pakistan Microfinance Network.

[11] Cieslik, K., M. Hudon and P. Verwimp, 2015, “Unruly entrepreneurs – Value creation and value capture by microfinance clients in rural Burundi”, CEB Working Paper #15/013, Universite Libre de Bruxelles.

[12] Chowdhury, P.R., 2005, “Group-lending: Sequential financing, lender monitoring and joint liability”, *Journal of Development Economics* 77, 415–439.

[13] de Mel, S., D. McKenzie and C. Woodruff, 2008, “Returns to capital in microenterprises: evidence from a field experiment”, *Quarterly Journal of Economics* 123(4), 1329–1372.

[14] De Quidt, J., T. Fetzer and M. Ghatak, 2016, “Group lending without joint liability”, *Journal of Development Economics* 121, 217-36.

[15] De Quidt, J., T. Fetzer and M. Ghatak, 2018, “Commercialization and the decline of joint liability credit”, *Journal of Development Economics* 134, 209-225.

[16] Dubey, P., Geanakoplos, J. and Shubik, M., 2005, “Default and punishment in general equilibrium”, *Econometrica* 73(1), 1-37.

[17] Field, E., R. Pande, J. Papp and N. Rigol, 2013, “Does the classic microfinance model discourage entrepreneurship among the poor? Experimental evidence from India”, *American Economic Review* 103(6), 2196-2226.

[18] Gangopadhyay, S., M. Ghatak and R. Lensink, 2005, “Joint liability lending and the peer selection effect”, *Economic Journal* 115, 1005–1015.

[19] Ghatak, M., 1999, “Group lending, local information and peer selection”, *Journal of Development Economics* 60, 27–50.

[20] Ghatak, M., 2000, “Screening by the company you keep: Joint liability lending and the peer selection effect”, *Economic Journal* 110, 601–631.

[21] Ghatak, M. and T. Guinnane, 1999, “The economics of lending with joint liability: theory and practice”, *Journal of Development Economics* 60, 195–228.

[22] Gine, X. and D. Karlan, 2014, “Group versus individual liability: Short and long term evidence from Philippine microcredit lending groups“, *Journal of Development Economics* 107, 65–83.

[23] Karlan, D., 2007, “Social connection and group banking”, *Economic Journal* 117(517), 52-84.

[24] Morduch, J., 1999, “The microfinance promise”, *Journal of Economic Literature* 37, 1569–1614.

[25] Mosley, P., 1986, “Risk, insurance and small farm credit in developing countries: A policy proposal”, *Public Administration and Development* 6, 309–319.

[26] Planet Rating, 2005, “CFPA (China Foundation for Poverty Alleviation) – Zuoquan and Fudan branches: Evaluation report”, www.planetrating.org

[27] Rai, A. and T. Sjostrom, 2004, “Is Grameen lending efficient? Repayment incentives and insurance in village economies”, *Review of Economic Studies* 71, 217–234.

[28] Rothschild, M. and J. Stiglitz, 1976, “Equilibrium in competitive insurance markets: an essay on the economics of imperfect information”, *Quarterly Journal of Economics* 90(4), 629–649.

[29] Stiglitz, J., 1990, “Peer monitoring and credit markets”, *World Bank Economic Review* 4, 351–366.

[30] Udry, C., 1990, “Credit markets in Northern Nigeria: Credit as insurance in a rural economy”, *World Bank Economic Review* 4, 251–269.

[31] Zhang, W., H. Li, and S. Ishida, 2010, "China's non-governmental microcredit practice: History and challenges", *Journal of Family and Economics* 31, 280-296.

Appendix A Proofs

Proof of Proposition 1

(a) Substituting $R = \frac{L}{p(2-p)}$ (shown in the main text) into the no strategic default constraint (2) yields

$$L \leq p(2-p)V \tag{16}$$

Therefore, problem (SP) is equivalent to

$$\max_{L} \ (p(k_i + k_j) - 2)L + 2p(2-p)V \ \text{ s.t. (16)} \tag{17}$$

The objective, (17) is strictly increasing in L, thus (16) binds at optimum. Hence, the loan terms solving problem (SP) are $L_S = p(2-p)V$ and $R_S = \frac{L_S}{p(2-p)} = V$.

(b) Note that the loan terms (L_S, R_S) do not depend on k_i and k_j. Hence, if k_i and k_j are unobservable by the lender, Proposition 1(a) still holds. To show this formally, re-write the objective as weighted sum of ij's total expected payoffs, with weights equal to the population shares of hh, ll and hl groups. Since the no-default and break-even constraints do not depend on k_i and k_j, the result in part (a) obtains.

Proof of Lemma 1

An $ij \in \{ll, hl, hh\}$ borrower pair prefers to form a bogus group instead of a standard group for given loan terms (L, R) if and only if $W_{ij}^{B}(L, R) > W_{ij}^{S}(L, R)$ which, using (SEP) and (BEP), is equivalent to condition (5).

Proof of Proposition 2

(a)-(b) These results follow directly from Lemma 1. At $R = R_S = V$ the r.h.s. of condition ((5)) is zero. For heterogeneous borrower pairs ($k_i > k_j$) the l.h.s. of condition ((5)) is positive, hence they prefer a bogus group. For homogeneous pairs ($k_i = k_j$), the l.h.s. of ((5)) is zero and so they are indifferent between forming bogus or standard group.

(c) By part (a), all heterogenous pairs prefer a bogus group. The lender's expected profit from lending to a bogus group at terms (L_S, R_S) is

$$2pR_S - 2L_S = 2\left(p\frac{L_S}{p(2-p)} - L_S\right) = -\frac{2(1-p)}{2-p}L_S < 0,$$

therefore bogus hl groups cause a loss to the lender.

Proof or Proposition 3

We start with the following observation:

Result 1: *For* $R \leq V$, $W^S(L, \frac{L}{p}) \leq W^B(L, \frac{L}{p})$ *for homogeneous groups if and only if* $L \geq pV$.

Proof: using (SEP) and (BEP), $2pk_iL - 2(2-p)L + 2p(2-p)V \leq 2pk_iL - 2L + 2pV$ is equivalent to $pV \leq L$.□

Call (L^*, R^*) the loan terms solving Problem OP.

Homogeneous groups

Suppose the lender faces a homogeneous borrower pair (ll or hh) and wants to induce a bogus group ($\tau = 0$). Then $R = \frac{L}{p}$ by the break-even constraint. *Result 1* then implies that the incentive constraint (IC) is incompatible with the no-default constraint ($R \leq V$, i.e., $L \leq pV$) unless $L = pV$ and $R = V$, in which case the borrowers are indifferent between the two group forms.

Now suppose the lender wants to induce a standard group ($\tau = 1$). The break-even condition implies $R = \frac{L}{p(2-p)}$, which substituted into the no-default constraint yields $L \leq p(2-p)V$. Constraint (IC) is satisfied for any R satisfying the no-default condition $R \leq V$ (see Lemma 1). The objective $W^S(L, \frac{L}{p(2-p)})$ equals $(2pk_i - 2)L + 2p(2-p)V$, which is strictly increasing in L. Consequently, the no-default constraint written in terms of the loan size, $L \leq p(2-p)V$ must bind at the optimum, implying $L^* = L_S = p(2-p)V$ and $R^* = R_S = V$.

To determine whether inducing standard or bogus group is optimal, compare the respective payoffs

$$W^S(p(2-p)V, V) = 2k_ip^2(2-p)V \text{ and } W^B(pV, V) = 2k_ip^2V.$$

Clearly, $W^S(p(2-p)V, V)$ is larger for any $k_i > 0$. Hence, setting $(L^*, R^*) = (L_S, R_S)$ which induces a standard group is always optimal for homogeneous borrower pairs.

Heterogeneous groups

Suppose first the lender wants to induce a standard group ($\tau = 1$). As before, the break-even constraint implies $R = \frac{L}{p(2-p)}$, which substituted into the no-default constraint implies $L \leq p(2-p)V = L_S$. Unlike in the homogeneous case above, constraint (IC) is no longer automatically satisfied for any R satisfying the no-default condition. To satisfy (IC), i.e., $W^S(L, R) \geq W^B(L, R)$ we need, using Lemma 1:

$$(k_h - k_l)L \leq 2(1-p)(V - \tfrac{L}{p(2-p)}) \text{ or,}$$
$$L \leq \frac{p(2-p)V}{1 + \frac{p(2-p)}{2(1-p)}(k_h - k_l)} \equiv L_E$$

It is easy to verify that $L_S \geq L_E$ whenever $k_h \geq k_l$. Therefore, the loan terms in this case are $L^* \equiv L_E$ and $R^* \equiv \frac{L_E}{p(2-p)}$.

Suppose instead the lender wants to induce a bogus group ($\tau = 0$). The break-even constraint implies $R = \frac{L}{p}$ and the no-default condition is $R \leq V$. By Lemma 1, constraint (IC) is clearly satisfied for $R = V$ and $L = pV$. Therefore, the offered loan terms are $(L^*, R^*) = (L_B, R_B)$.

To determine whether inducing standard or bogus group is optimal, compare the payoff of a standard group with loan terms (L_E, R_E) with that of a bogus group with terms (L_B, R_B). Choosing $L^* = L_E$ and

$R^* = R_E$ which induces a standard group is optimal if $W^S(L_E, R_E) \geq W^B(L_B, R_B)$, that is:

$$(p(k_l + k_h) - 2)\frac{p(2-p)V}{1+\frac{p(2-p)}{2(1-p)}(k_h - k_l)} + 2(2-p)pV \geq (2pk_h - 2)pV + 2pV \tag{CC3}$$

or equivalently, $f(k_l) \geq k_h$

where

$$f(k) \equiv \frac{1}{2}\left(k + c + \sqrt{(k+c)^2 - \frac{4k}{p}}\right) \text{ and } c \equiv \frac{2p^2 - 5p + 4}{p(2-p)}.$$

Choosing $L^* = L_B$ and $R^* = R_B$ (inducing a bogus group) is optimal otherwise, for $f(k_l) < k_h$.[29] Call $\bar{k} = f(k_l)$ as defined above.

The table in the statement of Proposition 3 summarize the results proven above.■

Proof of Proposition 4

Substituting for R_N and R_M from the break-even conditions (9) and (7), Problem UP simplifies to:

$$\max_{L_N, L_M, \tau_{ij}} \sum_{ij} q_{ij}(\tau_{ij} W_{ij}^S(L_N, \tfrac{L_N}{p(2-p)}) + (1-\tau_{ij})W_{ij}^B(L_M, \tfrac{L_M}{p})) \tag{18}$$

subject to

$$L_M \leq pV \tag{19}$$

$$L_N \leq p(2-p)V \tag{20}$$

$$\begin{aligned} &\tau_{ij} W_{ij}^S(L_N, \tfrac{L_N}{p(2-p)}) + (1-\tau_{ij})W_{ij}^B(L_M, \tfrac{L_M}{p}) \geq \\ &\geq \max\left\{W_{ij}^B(L_N, \tfrac{L_N}{p(2-p)}), W_{ij}^S(L_M, \tfrac{L_M}{p})\right\} \quad \forall ij \in \{hh, hl, ll\} \end{aligned} \tag{IC}$$

where, using (SEP) and (BEP),

$$W_{ij}^S(L_N, \tfrac{L_N}{p(2-p)}) = (p(k_i + k_j) - 2)L_N + 2p(2-p)V \tag{21}$$

$$W_{ij}^B(L_N, \tfrac{L_N}{p(2-p)}) = 2(pk_i - \tfrac{1}{2-p})L_N + 2pV \tag{22}$$

$$W_{ij}^B(L_M, \tfrac{L_M}{p}) = 2(pk_i - 1)L_M + 2pV \tag{23}$$

$$W_{ij}^S(L_M, \tfrac{L_M}{p}) = (p(k_i + k_j) - 2(2-p))L_M + 2p(2-p)V \tag{24}$$

Lemma A0: *Homogeneous groups are standard at the solution to Problem UP, that is* $\tau_{ll} = \tau_{hh} = 1$.

Proof: In Problem UP it is impossible to induce a bogus group (that is, $\tau_{ii} = 0$) for homogeneous borrower pairs, hh or ll. The reason is that for any offered contract (L, R) with $R \leq V$ we have $W_{ii}^S(L, R) \geq W_{ii}^B(L, R)$ (strictly if $R < V$), since for given (L, R) expected output is the same but the expected future

[29]Re-write $W^S(L_E, R_E) = W^B(L_B, R_B)$ as a quadratic equation in terms of k_h. It is easy to show that its larger root equals $f(k_l)$ while its smaller root is strictly smaller than k_l. Thus, for $k_h \geq k_l$ inequality (CC3) is satisfied if and only if $k_h \leq f(k_l)$.

value net of repayment is larger in a standard group. This implies that the incentive constraint (IC) cannot be satisfied for $\tau_{ii} = 0$ and hence $\tau_{ll} = \tau_{hh} = 1$ – homogeneous groups are always standard.$\square$

Given Lemma A0, to satisfy constraint (IC) for homogeneous groups the lender thus only needs to ensure that $W_{ii}^S(L_N, \frac{L_N}{p(2-p)}) \geq W_{ii}^S(L_M, \frac{L_M}{p})$. Depending on the model parameters k_h, k_l and p there are two possibilities for τ_{hl} and the loan menu $\mathcal{N}, \mathcal{M}$, considered respectively in Case A and Case B below.

Case A: The lender finds it optimal to induce heterogeneous hl groups to be bogus, that is $\tau_{hl} = 0$, or,

$$W_{hl}^B(L_M, \tfrac{L_M}{p}) > W_{hl}^S(L_N, \tfrac{L_N}{p(2-p)}) \tag{25}$$

Since $\tau_{ll} = \tau_{hh} = 1$, the incentive constraints (IC) in this case are:

$$\begin{aligned} W_{hl}^B(L_M, \tfrac{L_M}{p}) &\geq W_{hl}^S(L_M, \tfrac{L_M}{p}) & (26)\\ W_{hl}^B(L_M, \tfrac{L_M}{p}) &\geq W_{hl}^B(L_N, \tfrac{L_N}{p(2-p)}) & (27)\\ W_{ii}^S(L_N, \tfrac{L_N}{p(2-p)}) &\geq W_{ii}^S(L_M, \tfrac{L_M}{p}) & (28) \end{aligned}$$

We show that $L_M = pV$ (its maximum possible value since $R_M = \frac{L_M}{p} \leq V$) is optimal. Suppose not, that is $L_M < pV$. Using expressions (23) and (21),

$$\frac{dW_{hl}^B(L_M, \frac{L_M}{p})}{dL_M} = 2(pk_h - 1) > p(k_h + k_l) - 2(2-p) = \frac{dW_{hl}^S(L_M, \frac{L_M}{p})}{dL_M}$$

Therefore, by increasing L_M by a small amount (so that $L_M + \varepsilon < pV$) while holding L_N constant, constraints (26) and (27) remain satisfied.

For the moment ignore constraint (28), as if solving a relaxed optimization problem. We will then show that the solution L_N, L_M of this relaxed problem also satisfies (28), that is, (28) is not binding at the solution to Problem UP. Since $W^B(L_M, \frac{L_M}{p})$ is strictly increasing in L_M by (23), the objective function (18) weakly increases in L_M, all else equal – a contradiction with the assumed optimality of $L_M < pV$. Therefore, setting $L_M = pV = L_B$ and $R_M = V = R_B$ is optimal for the relaxed problem.

At $L_M = pV$, expressions (23) and (24) imply

$$W_{hl}^B(L_M, \tfrac{L_M}{p}) = 2k_h p^2 V \geq (k_h + k_l)p^2 V = W_{hl}^S(L_M, \tfrac{L_M}{p}). \tag{29}$$

Thus, since $k_h \geq k_l$, constraint (26) is satisfied. Constraint (27) is equivalent to

$$W_{hl}^B(pV, V) \geq W_{hl}^B(L_N, \tfrac{L_N}{p(2-p)}) \Leftrightarrow L_N \leq L_F \equiv \frac{pk_h - 1}{pk_h - \frac{1}{2-p}} pV \tag{30}$$

Condition (25) is equivalent to

$$W_{hl}^S(L_N, \tfrac{L_N}{p(2-p)}) < W_{hl}^B(pV, V) \Leftrightarrow L_N < L_3(hl) \equiv \frac{2pk_h - 4 + 2p}{p(k_h + k_l) - 2} pV \tag{31}$$

We also need $L_N \leq p(2-p)V = L_S$ from the no-default constraint. Using (21), the objective function (18) is weakly increasing in L_N. Hence L_N is the largest possible value satisfying both the incentive constraints and no-default constraint (note that a larger L_N also relaxes constraint (28) that we still have not verified). It is easy to show that $L_S > L_F$ for any $p \in (0,1)$ and $k_h > 0$; hence the no-default constraint (20) holds for any $L_N \leq L_F$. Overall, this implies that to satisfy (27) and (25) we must have $L_N \leq L_F$ and $L_N < L_3(hl)$. We now check that the remaining incentive constraint, (28) is satisfied. Remember from Lemma 1 that $W_{ii}^S(L,V) = W_{ii}^B(L,V)$ for any L, hence (28) is equivalent to

$$W_{ii}^S(L_N, \tfrac{L_N}{p(2-p)}) \geq W_{ii}^B(pV,V) \Leftrightarrow L_N \geq L_3(ii) \equiv \tfrac{pk_i-(2-p)}{pk_i-1}pV \tag{32}$$

Using the definitions of L_F in (30), $L_3(hl)$ in (31) and $L_3(ii)$ in (32), it is easy to verify directly that $L_3(hl) \geq L_3(hh) > L_3(ll)$ and that $L_F > L_3(hh)$, for any $k_h > k_l > 0$ and $p \in (0,1)$. Thus, using (32), constraint (28) is indeed satisfied for any candidate solution $L_N = \min\{L_F, L_3(hl)\}$. We solve for the value of L_N at the end of the proof.

Case B: The lender finds it optimal to induce heterogeneous hl groups to be standard, that is $\tau_{hl} = 1$, or,

$$W_{hl}^S(L_N, \tfrac{L_N}{p(2-p)}) \geq W_{hl}^B(L_M, \tfrac{L_M}{p}). \tag{33}$$

Since $\tau_{ll} = \tau_{hh} = 1$ from Lemma A0, the objective function (18) equals $\sum_{ij} q_{ij} W_{ij}^S(L_N, \frac{L_N}{p(2-p)})$, strictly increasing in L_N. The incentive constraints (IC) in Case B are:

$$W_{hl}^S(L_N, \tfrac{L_N}{p(2-p)}) \geq W_{hl}^B(L_N, \tfrac{L_N}{p(2-p)}) \tag{34}$$
$$W_{hl}^S(L_N, \tfrac{L_N}{p(2-p)}) \geq W_{hl}^S(L_M, \tfrac{L_M}{p}) \tag{35}$$
$$W_{ii}^S(L_N, \tfrac{L_N}{p(2-p)}) \geq W_{ii}^S(L_M, \tfrac{L_M}{p}) \tag{36}$$

Constraint (34) is equivalent to

$$L_N \leq \frac{p(2-p)V}{1+\frac{p(2-p)}{2(1-p)}(k_h-k_l)} = L_E \tag{37}$$

where we used the definition of L_E in Proposition 3. We showed in Proposition 3 that $L_E < L_S = p(2-p)V$, hence the no-default constraint holds for any $L_N \leq L_E$. We will show that condition (33) and constraints (35) and (36) are all satisfied at $L_M = pV = L_B$ and $R_M = V$ (since in Case B no bogus groups are optimally induced, this is without loss of generality). Note that at $L_M = pV$, condition (33) is equivalent to

$$W_{hl}^S(L_N, \tfrac{L_N}{p(2-p)}) \geq W_{hl}^B(pV,V) \iff L_N \geq L_3(hl) \tag{38}$$

where $L_3(hl)$ was defined as in (31).

The loan terms

To complete the analysis, we determine solution of Problem UP for any values of the model parameters p, k_h, k_l. First, it is easy to show that,

$$L_3(hl) \leq L_E \iff k_h \leq f(k_l) \equiv \bar{k} \tag{39}$$

where $\bar{k} \equiv f(k_l)$ and the function $f(k)$ was defined in the proof of Proposition 3. Second, we can also show[30]

$$L_3(hl) > L_F \iff k_h > f(k_l). \quad (40)$$

Parameters $k_h \in [k_l, \bar{k}]$

Result (37) and the monotonicity of borrowers' payoffs imply that setting $L_N = L_E$ maximizes the objective (18) subject to constraint (34). Results (38) and (39) imply that condition (33) is satisfied for $L_N = L_E$ and $L_M = pV = L_B$. Constraint (35) is satisfied at $L_M = pV$ since condition (33) is satisfied and since $W_{hl}^B(L_M, \frac{L_M}{p}) \geq W_{hl}^S(L_M, \frac{L_M}{p})$ by (29). Finally, using $L_E \geq L_3(hl)$ by (39), $L_3(hl) \geq L_3(hh) > L_3(ll)$ (directly verified) and (32) together imply that constraint (36) is also satisfied. We have thus shown that $L_N = L_E$ and $R_N = \frac{L_N}{p(2-p)}$ solve Problem UP for $k_h \in (k_l, \bar{k})$ and that all borrower pairs choose loan $\mathcal{N} = (L_N, R_N)$ and form standard groups. We set $L_M = pV = L_B$ and $R_M = V = R_B$ without loss of generality (loan $\mathcal{M}$ is not used).

Parameters $k_h > \bar{k}$

Result (30) and the monotonicity of the group payoff in (22) imply that setting $L_N = L_F$ maximizes the objective subject to constraint (27). Results (31) and (40) imply that condition (25) is satisfied at $L_N = L_F$ and $L_M = pV = L_B$. The remaining incentive constraints (26) and (28) were shown to be satisfied in (29) and (32). Putting all these results together, we have shown that the solution to Problem UP for $k_h > \bar{k}$ is the loan menu $(\mathcal{N}, \mathcal{M})$ with $\mathcal{N} = \mathcal{F} \equiv (L_F, \frac{L_F}{p(2-p)})$ and $\mathcal{M} = \mathcal{B} = (L_B, \frac{L_B}{p})$. Homogeneous borrower pairs choose loan $\mathcal{F}$ and form standard groups while heterogeneous pairs choose loan $\mathcal{B}$ and form bogus groups.

Proof of Lemma 2

From the results in Section 3 the maximum feasible payoff (subject to no strategic default and breaking even) of a standard group is $W_{ij}^S(L_S, R_S)$ and the maximum feasible payoff of a bogus group is $W_{ij}^B(L_B, R_B)$. The lender would optimally offer loan terms depending on the model parameters and the efficient group form for any given borrower pair. We already showed that homogeneous groups have no incentive to be bogus, since $W_{ii}^S(L_S, R_S) \geq W_{ii}^B(L_B, R_B)$ for $ii = ll$ or hh. Hence the lender would optimally offer loan $\mathcal{S}$ to homogeneous borrower pairs, for any model parameters satisfying Assumption 1.

Consider now heterogeneous borrower pairs, hl. Comparing $W_{hl}^B(L_B, R_B)$ with $W_{hl}^S(L_S, R_S)$, we have

$$\begin{aligned} W_{hl}^B(L_B, R_B) &= 2pk_h pV > p(k_h + k_l)p(2-p)V = W_{hl}^S(L_S, R_S) \\ &\iff k_h > \tfrac{2-p}{p} k_l. \end{aligned}$$

Note that, for given k_l, the value $\frac{2-p}{p} k_l$ is strictly larger than the threshold $\bar{k} = f(k_l)$ defined in Propositions 3 and 4. This is true since $\bar{k}$ is the value of k_h that equalizes $W_{hl}^B(L_B, R_B)$ with $W_{hl}^S(L_E, R_E)$, while $k_h = \frac{2-p}{p} k_l$ equalizes $W_{hl}^B(L_B, R_B)$ with $W_{hl}^S(L_S, R_S)$ and since $W_{hl}^S(L_S, R_S) > W_{hl}^S(L_E, R_E)$ and $W_{hl}^B(L_B, R_B) = 2p^2 k_h V$ is increasing in k_h. Taken together, these results imply that for $k_h \leq \frac{2-p}{p} k_l$ it is optimal to offer loan $\mathcal{S}$ to hl borrower pairs, while if $k_h > \frac{2-p}{p} k_l$ the lender offers loan $\mathcal{B}$.

[30]Obviously $L_E \neq L_F$ in general, however, for any given k_l, the expressions $L_3(hl)$, L_F and L_E taken as functions of k_h cross at the same point (at $k_h = f(k_l)$) and thus $L_3(hl)$ is larger than both L_F and L_E whenever $k_h \geq f(k_l)$ and smaller otherwise.

Proof of Lemma 3

Using the expressions for $W_{ij}^B(L,R)$ and $W_{ij}^S(L,R)$ derived earlier in (BEP) and (SEP), we re-write the lender's Problem UP′ equivalently as:

Problem UP2

$$\max_{L_N,R_N,L_M,R_M} q_{ll}(2pk_lL_N + 2p(2-p)(V-R_N)) + q_{hh}(2pk_hL_N + 2p(2-p)(V-R_N)) \tag{41}$$
$$+q_{hl}(2pk_hL_M + 2p(V-R_M))$$

subject to:

$$R_M \leq V,\ \ R_N \leq V \ \ \text{(no default)}$$
$$(1-q_{hl})p(2-p)R_N + q_{hl}pR_M \geq (1-q_{hl})L_N + q_{hl}L_M \ \ \text{(break even)} \tag{42}$$
$$2pk_hL_M + 2p(V-R_M) \geq p(k_h+k_l)L_N + 2p(2-p)(V-R_N) \tag{43}$$
$$2pk_hL_M + 2p(V-R_M) \geq p(k_h+k_l)L_M + 2p(2-p)(V-R_M) \tag{44}$$
$$2pk_hL_M + 2p(V-R_M) \geq 2pk_hL_N + 2p(V-R_N) \tag{45}$$
$$2pk_iL_N + 2p(2-p)(V-R_N) \geq 2pk_iL_M + 2p(2-p)(V-R_M) \text{ for } i = l,h \tag{46}$$

Constraints (43)–(46) are the incentive/selection constraints. The constraint set of Problem UP2 is non-empty (e.g., loans $\mathcal{F},\mathcal{B}$ from Proposition 4 satisfy it). We will first prove by contradiction that at UP2's solution it must be that $R_M = V$ and $R_N = V$ (the no-default constraints bind).

Suppose that $R_M < V$ at the solution to Problem UP2 and consider increasing R_M by a small amount ε (so that $R_M + \varepsilon \leq V$) and simultaneously increasing L_M by a number just slightly larger than $\frac{\varepsilon}{k_h}$ while holding R_N and L_N fixed. The objective increases by the choice of variations in R_M and L_M. The left hand sides of constraints (43) and (45) increase for the same reason, hence they remain satisfied. Constraint (44) remains satisfied too, since we can re-write it as $(k_h - k_l)L_M \geq 2(1-p)(V-R_M)$. Constraint (46) stays satisfied since its r.h.s. decreases. Finally, note that the left-hand (revenue) side of the break-even constraint (42) goes up by $q_{hl}p\varepsilon$ which is larger than the increase of its right-hand (outlays) side, since $pk_h > 1$ by Assumption 1. In sum, these results imply that a larger borrower payoff can be achieved in an incentive-compatible and feasible way – a contradiction with the assumed optimality of R_M. This implies that $R_M = V$ at optimum.

Suppose now that $R_N < V$ at the solution of Problem UP2. At optimum the break-even constraint (42) must hold at equality, hence we can substitute it into the objective (41) and re-write the latter (dropping the constants) equivalently as:

$$\max_{L_N,R_N,L_M,R_M} q_{ll}(pk_l-1)L_N + q_{hh}(pk_h-1)L_N + q_{hl}(pk_h-1)L_M \tag{47}$$

Using $R_M = V$, breaking even implies:

$$L_M = \frac{(1-q_{hl})(p(2-p)R_N - L_N)}{q_{hl}} + pV \tag{48}$$

Incentive constraint (44) is always satisfied at $R_M = V$ since it is equivalent to $(k_h - k_l)L_M \geq 0$. Constraints (45) and (46) can be written as:

$$k_h(L_M - L_N) \geq V - R_N \geq \frac{k_i}{2-p}(L_M - L_N) \tag{49}$$

Still supposing that $R_N < V$, inequalities (49) imply $L_M > L_N$ and (since $k_h \geq k_l$) that (46) cannot bind for $k_i = k_l$.

Suppose (46) does not bind, that is, $V - R_N > \frac{k_h}{2-p}(L_M - L_N)$. Hold L_N fixed and consider a small increase in R_N (and corresponding increase in L_M in (48) to satisfy breaking even), so that (46) remains satisfied. The objective function (47) increases. Constraints (43) and (45) remain satisfied since their l.h.s. increases with L_M while their r.h.s. decreases with R_N. This implies that R_N and L_M can be increased, raising the objective and satisfying all constraints, until either (46) binds (for k_h) or until $R_N = V$, whichever happens first.

If $R_N = V$ happens first, then (49) implies $L_M = L_N$, and from the break-even constraint we find

$$L_M = L_N = \tilde{L} = ((1 - q_{hl})(2 - p) + q_{hl})pV$$

as defined in the Lemma 3 statement.

Suppose, instead, that $R_N < V$ at the solution to Problem UP2 (and hence $L_M > L_N$) but (46) binds, that is,

$$V - R_N = \frac{k_h}{2-p}(L_M - L_N) \tag{50}$$

Use (48) and (50) to express L_M and L_N in terms of R_N,

$$\begin{aligned} L_M &= [(2-p)(1-q_{hl})(p - \tfrac{1}{k_h})]R_N + q_{hl}pV + \tfrac{(1-q_{hl})(2-p)V}{k_h} \\ L_N &= [(2-p)(1-q_{hl})p + \tfrac{(2-p)q_{hl}}{k_h}]R_N + q_{hl}pV - \tfrac{q_{hl}(2-p)V}{k_h} \end{aligned}$$

Note that a small local increase in R_N raises both L_M and L_N, which raises the objective function (47). Substituting $R_M = V$ and the expression for $V - R_N$ from (50) into incentive constraint (43), the latter is equivalent to $2pk_hL_M \geq p(k_h + k_l)L_N + 2pk_h(L_M - L_N)$ which is true since $k_h \geq k_l$. Similarly, constraint (45) is equivalent to $2pk_hL_M \geq 2pk_hL_N + \frac{2pk_h}{2-p}(L_M - L_N)$ which is satisfied for $L_M > L_N$ since $1 > \frac{1}{2-p}$. Since we already showed that (44) is always satisfied at $R_M = V$ and since (46) and (42) are satisfied by construction via (48) and (50), this yields a contradiction with the assumed optimality of L_M, L_N and the assumption $R_N < V$. Hence, we must have $R_N = V = R_M$ and $L_M = L_N$ at optimum. Again, from the break-even constraint, we then find $\tilde{L}$ as defined in the Lemma 3 statement.

Proof of Proposition 5

Take a pair of borrowers with productivities k_i, k_j where $k_i = \max\{k_i, k_j\}$. If both receive individual loans with terms $(L_I, R_I) = (pV, V)$, then the group's total payoff is

$$\pi^I(k_i, k_j) = p^2(k_i + k_j)V,$$

which equals $2p^2 k_i V$ for a homogeneous pair (k_i, k_i). In contrast, by Proposition 3, a homogeneous pair (k_i, k_i) borrowing at the standard group loan terms (L_S, R_S) has payoff,

$$\pi^S(k_i, k_i) = 2p^2 k_i (2 - p)V,$$

which is clearly larger than the pair's total payoff with individual loans, $2p^2 k_i V$. Intuitively, joint liability allows risk diversification (members covering for each other), which reduces the risk of default and allows for lower interest rate and larger loan size, while keeping the required repayment the same as in an individual loan. Hence, homogeneous borrower pairs would have no incentive to switch to individual loans.

Consider now a heterogeneous pair (k_i, k_j) with $k_i > k_j$. If it forms a bogus group, its total payoff at (L_B, R_B) is

$$\pi^B(k_i, k_j) = 2p^2 k_i V$$

which is strictly larger than $\pi^I(k_i, k_j)$. Again, the two borrowers have no incentive to use individual loans (transferable utility is important for this result). Finally, suppose a heterogeneous group faces the loan terms (L_E, R_E) as in Proposition 3 case $k_h \leq \bar{k}$, and call its joint payoff $\pi^E(k_i, k_j)$. We know (see the proof of Proposition 3) that the group's payoff satisfies $W^S(L_E, R_E) \geq W^B(L_B, R_B)$, which implies $\pi^E(k_i, k_j) \geq \pi^B(k_i, k_j)$ and, since $\pi^B(k_i, k_j) > \pi^I(k_i, k_j)$ as shown before, the borrowers are once again better off with their group loan $\mathcal{E}$.

In sum, we conclude that, starting from the loan terms described in Proposition 3, if individual loans were made available at the best possible terms for the borrowers $(L_I, R_I) = (pV, V)$ and if payoffs are transferable within the group, no borrower group would prefer to switch to individual loans.

Appendix B Individual repayment decisions

B.1 Setting

Consider an alternative setting in which the group members make the decision to repay or default individually (non-cooperatively) instead of jointly as assumed in Sections 3 and 4.[31] Similarly to Besley and Coate (1995), think of the borrowers' repayment choices as a two-stage game. In the first stage, each borrower is asked by the lender to repay R and can either do that or report default. If both borrowers repay R or if both default in the first stage, their respective payoffs (described below) are realized and the game ends. The second stage is reached only if one borrower has repaid R in stage 1 while her partner has defaulted. In that case, the borrower who repaid is asked by the lender to pay additional R for her partner, as stipulated by the joint liability clause. Again, the borrower chooses to repay or default. We solve for the subgame-perfect Nash equilibrium of the described game.

Start with stage 2 – the decision facing a borrower with successful project who has repaid R in stage 1. It is never optimal to make a partial repayment (strictly between 0 or R) since either defaulting (repaying zero) and forfeiting the continuation value V, or repaying in full and securing the value V is dominant strategy

[31] For example, each borrower may only be able to verify her own project outcome, as in Armendariz de Aghion (1999), or within-group sanctions cannot be imposed.

(this assumes that the lender considers incomplete repayment a default). The borrower thus repays in stage 2 if her payoff, $-R+V$ is larger than the payoff of defaulting, 0 or:

$$R \leq V \quad \text{(S2R)}$$

Suppose (S2R) holds, so that each group member would repay if Stage 2 is reached and proceed by backward induction to stage 1. Conditional on project success, the borrowers play a simultaneous move game with normal form presented in Table 1. Only the payoffs for the row player i are listed; the payoffs for the column player j are symmetric across the diagonal. If a borrower's project fails she announces default. As before, partial repayment is dominated by either defaulting or repaying in full.

Table 1: Individual repayment – stage 1 game

	Repay	Default
Repay	$k_iL - R - (1-p)R + V$	$k_iL - 2R + V$
Default	$k_iL + pV$	k_iL

The repayment payoff in Table 1 reflects the fact that, assuming (S2R) holds, a borrower playing Repay in stage 1 while her partner also plays Repay, would only be asked to repay extra R in stage 2 with probability $1-p$ (if her partner's project fails).[32] Conditional on j playing Repay in stage 1, borrower i would also choose Repay when her project succeeds if her expected payoff from repaying R is larger than her payoff from strategically defaulting (repaying zero), that is, if $k_iL - R - (1-p)R + V \geq k_iL + pV$, or

$$R \leq \frac{1-p}{2-p}V. \tag{51}$$

Similarly, conditional on borrower j choosing Default, borrower i would choose Repay if paying back $2R$ over the two stages and securing the continuation value V results in a larger payoff than declaring default, that is, if $k_iL - 2R + V \geq k_iL$, or

$$R \leq \frac{1}{2}V. \tag{52}$$

Since $\frac{1-p}{2-p} < \frac{1}{2}$, condition (51) implies conditions (52) and (S2R). Thus, for $R \leq \frac{1-p}{2-p}V$ (Repay, Repay) is the unique Nash equilibrium of the stage 1 game (Repay is at least weakly dominant strategy), and repayment is also optimal at stage 2.[33] Comparing constraint (51) with constraint (2) in Section 3 (joint repayment choice) we see that assuming non-cooperative repayment decisions restricts the maximum feasible repayment (and hence the loan size) for which no strategic default can be supported. The intuition is that without coordination each borrower can free-ride on the repayment of the other group member, which increases the incentive for strategic default compared to the joint-decision setting.

[32] Here, consistent with the non-cooperative assumption, we assume the borrower makes her stage 1 repayment decision without observing her partner's project outcome. The analysis would not change qualitatively if instead the borrower knew that her partner's project has failed, in which case the no default condition becomes $R \leq \frac{(1-p)V}{2}$.

[33] See online Appendix B.3 for further discussion.

The analysis in Sections 3 and 4 can be re-done in a relatively straightforward way for the case with individual repayment decisions. We first show that, with standard groups only (as in Proposition 1), individual repayment decisions imply *strictly smaller* loan size, $L_{S'} < L_S$ (and, correspondingly, smaller repayment $R_{S'}$ and group payoff). The reason is the additional incentive for strategic default from being able to free ride on the repayment of the other member. Specifically, $L_{S'} = p(1-p)V$ and $R_{S'} = \frac{1-p}{2-p}V$ – see Proposition B-1 in Appendix B.2. The interest rate is the same as in Proposition 1 since it is determined by the break-even condition.

We then compare standard and bogus groups and show that the basic intuition from Section 3 still holds. Standard groups offer a diversification benefit and lower interest rate while expected output is larger in bogus groups. However, there is an additional dimension to the comparison, since the incentive for strategic default in a standard group is now stronger because of the free-riding possibility – see the discussion surrounding condition (51). In a bogus group this free-riding effect is absent, since the inactive partner has no income and cannot repay.[34] The additional free-riding effect is present in both heterogeneous and homogeneous pairs and results in an additional benefit from forming a bogus group.

As in Proposition 2, when bogus groups can form endogenously, offering standard loan terms to all groups can cause a loss to the lender (see Proposition B-2 below). This happens for borrowers with sufficiently heterogeneous projects.

We also characterize the constrained-optimal loan terms when the lender takes into account the endogenous choice of group form (see Proposition B-3 for the observable k's setting). Compared to Proposition 3, there exist more cases dependent on the parameter values for p, k_l and k_h. The basic intuition – that more heterogeneity in the productivities k_i and k_j makes bogus groups more likely – still holds. What is different from the joint-decision setting, however, is that with individual repayment decisions bogus groups may also be formed by homogeneous pairs. This happens either when the success probability p or when the project productivities are sufficiently high. Intuitively, the gains from forming a bogus group are the largest when the risk of failure is relatively low (large p) and/or when project productivity is large, so that the benefits from the reduced free riding incentive and larger loan size in a bogus group outweigh the loss of risk diversification. Standard groups are optimal otherwise.

For heterogeneous borrower pairs, either offering loan terms inducing bogus group is optimal, which happens when the productivity differential or level is sufficiently high enough; or loan terms inducing standard group are optimal, when the project productivities are relatively low and close to each other. As in Proposition 3, the loan terms inducing a bogus group do not depend on the values k_i, k_j. The only loan terms that depend on the productivity values k_i, k_j are $\mathcal{E}$ since, to deter an hl group from switching to bogus, it is necessary to take into account the output gain from changing the group form, which depends on k_h and k_l. Finally, as in Section 4, borrower welfare and productive efficiency are improved when the endogenous formation of bogus groups is incorporated in the lender's strategy and the borrowers are offered appropriate (choice of) loan terms.

[34]Note that deciding to strategically default in a bogus group does not require additional coordination compared to in a standard group. In a standard group the joint liability (JL) clause means that if member 1 decides to strategically default, the lender can go to member 2 and collect $2R$ if 2 declares success. In a bogus group JL implies that if 1 decides to default, the lender can still go to member 2 but cannot collect anything due to limited liability.

B.2 Results

Given loan terms (L, R) satisfying feasibility (1) and the no strategic default condition (51), the expected total payoff of standard group of type $ij \in \{hh, ll, hl\}$ equals

$$\bar{W}_{ij}^{S}(L, R) = p(k_i + k_j)L - 2p(2-p)R + 2p(2-p)V, \tag{SEP}$$

where we use $\bar{W}$ to indicate all payoffs in the individual default decision setting.

The lender's break even condition remains as in (3),

$$R = \tfrac{L}{p(2-p)}.$$

Standard groups

As in Section 3, we first characterize the loan terms if bogus groups are assumed away exogenously.

Proposition B-1
(a) With standard groups only and individual default decisions, the optimal loan terms are $\mathcal{S}' \equiv (L_{S'}, R_{S'})$ *with*

$$L_{S'} = p(1-p)V \text{ and } R_{S'} = \tfrac{1-p}{2-p}V.$$

(b) The loan terms $\mathcal{S}'$ *remain optimal if the borrowers' productivities are unobservable to the lender.*

The proof is very similar to the proof of Proposition 1 and hence omitted.

Bogus groups

To introduce bogus groups in the individual-default setting, we follow the microfinance literature and assume that the group members share social capital which can be used to enforce a (possibly non-monetary) transfer T between them in order to ensure that the bogus group cosigner ('ghost member') obtains at least the same payoff as she would in a standard group under repayment.[35] More precisely, assume that the bogus group *leader* (the member who invests $2L$) makes a default/repay decision individually, based on his own payoff as the borrowers in a standard group, however, he does so with the understanding that a transfer T must be made in either case. The repay/default trade-off is thus unaffected by the transfer. The cosigner has no decisions to make since her project is not funded and there is limited liability (she has no other wealth or income).

If the borrowers consider forming a bogus group, they optimally invest all funds ($2L$) into the higher productivity project. Assuming $k_i \geq k_j$ without loss of generality, all funds are invested in project i. Consider the bogus group leader's repay/default decision for given loan terms (L, R). As before, it is not optimal to repay partially. Repayment in stage 2 (reached if the leader repays in stage 1) is optimal as long as

$$R \leq V,$$

[35] It is easy to compute the required transfer amount and show that it is always feasible upon project success, both under repayment and strategic default. Alternatively, the transfer can be non-monetary.

yielding a payoff of $V-R$. Conditional on project success, the bogus group leader's payoff from repaying in stage 1 is larger than her payoff from defaulting if

$$k_i(2L) - R - T + (V - R) \geq k_i(2L) - T \text{ or, } R \leq V/2$$

The no-strategic-default condition over both stages 1 and 2 is then:

$$R \leq \tfrac{V}{2} \tag{53}$$

This implies that, for given loan terms (L, R), the joint expected payoff $\bar{W}^B_{ij}(L, R)$ of a bogus group with productivities k_i, k_j is:

$$\bar{W}^B_{ij}(L, R) = 2pk_iL - 2pR + 2pV \tag{54}$$

Breaking even implies $L = pR$ as in Section 3. Combined with the no-default condition (53), the payoff-maximizing loan terms for a bogus group (as if known and taken in isolation) are

$$L_{B'} \equiv \tfrac{pV}{2} \text{ and } R_{B'} \equiv \tfrac{V}{2}. \tag{55}$$

Comparing bogus and standard groups, the basic intuition from Section 3.3 still holds, as discussed in Section 5.2 in the main text.

It is easy to show that Lemma 1 still applies. The following Proposition is the counterpart of Proposition 2 and shows the consequence of endogenous bogus group formation in the lender were to offer terms $\mathcal{S}'$ to all borrowers.

Proposition B-2: *If loan $\mathcal{S}' = (L_{S'}, R_{S'})$ defined in Proposition B-1 is offered to all borrowers, then:*
(a) if

$$\Delta \equiv k_h - k_l > \tfrac{2}{p(2-p)}, \tag{56}$$

heterogeneous (hl) borrower pairs form bogus groups; otherwise hl pairs form standard groups;
(b) all homogeneous (hh or ll) borrower pairs form standard groups;
(c) bogus groups cause a loss to the lender.

Proof: see Appendix B.3.

The previous intuition (see Proposition 2) carries over – at loan terms $\mathcal{S}'$ designed for standard groups only, bogus groups are strictly optimal only for heterogeneous borrower pairs. An additional condition is now needed – the productivity differential across the two project in an hl pair must be large enough for the borrowers to form a bogus group. If condition (56) does not hold (that is, for k_h and k_l relatively close), then offering loan terms $(L_{S'}, R_{S'})$ would not cause a loss to the lender since all groups will be standard. However, as we show below, offering these terms to all borrowers may not be optimal since allowing (inducing) bogus groups can be efficiency-improving as larger amount of funds can be invested in the higher return project.

Endogenous group form and loan terms

Here we only analyze the case of observable project productivities. The analysis of the unobservable productivity setting is similar to that in Section 4.2 and is available upon request.

Proposition B-3: *Suppose repayment decisions are individual and the productivities k_i and k_j are observed by the lender. The optimal loan terms (L^*, R^*) for an $ij \in \{hh, ll, hl\}$ group are:*
(a) homogeneous, $k_i = k_j$ (hh or ll) groups: if $p(2p-1)k_i > 1$ then $L^* = \frac{pV}{2} \equiv L_{B'}$, $R^* = \frac{V}{2} \equiv R_{B'}$ and the group is bogus ($\tau^* = 0$) while if $p(2p-1)k_i \leq 1$ then $L^* = p(1-p)V = L_{S'}$, $R^* = \frac{(1-p)V}{2-p} = R_{S'}$ and the group is standard ($\tau^* = 1$).
(b) heterogeneous, $k_i > k_j$ (hl) groups: depending on the parameter values,[36] either $L^* = L_{B'}$ and $R^* = R_{B'}$ and the group is bogus or $L^* = \min\{L_{S'}, L_E\}$, $R^* = \frac{L^*}{p(2-p)}$ and the group is standard, where L_E is as defined in Proposition 3.

Proof: see Appendix B.3.

The following Corollary shows the exact mapping between the model parameters (k_i, k_j, p), the loan terms (L^*, R^*) and the chosen group form.

Corollary B. *Let $R_E \equiv \frac{L_E}{p(2-p)}$, $\hat{k} \equiv \frac{2-3p}{p(2-p)(2p-1)}$, $\tilde{k} \equiv \frac{1}{p(2p-1)}$, $d(k) \equiv \frac{2}{p(2-p)} + k$, $g(k) \equiv \frac{1}{p^2} + \frac{1-p}{p}k$ and $\bar{k}^I$ be the larger root of $\bar{W}^S(L_E, R_E) = \bar{W}^B(L_{B'}, R_{B'})$ written as quadratic equation in k_h.[37] Calling $\mathcal{S}' \equiv (L_{S'}, R_{S'})$, $\mathcal{B}' \equiv (L_{B'}, R_{B'})$ and $\mathcal{E} = (L_E, R_E)$, the loan terms for all possible group types and parameter values are:*

Parameter conditions	**Loan terms and group form**		
	ll pairs	hh pairs	hl pairs
(a) high productivities or high p: $p(2p-1)k_l > 1$	*$\mathcal{B}'$, bogus*	*$\mathcal{B}'$, bogus*	*$\mathcal{B}'$, bogus*
(b) high k_h and low k_l: $p(2p-1)k_h > 1 \geq p(2p-1)k_l$	*$\mathcal{S}'$, standard*	*$\mathcal{B}'$, bogus*	*$\mathcal{B}'$, bogus*
(c) low productivities or low p: $p(2p-1)k_h \leq 1$ and			
i) $k_h \in [k_l, \min\{g(k_l), d(k_l)\}]$	*$\mathcal{S}'$, standard*	*$\mathcal{S}'$, standard*	*$\mathcal{S}'$, standard*
ii) $k_h \in (d(k_l), \bar{k}^I] \wedge \{(k_l < \hat{k} \wedge p \in (\frac{1}{2}, \frac{4}{7})) \vee p \leq \frac{1}{2}\}$	*$\mathcal{S}'$, standard*	*$\mathcal{S}'$, standard*	*$\mathcal{E}$, standard*
iii-1) $k_h > g(k_l) \wedge k_l \geq \hat{k}$	*$\mathcal{S}'$, standard*	*$\mathcal{S}'$, standard*	*$\mathcal{B}'$, bogus*
iii-2) $k_h > \bar{k}^I \wedge \{(k_l < \hat{k} \wedge p \in (\frac{1}{2}, \frac{4}{7})) \vee p \leq \frac{1}{2}\}$	*$\mathcal{S}'$, standard*	*$\mathcal{S}'$, standard*	*$\mathcal{B}'$, bogus*

Proof: see Appendix B.3.

B.3 Proofs and details

(Repay, Default) equilibrium

In general, all pure strategy Nash equilibria of the stage 1 game described in Section B.1 for given (L, R) are: (i) (Repay, Repay) if $R \leq \frac{1-p}{2-p}V$; (ii) (Repay, Default) or (Default, Repay) if $\frac{1-p}{2-p}V < R \leq \frac{1}{2}V$; and (iii) (Default, Default) if $R > \frac{1}{2}V$. In the main text we focus on the (Repay, Repay) equilibrium. A lender could potentially offer a loan with $\frac{1-p}{2-p}V < R \leq \frac{1}{2}V$ inducing the (Repay, Default) outcome it stage 1 and break

[36] The exact conditions are shown in Corrolary B.
[37] The threshold $\bar{k}^I$ depends on k_l and p – see the proof of Proposition B-3 for details.

even by setting $L = pR$. However, we prove that it is never optimal for the lender to induce this equilibrium for standard groups. Indeed, note that in a (Repay, Default) equilibrium, switching from standard to bogus group does not affect the repayment probability (it is p in both cases), the interest rate ($1/p$ in both), and the expected continuation value ($2pV$ in both). However, forming a bogus group raises expected output from $p(k_i + k_j)L$ to $2pk_iL$, while supporting the same maximum loan size $\frac{pV}{2}$. Therefore, it is always (weakly) better for the borrowers to form a bogus group if facing loan terms inducing (Repay, Default) equilibrium. This implies that our focus on the (Repay, Repay) equilibrium in Section B.1 is not restrictive since the maximum payoff in (Repay, Default) equilibrium is always weakly lower than the bogus group payoff at loan terms (L_B, R_B).[38]

Proof of Proposition B-2:

(a) Consider a heterogeneous (hl) borrower pair, that is, $k_i = k_h$ and $k_j = k_l$. Substituting in the loan terms $(L_{S'}, R_{S'})$ derived in Proposition B-1, it is easy to verify that condition (5) from Lemma 1 implies that bogus groups are payoff-maximizing if and only if inequality (56) in part (a) holds. Otherwise, Lemma 1 implies that forming standard group is optimal.

(b) Using Lemma 1, since for homogeneous pairs ($k_i = k_j$) the l.h.s. of (5) is zero while its r.h.s. is positive at $(L_{S'}, R_{S'})$, it is always payoff-maximizing for a homogeneous pair to form a standard group.

(c) Shown analogously to Proposition 2(c).■

Proof of Proposition B-3:

As in Section 4.1, the lender chooses loan terms (L, R), contingent on the observed pair type ij (omitted to save on notation), to maximize the group payoff subject to incentive compatibility, no-default and break-even constraints:

$$\max_{L,R,\tau\in\{0,1\}} \tau\bar{W}^S(L,R) + (1-\tau)\bar{W}^B(L,R) \tag{OP}$$

$$\text{s.t. } \tau\bar{W}^S(L,R) + (1-\tau)\bar{W}^B(L,R) \geq \tau\bar{W}^B(L,R) + (1-\tau)\bar{W}^S(L,R) \quad \text{(IC)}$$

$$R \leq \tau\frac{(1-p)V}{2-p} + (1-\tau)\frac{V}{2} \quad \text{(no default)}$$

$$R = \tau\frac{L}{p(2-p)} + (1-\tau)\frac{L}{p} \quad \text{(break even)}$$

and where the payoffs for a standard group, $\bar{W}^S(L,R)$ and bogus group, $\bar{W}^B(L,R)$ are:[39]

$$\bar{W}^S(L,R) = p(k_i + k_j)L - 2p(2-p)R + 2p(2-p)V \tag{57}$$

and

$$\bar{W}^B(L,R) = 2pk_iL - 2pR + 2pV \tag{58}$$

Above we used the result that, for any k_i, k_j, p, the (Repay, Default) equilibrium in a standard group is

[38] Additionally, supporting the (Repay, Default) equilibrium may be hard in practice – it either means that one borrower is allowed to consistently free ride on the repayment decision of the other, or that the borrowers take turns, which requires coordination and commitment at odds with the non-cooperative setting.

[39] We exhibit the total payoff under all possible Nash equilibria for standard groups even though the (Repay, Default) and (Default, Default) equilibria cannot obtain at optimum. The reason is that the hidden action of borrowers choosing the group form requires evaluating deviations from the prescribed behaviour.

always weakly dominated by the (Repay) equilibrium in a bogus group – see Appendix B.1.[40] Hence, we write the no-default and break-even constraints in problem (OP) only for the (Repay, Repay) equilibrium whenever the lender wants to induce a standard group ($\tau = 1$). As before, the break-even constraint must hold at equality – if the lender made positive profit, then L can be increased or R reduced to increase the objective function. Constraint (IC) has the same interpretation as in Section 4.1.

To prove Proposition B-3 we will use two auxiliary results.

Remark 1: Comparing the payoffs $\bar{W}^S(L,R)$ and $\bar{W}^B(L,R)$ from (57) and (58) for given loan terms (L,R) with $R \leq \frac{(1-p)V}{2-p}$, an ij pair would optimally choose form bogus group if and only if,

$$(k_i - k_j)L > 2(1-p)(V-R) \tag{BC}$$

and form standard group otherwise.

Remark 2: It is easy to see from the expressions for $\bar{W}^S(L,R)$ and $\bar{W}^B(L,R)$ that forming a bogus group is always optimal for any pair ij facing loan terms (L,R) with $R > \frac{(1-p)V}{2-p}$.

(a) Homogeneous groups

Suppose the lender faces a homogeneous borrower pair (ii with $i = l$ or h) and wants to induce a standard group (choose $\tau = 1$). The break-even constraint implies $R = \frac{L}{p(2-p)}$, which substituted into the no-default constraint yields $L \leq p(1-p)V$. Constraint (IC) is satisfied for any R satisfying the no-default constraint $R \leq \frac{(1-p)V}{2-p}$ since condition (BC) does not hold for $k_i = k_j$ (see *Remark 1*). The objective $\bar{W}^S(L, \frac{L}{p(2-p)})$ equals $(2pk_i - 2)L + 2p(2-p)V$, which is strictly increasing in L. Consequently, the no-default constraint, $L \leq p(1-p)V$ must bind at optimum, implying $L^* = L_{S'} = p(1-p)V$ and $R^* = R_{S'} = \frac{(1-p)V}{2-p}$.

Suppose now the lender wants to induce a bogus group (choose $\tau = 0$). The break-even constraint implies $L = pR$, which substituted into the no-default constraint gives $L \leq \frac{pV}{2}$. The objective $\bar{W}^B(L, \frac{L}{p})$ equals $(2pk_i - 2)L + 2pV$ which is strictly increasing in L. Hence, as long as (IC) is not violated, it is optimal to have the no-default constraint bind and so $L^* = \frac{pV}{2} = L_{B'}$ and $R^* = \frac{V}{2} = R_{B'}$. Constraint (IC) is indeed not violated at $(L_{B'}, R_{B'})$ since forming a bogus group is optimal at $R = R_{B'} = \frac{V}{2} > \frac{(1-p)V}{2-p}$ (see *Remark 2*).

To decide whether choosing $\tau = 0$ or $\tau = 1$ is optimal, the lender compares the payoffs from inducing a standard group with loan terms $(L_{S'}, R_{S'})$ vs. bogus group with loan terms $(L_{B'}, R_{B'})$. Choosing $\tau = 1$, $L^* = L_{S'}$ and $R^* = R_{S'}$ is optimal when $\bar{W}^S(L_{S'}, R_{S'}) \geq \bar{W}^B(L_{B'}, R_{B'})$ which is equivalent to:

$$(2pk_i - 2)L_S + 2(2-p)pV \geq (2pk_i - 2)L_B + 2pV \text{ or,}$$
$$p(2p-1)k_i \leq 1.$$

Choosing $\tau = 0$, $L^* = L_{B'}$ and $R^* = R_{B'}$ is optimal otherwise.

(b) Heterogeneous groups

Suppose the lender faces a heterogeneous (k_h, k_l) borrower pair and considers inducing standard group ($\tau = 1$). The break-even constraint implies $R = \frac{L}{p(2-p)}$, which substituted into the no-default constraint

[40]Obviously, setting (L,R) to induce (Default, Default) in a standard group or (Default) in a bogus group is not compatible with the break-even condition, so it is not optimal either.

implies $L \leq p(1-p)V = L_{S'}$. Unlike for homogeneous pairs, constraint (IC) is not automatically satisfied for any R satisfying the no-default condition. Indeed, for (IC) to hold, that is $\bar{W}^S(L,R) \geq \bar{W}^B(L,R)$, using (BC) we need,

$$p(k_h - k_l)L \leq 2p(1-p)(V - \tfrac{L}{p(2-p)}) \text{ or,}$$
$$L \leq \frac{p(2-p)V}{1+\frac{p(2-p)}{2(1-p)}(k_h-k_l)} = L_E$$

Therefore, the optimal loan terms in this case are $L_N \equiv \min\{L_E, L_{S'}\}$ and $R_N \equiv \frac{L^*}{p(2-p)}$.

Suppose instead the lender wants to induce a bogus group ($\tau = 0$). The analysis is analogous to that for homogeneous pairs. Constraint (IC) is satisfied at $L = \frac{pV}{2}$ and $R = \frac{V}{2}$. Therefore, the optimal loan terms in this case are $L^* = L_{B'}$ and $R^* = R_{B'}$, as before.

To choose $\tau = 0$ or $\tau = 1$ the lender compares the group payoff from inducing a standard group with loan terms (L_N, R_N) vs. a bogus group with loan terms $(L_{B'}, R_{B'})$. There are two cases, depending on whether L_E or $L_{S'}$ is smaller. $L_E \geq L_{S'}$ is equivalent to

$$\frac{p(1-p)V}{\frac{1-p}{2-p}+\frac{p}{2}(k_h-k_l)} \geq p(1-p)V \iff \frac{2}{p(2-p)} \geq k_h - k_l \tag{CC1}$$

which is the converse of the condition in Proposition B-2.

Suppose first that condition (CC1) is satisfied, that is

$$k_h \leq d(k_l) = k_l + \frac{2}{p(2-p)}$$

and hence $L_N = L_{S'} \leq L_E$. Then, choosing $\tau = 1$, $L^* = L_{S'}$ and $R^* = R_{S'}$ is optimal if $\bar{W}^S(L_{S'}, R_{S'}) \geq \bar{W}^B(L_{B'}, R_{B'})$ which is equivalent to

$$(p(k_l + k_h) - 2)L_S + 2(2-p)pV \geq (2pk_h - 2)L_B + 2pV \text{ or,} \tag{CC2}$$
$$k_h \leq g(k_l) \equiv \frac{1}{p^2} + \frac{1-p}{p}k_l$$

and $\tau = 0$, $L^* = L_{B'}$ and $R^* = R_{B'}$ is optimal otherwise.

If inequality (CC1) does not hold, that is if $k_h > d(k_l)$, then $L_N = L_E < L_{S'}$ and $R_N = R_E \equiv \frac{L_E}{p(2-p)}$. Choosing $\tau = 1$, $L^* = L_N$ and $R^* = R_N$ is optimal if $\bar{W}^S(L_E, R_E) \geq \bar{W}^B(L_{B'}, R_{B'})$ which is equivalent to

$$(p(k_l+k_h)-2)\frac{p(1-p)V}{\frac{1-p}{2-p}+\frac{p}{2}(k_h-k_l)} + 2(2-p)pV \geq (2pk_h - 2)\frac{pV}{2} + 2pV \text{ or } \bar{f}(k_l) \geq k_h \tag{CC3}$$

where[41]

$$\bar{f}(k) \equiv \frac{1}{2}\left(k + c + \sqrt{(k+c)^2 - \frac{8(1-p)}{p^2(2-p)} - \frac{4k}{p}}\right) \text{ and } c \equiv \frac{4p^2-11p+8}{p(2-p)}.$$

Choosing $\tau = 0$, $L^* = L_{B'}$ and $R^* = R_{B'}$ is optimal otherwise, for $\bar{f}(k_l) < k_h$.

Remark 3. Note that, since $L_E < L_{S'}$, then $k_h > d(k_l)$ implies $\bar{W}^S(L_{S'}, R_{S'}) > \bar{W}^S(L_E, R_E)$, which

[41] One can show that the smaller root of the quadratic equation of which $\bar{f}(k)$ is the larger root is strictly smaller than k. Thus, for $k_h \geq k_l$ inequality (CC3) is satisfied iff $k_h \leq \bar{f}(k_l)$.

means that condition (CC3) is tighter than (implies) condition (CC2).

In sum, we obtain:

(a) if $k_h \leq d(k_l)$ then it is optimal to offer loan terms $\mathcal{S}'$ if $k_h \leq g(k_l)$ and $\mathcal{B}'$ if $k_h > g(k_l)$

(b) if $k_h > d(k_l)$ then it is optimal to offer loan terms $\mathcal{E}$ if $k_h \leq \bar{f}(k_l)$ and $\mathcal{B}'$ if $k_h > \bar{f}(k_l)$.■

Proof of Corollary B

The results for homogeneous groups (Table columns ll and hh) depend only on whether $p(2p-1)k_i > 1$ or $p(2p-1)k_i \leq 1$ and are implied directly by Proposition B-3(a). Hence, we only discuss heterogeneous groups below. Call $\hat{k} \equiv \frac{2-3p}{p(2-p)(2p-1)}$ and $\tilde{k} \equiv \frac{1}{p(2p-1)}$, defined for $p > 1/2$. The following results are easily shown (all functions are defined in the proof of Proposition B-3).

Result 1. Suppose $p \leq 1/2$. Then, $(2p-1)\hat{k} \leq 0$, $d(k) < g(k)$ and $d(k) < \bar{f}(k)$ for any $k \geq 0$.

Proof: The first two statements are easy to check directly. To prove that $d(k) < \bar{f}(k)$, rewrite $k + c$ in the definition of $\bar{f}(k)$ as $d(k) + n$ where $n \equiv \frac{3}{p} - 4$. Then, $d(k) < \bar{f}(k)$ is equivalent to:

$$d(k) < \frac{1}{2}\left(d(k) + n + \sqrt{(d(k)+n)^2 - \frac{4k}{p} - \frac{8(1-p)}{p^2(2-p)}}\right)$$

or $d(k) - n < \sqrt{(d(k)+n)^2 - \frac{4k}{p} - \frac{8(1-p)}{p^2(2-p)}}$, which is equivalent to $\frac{k}{p} + \frac{2(1-p)}{p^2(2-p)} < d(k)n = (k_{\min}+k)(\frac{3}{p} - 4)$, or $2k(2p-1) < k_{\min}(2-3p)$ which is true for $p \leq 1/2$.

Result 2. Suppose $p > 1/2$. Then $d(\hat{k}) = g(\hat{k}) = \bar{f}(\hat{k})$.

Proof: $d(\hat{k}) = g(\hat{k})$ is verified directly. To show that $d(\hat{k}) = \bar{f}(\hat{k})$, follow the proof of *Result 1* above and use $k_{\min} = \frac{2}{p(2-p)}$ in the last step.

Result 3. Suppose $p > 1/2$. Then $\hat{k} \leq k_{\min}$ if $p \geq 4/7$ and $\hat{k} > k_{\min}$ if $p \in (1/2, 4/7)$.

Result 4. Suppose $p > 1/2$. If $k \geq \hat{k}$, then $d(k) \geq \bar{f}(k)$ and $d(k) \geq g(k)$. If $k < \hat{k}$, then $d(k) < \bar{f}(k)$ and $d(k) < g(k)$.

Result 5. Suppose $p > 1/2$. Then, $k > \tilde{k} \Longleftrightarrow k > g(k)$.

Result 6. Suppose $p > 1/2$. Then, $\tilde{k} > \hat{k}$ and $\tilde{k} < k_{\min} \Leftrightarrow p > 4/5$.

We proceed with the proof of Corollary B. Parts (a)–(c) below refer to the corresponding Table lines in the corollary statement.

(a) The condition $p(2p-1)k_l > 1$ (equivalent to $k_l > \tilde{k}$) can only hold for $p > 1/2$. Note also that if $p > 4/5$ then $\tilde{k} < k_{\min}$ by *Result 6* and hence $p(2p-1)k_l > 1$ is satisfied for any $k_l \geq k_{\min}$. Using *Result 5*, $k_l > \tilde{k}$ implies $k_h \geq k_l > g(k_l)$, and hence, by (CC2), it is optimal to offer loan terms $\mathcal{B}'$ for any such k_h, k_l.

(b) As explained in (a), this case is impossible for $p > 4/5$. It is also impossible for $p \leq 1/2$ because $p(2p-1)k_h > 1$ cannot hold. Assuming $p \in (\frac{1}{2}, \frac{4}{5}]$, the inequality $k_h > \tilde{k}$ implies (by *Result 5*) $k_h > g(k_h) \geq g(k_l)$, that is, $k_h > g(k_l)$ for any such k_h, k_l. Note also that, in the case $k_h > d(k_l)$, the inequality $k_h > g(k_l)$ (the negation of CC2) implies $k_h > \bar{f}(k_l)$, which is the negation of the stricter inequality (CC3). Using Proposition B-3(b), in either of these cases it is optimal to offer loan terms $\mathcal{B}'$ to hl pairs for any such k_h, k_l.

(c) The following sub-cases depending on the value of p are possible:

(c-i) Suppose $p \leq 1/2$, in which case $p(2p-1)k_h \leq 1$ holds for any $k_h \geq k_{\min}$. Then, by

Result 1 and since $d(k) > k$ for all $k > 0$, using Proposition B-3(b), the lender offers loan terms $\mathcal{S}'$ if $k_h \in [k_l, d(k_l)]$, loan terms $\mathcal{E}$ if $k_h \in (d(k_l), \bar{f}(k_l)]$, and loan terms $\mathcal{B}'$ if $k_h > \bar{f}(k_l)$. Calling $\bar{k}^I \equiv \bar{f}(k_l)$, these cases correspond to lines (i), (ii) and (iii-2) in part (c) of the Table.

(c-ii) Suppose $p \in (\frac{1}{2}, \frac{4}{7})$. Then, by *Result 3*, $\hat{k} > k_{\min}$. If $k_l \in [k_{\min}, \hat{k})$, then from *Result 4*, $d(k_l) < \bar{f}(k_l)$ and $d(k_l) < g(k_l)$. Hence, using Proposition B-3(b), the lender offers loan terms $\mathcal{S}'$ if $k_h \in [k_l, d(k_l)]$, loan terms $\mathcal{E}$ if $k_h \in (d(k_l), \bar{f}(k_l)]$, and loan terms $\mathcal{B}'$ if $k_h > \bar{f}(k_l)$. These cases map to lines (i), (ii) and (iii-2) in part (c) of the Table. Alternatively, if $k_l \geq \hat{k}$, then $d(k_l) \geq \bar{f}(k_l)$ and $d(k_l) \geq g(k_l)$ by *Result 4*, and so it is optimal to offer $\mathcal{S}'$ for $k_h \in [k_l, g(k_l)]$ and offer $\mathcal{B}'$ for $k_h > g(k_l)$. These correspond to lines (i) and (iii-1) in the Table.

(c-iii) Suppose $p \in [\frac{4}{7}, \frac{4}{5}]$. By *Result 3*, $\hat{k} \leq k_{\min}$ and so $k_l \geq \hat{k}$. Thus, by *Result 4*, $d(k_l) \geq \bar{f}(k_l)$ and $d(k_l) \geq g(k_l)$. If $k_l > \tilde{k}$, then, by *Result 5*, $k_l > g(k_l)$ and so $k_h > g(k_l)$. Hence, the lender offers loan terms $\mathcal{B}'$ (line iii-1 in the Table). If, instead, $k_l \in [k_{\min}, \tilde{k}]$ then, by *Result 5*, $g(k_l) \geq k_l$ and the lender offers loan terms $\mathcal{S}'$ for $k_h \in [k_l, g(k_l)]$ and loan terms $\mathcal{B}'$ for $k_h > g(k_l)$, corresponding to lines (i) and (iii-1) in the Table.

Finally, if $p > 4/5$, then only case (a) is possible since $p(2p-1)k_l > 1$ for any $k_l \geq k_{\min}$.□

Appendix C n-member groups

Our analysis can be extended to n-member groups where $n > 2$. As in Section 3, it is easy to show that with joint repayment decisions both bogus and standard groups repay when $R \leq V$. Denote the repayment probability of a standard group by p_S where $p_S \geq p$. The value of p_S depends on the feasibility constraint, that is, how much extra liability a member is able to cover when her partner(s) fail(s). For example, suppose that a borrower can repay up to mR in total liability when other members fail, where $m \leq n$. In our baseline model we assumed $m = n = 2$, and so $p_S = 1 - (1-p)^2 = p(2-p)$. In general, the expected joint payoffs in a standard and a bogus groups are:

$$W^S(L, R) = p \sum_{i=1}^{n} k_i L + np_S(V - R) \text{ and } W^B(L, R) = npk_{max}L + np(V - R),$$

where $k_{\max}$ is the largest productivity value in the group and superscripts S and B denote standard and bogus group as before. For given loan terms (L, R) a group would choose to be bogus if

$$npk_{max}L + np(V - R) \geq p \sum_{i} k_i L + np_S(V - R)$$

$$\Leftrightarrow p(k_{max} - \frac{1}{n} \sum_{i=1}^{n} k_i)L \geq (p_S - p)(V - R)$$

The left hand side reflects the expected output gain from forming a bogus group and raising the average project productivity from $\frac{1}{n} \sum_{i=1}^{n} k_i$ to $k_{\max}$. The right hand side corresponds to the net gain from forming a standard group and raising the probability of repayment and obtaining the future value V from p to p_S. Note that for $n = 2$ and $p_S = 1 - (1-p)^2$ we obtain Lemma 1. The rest of our previous analysis can be extended for $n > 2$ in a similar way, with the complication of additional parametric cases and algebra – full details are available upon request.

Appendix D

Table A1: Sample statistics

	Bogus groups		Standard groups		Bogus minus standard	
	mean	s.d.	mean	s.d.	mean	p-value
group size	4.36	0.74	4.36	0.58	-0.01	0.97
loan size	7,225	1,498	7,064	1,743	161	0.68
required monthly repayment	729	137	706	174	22	0.55
age	43.4	5.06	44.3	4.83	-0.83	0.51
% married	94	14	96	14	-3	0.48
% intended loan use = agriculture	79	37	77	35	2	0.81
% intended loan use = consumption	4	14	0	0	4	0.19
% intended loan use = business[#]	17	32	23	35	-6	0.44
% without education	1	4	1	5	-0	0.74
% with primary school	25	29	28	30	-3	0.66
% with junior school	71	31	63	35	8	0.30
% with high school	3	1	3	9	-1	0.85
% with college or above	0	0	4	15	-4	0.03
% Hanzu	30	33	29	27	1	0.93
% Manchu	70	33	70	28	-0	0.99
% Mongol	0	3	1	4	-1	0.47
sample size	59		22			

Note: #business use = manufacturing, services, wholesale or transportation.