JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. (2012)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.2276

HOW IMPORTANT ARE ENDOGENOUS PEER EFFECTS IN GROUP LENDING? ESTIMATING A STATIC GAME OF INCOMPLETE INFORMATION

SHANJUN LI[a*], YANYAN LIU[b*] AND KLAUS DEININGER[c]

[a] *Dyson School of Applied Economics and Management, Cornell University, Ithaca, NY, USA*
[b] *International Food Policy Research Institute (IFPRI), Washington, DC, USA*
[c] *World Bank, Washington, DC, USA*

SUMMARY

We quantify the importance of endogenous peer effects in group lending programs by estimating a static game of incomplete information. Endogenous peer effects describe how one's behavior is affected by the behavior of her peers. Using a rich dataset from a group lending program in India, our empirical analysis presents a robust finding of large peer effects. The preferred model suggests that the probability of a member making a full repayment would be 12 percentage points higher if all the fellow members were to make full repayment compared with a scenario in which none of the other members repay in full. We find that peer effects would be overestimated without controlling for unobserved group heterogeneity and that inconsistencies exist in the estimated effects of other variables without modeling peer effects and unobserved heterogeneity. Copyright © 2012 John Wiley & Sons, Ltd.

Received 13 November 2009; Revised 5 February 2012;

Supporting information may be found in the online version of this article.

1. INTRODUCTION

Since the establishment of the Grameen Bank in Bangladesh in 1976, the practice of group lending has been widely adopted in microfinance programs in developing countries as an important tool to provide credit to the poor. Different from conventional individual lending, a loan in group lending (or joint liability) is granted to a group of borrowers and the whole group is liable for the debt of any individual member in the group. This practice allows the lenders to largely rely on accountability and mutual trust among group members, rather than financial collateral to insure against default. Given that the poor often do not have appropriate financial collateral to offer, group lending programs offer a feasible and even profitable channel to extend credit to the poor, who are usually kept out of traditional banking systems.

This paper provides, to our knowledge, the first structural analysis of endogenous peer effects in group lending programs. We model repayment decisions of group members using a static game of incomplete information and estimate the game based on a rich dataset from a group lending program in India. Numerous theoretical studies exist with the aim of explaining the success of group lending. Most of these studies employ a game-theoretical framework where members in a group are assumed to make their repayment decisions strategically.[1] The success of group lending has been attributed to, among other things, the ability of such groups to mitigate adverse selection and moral hazard through peer effects. Peer

* Correspondence to: Shanjun Li, Dyson School of Applied Economics and Management, Cornell University, 424 Warren Hall, Ithaca, NY 14853, USA. E-mail: sl2448@cornell.edu
* Correspondence to: Yanyan Liu, International Food Policy Research Institute, 2033 K Street N.W., Washington DC 20006, USA. E-mail: Y.Liu@cgiar.org

[1] See, for example Banerjee *et al*. (1994); Besley and Coate (1995); de Aghion (1999); Ghatak (1999); Stiglitz (1990). Morduch (1999) provides a review of theoretical perspectives and early empirical evidence. Ahlin and Townsend (2007) present an empirical test of four of the theoretical models.

effects could operate and manifest through peer selection, peer monitoring and peer pressure, all of which could mitigate the information asymmetry problem in credit markets and are less costly than the tools available to formal institutions in achieving the same goals. The process of peer selection tends to screen the more risky households out of a group lending program. Through peer monitoring, members in a group can effectively monitor others' usage of a loan and reduce *ex ante* moral hazard (e.g. risky investment). Peer pressure refers to the influence peers can exert on enforcing repayment and mitigating *ex post* moral hazard (e.g. deliberate default). The effectiveness of these channels hinges on the premise that group members living in close-knit poor communities can effectively identify, as well as punish, irresponsible members and deliberate defaulters through social penalties.

From empirical standpoints, peer effects can be categorized into endogenous peer effects and contextual peer effects depending on the channel through which peer effects operate (Manski, 2000; Brock and Durlauf, 2007). Endogenous peer effects capture the fact that one's behavior (e.g. repayment decision) could be directly affected by the behavior of her peers. Contextual peer effects relate to how characteristics of a group affect its members' decisions. The distinction of these two types of peer effects has important policy implications because endogenous peer effects give rise to 'multiplier effects' through the feedback in member behaviors whereas contextual effects do not. In the remainder of the text, we use peer effects and endogenous peer effects interchangeably and refer to contextual peer effects as just contextual effects.

Despite a rich theoretical literature, there is a lack of empirical work on microfinance, especially in examining the importance of endogenous peer effects. Although it is recognized that strategic interactions among members and endogenous peer effects could be critical elements in group lending programs, they have not been modeled explicitly in existing empirical studies. Most empirical studies treat a group as a decision maker and employ single-agent choice models such as logit and tobit. They use various indicators on group heterogeneity and social ties to capture peer effects. Sharma and Zeller (1997) study how the proportion of members in the group that are relatives affects repayment rate at the group level. The performance of group loans has also been examined against the degree of homogeneity defined over social economic characteristics and family ties by Paxton *et al*. (2000), and social ties measured in various ways by Wydick (1999) and Hemnes *et al*. (2005). Unlike the aforementioned studies, Karlan (2007) examines repayment decisions of *individual* members and uses geographic proximity to capture contextual peer effects.

The approach undertaken thus far in the empirical literature can probably be attributed to the following two reasons. First, incorporating strategic interactions into a discrete choice model is empirically challenging. As will be discussed in more detail in the model and estimation sections, it inevitably produces a nonlinear model with an endogenous variable that characterizes the repayment decisions of other members in the group. Second, data of group lending programs with detailed member information that are suitable for a game-theoretical framework are not easy to obtain.

Unlike previous empirical studies, our focus is to examine endogenous peer effects by explicitly modeling strategic interactions among group members in the repayment stage. We model repayment decisions of group members in a static game of incomplete information in which members make their repayment decisions simultaneously based on their individual characteristics (some of which are unobserved by other members), group-level characteristics, as well as their expectations of other members' repayment decisions. We estimate the game by the simulated maximum likelihood method with a nested fixed-point algorithm, which recovers equilibrium repayment probabilities for all members in the game. These repayment probabilities are then used to form the likelihood function. Our estimation strategy follows the growing literature in estimating discrete-choice games such as entry games in industrial organization where one firm's payoff from entry is affected by other firms' entry decisions.

Using a rich dataset from a group lending program in Andhra Pradesh in India, our structural estimation suggests strong endogenous peer effects in repayment decisions of program participants: the probability that a group member making a full repayment would be 12 percentage points higher

DOI: 10.1002/jae

if the member is in a group where all the fellow members repay in full than if she is in a group where no fellow members make full repayment, *ceteris paribus*. The endogenous peer effects give rise to a multiplier effect of about three based on our parameter estimates. Our empirical results also highlight the importance of explicitly modeling peer effects and controlling for unobserved group heterogeneity in empirical studies of group lending programs by showing that, without doing so, large inconsistencies could arise in the estimated effects of other variables on repayment decisions.

2. GROUP LENDING AND PEER EFFECTS

The origin of group lending can be traced back to 1976 when the 2006 Nobel Peace Prize winner, Muhammad Yunus, started the Grameen Bank Project, a lending project in several villages in Bangladesh. The goal of the project is to examine the feasibility of a credit delivery system (Grameen Bank) specifically targeted to the rural poor, who often do not have financial collateral and cannot obtain credit from conventional banks. Instead of requiring collateral, group lending employs a group-based credit approach and relies on peer effects within groups to ensure repayment. The project has achieved great success in delivering credit to the poor while attaining an almost 100% repayment rate. The achievement of Grameen Bank in Bangladesh has inspired similar endeavors in more than 40 developing countries, including one of Bangladesh's neighbors: India.

In 1992, India's National Bank for Agriculture and Rural Development organized 500 self-help groups (SHGs) composed of only women as a pilot program for delivering credit to the poor. Since then, the SHG program has witnessed tremendous growth that brought about one of the world's largest and fastest-growing networks for microfinance. In 2007, some 40 million households were organized in more than 2.8 million SHGs that borrowed more than US $1 billion of credit from banks in 2006–07 alone (Reserve Bank of India, 2008). Cumulative credit disbursed to SHGs amounted to some US $4.5 billion (or about 10% of total rural credit) in India (Garikipati, 2008).

The SHG model in India combines saving generation and micro-lending with social mobilization. In this model, women who live in the same village voluntarily form SHGs with the understanding of a joint liability mechanism. A typical SHG consists of 10–20 members who meet regularly to discuss social issues and activities and, during these meetings, deposit a small thrift payment into a joint bank account. Once enough savings have been accumulated, group members can apply for internal loans that draw on accumulated savings at an interest rate to be determined by the group. Having established a record of internal saving and repayment, the group can become eligible for loans through a commercial bank, normally at a fixed ratio (typically starting at 4:1) to its equity capital.

The microfinance groups under study are located in Andhra Pradesh in India. Besides thrift savings and obtaining credits, SHGs in Andhra Pradesh also work as local institutions that carry out implementation of government programs in a variety of areas, such as distributing subsidized rice credit, life and property insurance, and pensions. In Andhra Pradesh, banks carry out microfinance business in non-overlapping territories, so that a group can only borrow from one bank.[2] Moreover, a bank only allows a group to have one outstanding loan. Once a loan is obtained by a group, it is immediately allocated among the members (mostly on an equal basis) with the repayment terms (such as interest rate, length, number of installments, etc.) set by the bank. The group cannot obtain loans from the bank in the future until the group has fully repaid the loan.

Previous literatures have discussed several mechanisms through which peers influence a member's repayment decisions (Besley and Coate, 1995; Morduch, 1999; Karlan, 2007). Positive peer effects, which imply that a higher repayment rate of other members increases one's own repayment likelihood, can be brought about through increasing the cost of defaulting, encouraging more diligent

[2] Only when a group cannot obtain a loan from the bank specialized in the area where the group is located can the group apply for a loan from other banks. However, the chances of a group obtaining a loan from other banks are rather slim.

DOI: 10.1002/jae

work ethics, and inspiring reciprocity and solidarity within groups. Members in a group are neighbors who know each other well, so they might be able to observe each other's usage of funds and distinguish deliberate default and default due to irresponsible behaviors (such as investing in projects that are too risky, spending on alcohol and tobacco) from default due to unexpected negative shocks. The repaying members can thus impose social penalties to increase the cost of deliberate default and default due to irresponsible behaviors. Social penalties can take the forms of ostracism, not providing help in their production and other activities in the future, and so on. These penalties could be severe in close-knit poor communities where people rely on each other in their daily lives and, to an even larger extent, during times of distress. On the other hand, a member who defaults due to unexpected negative shocks is likely to be forgiven and covered by her peers, which can give her a high incentive to pay back if her situation gets better.

Nevertheless, another mechanism that has been raised in the literature (Besley and Coate, 1995) can result in negative peer effects. This mechanism suggests that some 'bad', nonpaying members may 'free ride' off good, paying members by relying on the paying members' help to repay the loan even though they have the ability to repay on their own; that is, they would repay in individual lending. The fact that the SHGs in Andhra Pradesh also serve as the organization base for programs and activities other than group lending implies that the potential social penalties can be very severe and that the free-rider problem is likely to be small: free-riders are likely to be kept out of the groups through intensive peer selection process or could be dismissed from the group at a later stage and lose access to other programs implemented by the groups.[3] Nevertheless, in our empirical estimation, we do not restrict the direction of peer effects a priori. The positive and significant peer effects found from our estimation imply that a higher repayment (default) rate of other members increases one's own repayment (default) likelihood. This finding confirms that the free-rider problem is dominated by other mechanisms, if it exists at all.

3. MODEL AND IDENTIFICATION

In this section, we first lay out a model to characterize household decisions in group lending. We then discuss the issue of identification.

3.1. Model

In a group lending program, households form groups to get loans from a lender such as a commercial bank. The loan is extended to the group and divided among members. The group as a whole is held liable should one or more members fail to make a repayment. We index group loan by g and member (i.e. household) by i. We denote the choice set of a member i in group loan g by $A_{gi} = \{0, 1\}$, where 1 represents a full repayment and 0 otherwise. Let the Cartesian product $A_g = \times_i A_{gi}$ denote the possible actions of all borrowers and define $a_g = \left(a_{g1}, a_{g2}, \ldots, a_{gN_g}\right)$ as an element in A_g, where N_g is the number of members that participated in group loan g. Let y_{gi} be the characteristics of member i and y_g denote member characteristics of all participating members. We assume that y_g is observed by all members in the group while allowing some components of y_g to be unobserved by researchers.

The utility of member i in group g after the realization of repayment decisions by all members in the loan is

[3] Olson (1965) argues that the free-riding problem can be overcome through the provision of selective benefits which are conferred only to those who contribute to the collective good. In our context, these programs in the spirit of selective benefits include the Rice Credit Line program, which provides in-kind credit for subsidized rice, a pension program, and a job training program.

DOI: 10.1002/jae

$$U_{gi}\left(a_{gi}, a_{-gi}, y_g, \varepsilon_{gi}\right) = U_{gi}\left(a_{gi}, a_{-gi}, y_g\right) + \varepsilon\left(a_{gi}\right) \tag{1}$$

where a_{gi} is the action taken by member i, and a_{-gi} is a vector of actions of other members in the same loan. $\varepsilon(a_{gi})$ is a stochastic preference shock that is i.i.d. across i and additively separable in the utility function as in a standard random utility model. We assume that $\varepsilon(a_{gi})$ is observed only by member i. The key feature of the utility function is the presence of actions taken by others in the loan, a_{-gi}. With $\varepsilon(a_{gi})$ being private information, the above model is a static game with incomplete information. A pure-strategy Bayesian Nash equilibrium in such a game is defined by $a_g^* = \left(a_{g1}^*(\varepsilon_{g1}), a_{g2}^*(\varepsilon_{g2}), \ldots, a_{gN}^*(\varepsilon_{gN})\right)$, where $a_{gi}^*(\varepsilon_{gi}) \in \{0, 1\}$ maximizes the expected utility:

$$U_{gi}\left(a_{gi}, y_g, \varepsilon_{gi}\right) = \int U_{gi}\left(a_{gi}, a_{-gi}^*(\varepsilon_{-gi}), y_g\right) f(\varepsilon_{-gi}) \mathrm{d}\varepsilon_{-gi} + \varepsilon_{gi} \tag{2}$$

for any $i \in \{1, 2, \ldots, N\}$.

We normalize the utility of member i from loan default to be zero, and assume that the normalized ε_{gi} (i.e. $\varepsilon(a_{gi} = 1) - \varepsilon(a_{gi} = 0)$ in previous notations) has a continuous distribution. We further delineate y_g and specify the utility function to take the following linear form:

$$U_{gi}\left(a_{gi} = 1, a_{-gi}, y_g, \varepsilon_{gi}\right) = \gamma \frac{1}{N_g - 1} \sum_{j \neq i} a_{gj} + x_{gi}\beta + z_g\eta + \xi_g + \varepsilon_{gi} \tag{3}$$

We denote $x_g = \{x_{g1}, x_{g2}, \ldots, x_{gN}\}$ and $y_g = \{x_g, z_g, \xi_g\}$, with each component to be discussed below. Because of the normalization, the above function captures how member i's utility differs between the two choices (repay in full and default).

The actions of other members in the loan are summarized in a single variable $\frac{1}{N_g - 1}\sum_{j \neq i} a_{gj}$, the proportion of members other than i who make full repayments. Following the literature, we use this variable to capture endogenous peer effects as it describes how one member's decision is affected by other members' decisions (Manski, 1993; Brock and Durlauf, 2001). Peer effects can arise through multiple channels, as discussed in Section 2. Since the default of member i can hinder the ability of other members to obtain credit in the future, other members may impose social penalties to member i in various forms to enforce repayment. The cost of default (or the benefit of repayment) of member i should be higher when more other members choose to repay, holding everything else constant. A positive coefficient γ implies positive peer effects, and the larger γ is, the stronger are the peer effects, *ceteris paribus*. Strong positive peer effects could mitigate the adverse effect from adverse selection and moral hazard and help explain the success of many group lending programs. In addition, stronger peer effects give rise to a large multiplier effect through the feedback in member repayment decisions and imply that policy interventions could exert a larger effect than otherwise. Although the above discussion on social penalties suggests positive peer effects and hence complementarity among group members in making full repayment, the free-rider problem as discussed in the previous section implies the opposite. In the estimation, we do not restrict the direction of the coefficient on the peer effect term, γ.

The vector x_{gi} represents observed member characteristics, while unobserved member characteristics are summarized by a scalar ε_{gi}. z_g is a vector of observed group-level characteristics, which include leave-me-out average member characteristics, $\frac{1}{N_g - 1}\sum_{j \neq i} x_{gj}$, and their quadratic terms. These variables provide contextual effects as well as help controlling for group-level unobservables.[4] ξ_g summarizes group-level characteristics that are observed by members but not by researchers. The distinction

[4] To deal with unobserved heterogeneity, Bajari *et al*. (2010) make the assumption that group unobservables have a fixed-effects presentation but is an unknown but smooth function of the observed variables. A similar assumption has been made in the literature (Newey, 1994; Papke and Wooldridge, 2008).

DOI: 10.1002/jae

between ξ_g and z_g is immaterial from group members' perspective; however, it has important implications for identification and estimation, to be discussed in the following sections.

Group-level unobservable ξ_g may capture group solidarity and reciprocity, weather conditions, risk type, and other common shocks that affect repayment. Some of these components could be the basis for group formation. It is important to incorporate these unobservables in empirical models for the following reasons. First, the information advantage of group members over outsiders, in addition to being plausible, is essential in justifying the idea of group lending (versus individual lending). More importantly, failure to control for these unobservables would lead to overestimation of the importance of peer effects in increasing the repayment rate. For example, in the event of a group's suffering from a common negative shock (e.g. an adverse weather condition), we would wrongly attribute the lowered repayment of a member to peer effects if we do not control for the common shock.

The key difference between our model and previous empirical models in the literature on group loans lies in the fact that we examine the repayment decision of individual members and explicitly incorporate (endogenous) peer effects while controlling for contextual effects. Although there exists a large theoretical literature that hypothesizes the potential of group lending in reducing information asymmetry problem through peer effects, previous empirical literature has largely focused on the repayment decision at the group level where peer effects are proxied by indicators on group heterogeneity and social ties.[5] A typical regression framework is to regress group loan outcomes on group characteristics. None of the previous empirical studies explicitly models and quantifies the importance of endogenous peer effects, which is the focus of our paper.

The above utility specification implies that when making repayment decisions each member needs to make and utilize their subjective beliefs about other members' decisions. We assume that these subjective beliefs are rational, given available information on member and group characteristics. With this assumption, we can use mathematical expectation in lieu of subjective expectation. The expected utility (or *ex ante* utility with respect to others' decisions) of member i becomes

$$\begin{aligned} U_{gi}(a_{gi}=1, y_g, \varepsilon_{gi}) &= E_{a_{-gi}}\left[U_{gi}(a_{gi}=1, a_{-gi}, y_g, \varepsilon_{gi}) \middle| y_g, \varepsilon_{gi}\right] \\ &= \gamma \frac{1}{N_g-1}\sum_{j\neq i} E(a_{gj}|y_g) + x_{gi}\beta + z_g\eta + \xi_g + \varepsilon_{gi} \\ &= \gamma \frac{1}{N_g-1}\sum_{j\neq i} \mathrm{Prob}(a_{gj}=1|y_g) + x_{gi}\beta + z_g\eta + \xi_g + \varepsilon_{gi} \end{aligned} \tag{4}$$

Therefore, household i will choose to fully repay the loan if and only if $U_{gi}(a_{gi}=1, y_g, \varepsilon_{gi}) > 0$. The optimal choice by member i implies that the *ex ante* probability of a full repayment (before the realization of private shock, ε_{gi}) is given by

$$P_{gi} = \mathrm{Prob}(a_{gi}=1|y_g) = \Phi\left(\gamma \frac{1}{N_g-1}\sum_{j\neq i} P_{gj} + x_{gi}\beta + z_g\eta + \xi_g\right) \tag{5}$$

where $\Phi()$ is the cumulative distribution of ε_{gi}.

Denote the probabilities of full repayment for all members in the group loan g by $P_g = (P_{g1}, P_{g2}, \ldots, P_{gN})$. The probabilities that are consistent with the Bayesian Nash equilibrium are therefore defined by the fixed point of the mapping $P_g = M(P_g)$. $M(.) : [0,1]^{Ng} \to [0,1]^{Ng}$ is a continuous function whose single dimension is represented by equation (5). The existence of a fixed point to the above function

[5] Hemnes and Lensink (2007) provide a good review of a series of recent empirical papers on this topic.

follows directly from Brouwer's fixed-point theorem. Nevertheless, the uniqueness of the fixed point is not guaranteed and the implication on estimation will be discussed below.

3.2. Identification

In this section, we discuss the identification of the model described by equations (4) and (5). First, we treat the group-level unobservable, ξ_g, as random effects rather than fixed effects. That is, we assume ξ_g is uncorrelated with observed variables x_{gi} and z_g. Although treating ξ_g as fixed effects would be less restrictive, Brock and Durlauf (2007) show that binary choice models with social interactions and unobserved group-level fixed effects cannot be identified under common regularity conditions. It is worth noting that, unlike the random-effects model without strategic interactions where the unobservable is assumed to be uncorrelated with *all* explanatory variables, the common unobservable ξ_g in our model is nonetheless correlated with the key explanatory variable $\sum_{j\neq i} P_{gj}$, an equilibrium outcome.

Without loss of generality, we assume $E(\xi_g)=0$. We further assume that the distribution of ξ_g belongs to the family of distributions with the scale property, i.e. $\xi_g=\sigma v_g$, where σ is a parameter to be estimated and v_g follows some known distribution. We let $m^e_{gi}=\frac{1}{N_g-1}\sum_{j\neq i} P_{gj}$ and define $\mathrm{supp}(x,z)$ as the joint support of the distribution of (x_g, z_g). We then rewrite equation (5) as

$$P_{gi}=\Phi\left(\gamma m^e_{g_i}+x_{gi}\beta+z_g\eta+\sigma v_g\right) \tag{6}$$

Following Brock and Durlauf (2001), we define the model as being globally identified if for all parameter pairs $(\gamma,\beta,\eta,\sigma)$ and $(\bar{\gamma},\bar{\beta},\bar{\eta},\bar{\sigma})$, it must be that $(\gamma,\beta,\eta,\sigma)=\left(\bar{\gamma},\bar{\beta},\bar{\eta},\bar{\sigma}\right)$ when the following equations (7) and (8) hold for any $(x_g, z_g)\in \mathrm{supp}(x,z)$:

$$\begin{aligned}&\int\left\{\Pi_{i=1}^{N_g}\left[\Phi\left(\gamma m^e_{gi}+x_{gi}\beta+z_g\eta+\sigma v_g\right)\right]^{a_{gi}}\left[1-\Phi\left(\gamma m^e_{gi}+x_{gi}\beta+z_g\eta+\sigma v_g\right)\right]^{\left(1-a_{gi}\right)}\right\}\mathrm{d}F(v)\\ &=\int\left\{\Pi_{i=1}^{N_g}\left[\Phi\left(\bar{\gamma} m^e_{gi}+x_{gi}\bar{\beta}+z_g\bar{\eta}+\bar{\sigma} v_g\right)\right]^{a_{gi}}\left[1-\Phi\left(\bar{\gamma} m^e_{gi}+x_{gi}\bar{\beta}+z_g\bar{\eta}+\bar{\sigma} v_g\right)\right]^{\left(1-a_{gi}\right)}\right\}\mathrm{d}F(v)\end{aligned}$$

for any $a_g\in A_g$, where $F(v)$ is the c.d.f. of v_g, and

$$\begin{aligned}m^e_{gi}&=\frac{1}{N_g-1}\sum_{j\neq i}\Phi\left(\gamma m^e_{gj}+x_{gi}\beta+z_g\eta+\sigma v_g\right)\\ &=\frac{1}{N_g-1}\sum_{j\neq i}\Phi\left(\bar{\gamma} m^e_{gj}+x_{gi}\bar{\beta}+z_g\bar{\eta}+\bar{\sigma} v_g\right)\end{aligned}$$

for any $i=1,2,\ldots,N_g$.

Without group-level unobservables, m^e_{gi} can be taken as given for the purpose of identification because it is a function of observed variables only, as is done in Brock and Durlauf (2001). If we were to take m^e_{gi} as given in our model, the identification problem boils down to a regular random-effects discrete-choice model which is well established (Chamberlain, 1980; Heckman, 1981). However, from equation (5), m^e_{gi} is a function of x_g, z_g, as well as v_g, which is unobservable to researchers. Because m^e_{gi} does not have a closed-form solution and is a function of v_g, formally deriving sufficient identification conditions from equation (7) is difficult, even after we impose specific distributions for v_g and ε_{gi}. Therefore, we rely on Monte Carlo simulations in Section 5.2 to inform identification of our

DOI: 10.1002/jae

model as well as the performance of our estimation method described in Section 5.1. The simulations provide very positive evidence on both fronts.

To intuitively understand the identification of peer effects coefficient γ, imagine that there are two otherwise identical households except that they are in different group loans, and further assume there are no common unobservables. The repayment decisions of the two households may be different solely due to the difference in expected repayment decisions of their peers in the two loans. Intuitively, peer effects are identified from the difference in the two households' repayment decisions in relation to the difference in their peers' expected repayment decisions. In the presence of common unobservables, the group loans with members who make different repayment decisions are essential because the model would not be identified without these loans. For example, drivers of group members making the same decisions in the data could be very large positive shocks for the groups with full repayment or very large negative shocks for the groups that default, even for different sets of parameters. The identification of the distribution of group-level unobservable is based on within-group correlation of repayment decisions among members. A strong correlation (beyond what is implied by common observables) implies a large dispersion of group-level unobservables and vice versa.

4. DATA

Our data are mainly from an SHG survey conducted from August to October 2006 in Andhra Pradesh in India. The survey covers more than 3000 SHGs from eight districts which were chosen to represent the state's three macro-regions (Rayalaseema, Telangana, and Coastal AP). The survey contains demographic characteristics of group members (i.e. households) such as caste, occupation, housing condition, land and livestock ownership, and education background. More importantly, the survey recorded from SHG account books the information on all loans that were taken between June 2003 and June 2006. The information includes the terms of each loan and how much a loan had been repaid by each member at the time the survey was conducted. After the survey was done, enumerators obtained each member's poverty category ('very poor', 'poor', and 'middle class') from a census dataset according to household location and name.[6]

In this analysis, we investigate 'expired' group loans from commercial banks which had passed their due date by the time of the survey. For the purpose of this study, we need to identify each loan taken by a group as a whole as well as how the loan was allocated among group members. In the survey instrument, information on the loans to the group (group loans) and the allocation of loans to members (member loans) were recorded in two different sections. For member loans, we have information on loan source, borrowing date (month and year), total amount, loan terms, amount repaid, and amount currently overdue. For group loans, we have the same information as member loans in addition to the information on how many members the loan was allocated to. However, no identifiers exist to link the group loans with corresponding member loans. That is, we do not know directly which member loans belong to which group loan. If key information (borrowing date and source) on each member loan was correctly recorded, we can rely on the member loan information only to identify all member loans that belong to the same group loan because a group was not allowed to borrow two loans at the same time from commercial banks. However, enumerators may have made mistakes when copying information from account books or at the data entry stage. If a member loan was missing or any

[6] The census dataset contains household poverty status which incorporates both economic and social status. The households defined as '*very poor*' are those who can eat only when they get work and who lack shelter, proper clothing, respect in society, and cannot send their children to school. The '*poor*' have no land, live on daily wages, and need to send school-going children to work in times of crisis. The '*middle class*' have some land and proper shelter, are recognized in society, have access to bank credit as well as public services and can send their children to schools.

DOI: 10.1002/jae

information on borrowing date or source was wrongly recorded, we cannot correctly identify a group loan based on the information of member loans only.

To deal with this issue, we match the data from the two sections (group loan and member loan) based on borrowing date, source, and total amount of each group loan. We require that the date, source, and loan amount be correctly recorded for each member in the same group in order to obtain a match. Out of the 1916 'expired' loans from commercial banks, 1632 can be matched. We drop the 284 group loans that cannot be matched. In addition, the enumerators could not obtain the poverty category for all members in 513 groups who have 625 'expired' bank loans. Thus these loans are also dropped. The remaining 1007 group loans from 814 groups form the basis for our analysis.[7]

Because the loan data were copied from account books, we are confident that the 284 no-matching cases are due to data entry errors in account books or made by enumerators. On the other hand, the issue of missing poverty category occurring in the 625 loans is due either to the failure to match the SHG survey data with the census data or to errors by enumerators. We believe that the problems in both cases are unlikely to be correlated with some unobserved factors underlying repayment decisions. We formally investigate whether the two data selection procedures would introduce bias into our estimation results. First, in one of the robustness checks discussed in Section 6.2, we add the 625 loans (dropped due to missing poverty status) to the 1007 loans for estimation. The estimation results using the larger sample are very similar to those from the 1007 loans. Second, we compare member characteristics between those in the 284 loans (dropped due to non-matching) to the members in the other 1632 loans. We find member characteristics to be similar across these two groups as shown in supporting information Table A.1 in the online Appendix.

We next briefly describe the basic characteristics of the 814 groups and 1007 group loans. These groups have 12 members on average, with the smallest group having 7 members and the largest having 20 members. The groups are from all of the three macro-regions in the province: about 26% of the groups are from Telangana, 38% from Rayalaseema, and 36% from Coastal AP. We define *default* as failure to make a full repayment at the survey time if the loan was past due by then. Among the 1007 expired group loans, 76% were fully repaid by all members to whom the loan was allocated, 7.5% were fully repaid by some members but defaulted by others, and 16.3% were defaulted by all members. The average loan size was 34,000 rupees (about US $682) and a loan was allocated to 11 members on average.

The 1007 group loans were allocated to a total of 11,037 member loans overall. Panel 1 of Table I presents summary statistics for member characteristics and panel 2 summarizes terms of the 11,037 member loans. Approximately 30% of the members belong to a scheduled tribe or a scheduled caste and 24% are literate. About 6.3% are disabled or have family members who are disabled. About 26% of the members are from very poor households, 54% from poor households, and 20% from middle-class households. About 60% households own some land and 43% own some livestock. Thirty-eight percent live in pucca houses, 37% in semi-pucca houses, and 25%t in kutcha houses.[8] About 62% are agricultural laborers who do not own land or own such small amount of land that they have to provide agricultural labor for others; 15% are self-employed agricultural workers; and 23% have other occupations (such as those self-employed in non-agricultural sectors, employed in non-agricultural sectors and housewives). The data show that most SHG members are from poor and vulnerable households. This is in line with the program's goal to target the rural poor. The average size

[7] We have no missing individual data for the loans in the final sample for analysis. If the initial sample size were small or the data were of lower quality, e.g. missing data issue exists for most of the observations, one might need to only drop individual observations with missing data and keep all other observations within the same loan. In that case, one would face the selection problem discussed in Ammermueller and Pischke (2009) and Sojourner (2009).

[8] A pucca house has walls and roof made of materials such as burnt bricks, stones, cement concrete, and timber while a kutcha house uses less sophisticated material such as hays, bamboos, mud, and grass. A semi-pucca house uses a combination of materials for the other two types.

DOI: 10.1002/jae

Table I. Summary statistics

Variable	Mean	SD	Min.	Max.
Member characteristics (11)				
If belonging to scheduled tribe	0.296	0.457	0.00	1.00
If literate	0.239	0.427	0.00	1.00
If any household member disabled	0.063	0.244	0.00	1.00
If being poorest family	0.261	0.439	0.00	1.00
If not being poor	0.195	0.396	0.00	1.00
If owning land	0.599	0.490	0.00	1.00
If owning livestock	0.427	0.495	0.00	1.00
If living in pucca house	0.378	0.485	0.00	1.00
If living in kacha house	0.247	0.432	0.00	1.00
If self-employed agric. worker	0.152	0.359	0.00	1.00
If agric. laborer	0.623	0.485	0.00	1.00
Loan characteristics (8)				
Amount of loan (10,000 rupees)	0.311	0.261	0.04	4.00
Annual rate of interest	12.476	2.503	0.00	25.00
Length of loan (year)	1.076	0.414	0.06	2.50
Repayment frequency at least monthly	0.975	0.156	0.00	1.00
If due in 2005	0.491	0.500	0.00	1.00
If due in 2006	0.419	0.493	0.00	1.00
Located in Telangana	0.265	0.441	0.00	1.00
Located in Rayalaseema	0.312	0.463	0.00	1.00
Group characteristics (24)				
Number of members/10	1.221	0.294	0.20	2.00
(Number of members/10) squared	1.576	0.679	0.04	4.00
(Leave-me-out) average member char. (11)				
Quadratic of average char. (11)				

Note: The number of observations (loan member) is 11,037. There are 1007 loans: 768 were paid in full by all members (8485 observations for these loans); 164 were defaulted by all members (1817 observations); and the remaining 75 we paid in full by some but not all members (735 observations). The base group for poverty status is being poor (the middle group). The base group for housing type is semi-pucca house (the middle group). The base group for employment status is other occupations. The base group for the due date is 2004.

of member loans was 3110 rupees (about US \$62). The average annual rate of interest was 12.5%, which is much lower than the prevailing rate of moneylenders in India. The average duration of a loan was about one year. The majority of loans (97.5%) required the groups to make a repayment at least monthly, if not more frequently.

5. EMPIRICAL STRATEGY AND MONTE CARLO EVIDENCE

In this section, we first lay out our estimation strategy. We then present a Monte Carlo study to understand model identification and the performance of the estimation method.

5.1. Estimation Strategy

To take our empirical model to the data, we assume the normalized preference shock ε_{gi} to have a logistic distribution. Normalizing the utility from loan default to zero, the cumulative distribution function $\Phi()$ in equation (5) becomes the logistic function:

$$P_{gi} = \text{Prob}\left(a_{gi} = 1 \middle| x_g, z_g, \xi_g\right) = \frac{exp\left(\gamma \frac{1}{N_g - 1} \sum_{j \neq i} P_{gj} + x_{gi}\beta + z_g\eta + \xi_g\right)}{1 + exp\left(\gamma \frac{1}{N_g - 1} \sum_{j \neq i} P_{gj} + x_{gi}\beta + z_g\eta + \xi_g\right)} \tag{7}$$

There are two challenges in taking the choice probabilities defined by equation (7) to the data. The first one is fundamental in incorporating strategic interactions into the discrete-choice model. It lies in the fact that one of the explanatory variables is unobserved, which is member i's expectation about the average repayment rate among all the fellow members $\frac{1}{N_g-1}\sum_{j\neq i} P_{gj}$. Although the observed outcome, $\frac{1}{N_g-1}\sum_{j\neq i} a_{gj}$, is a natural choice for the expectation variable, it is correlated with the individual error term, ε_{gi}. Owing to the nonlinear nature of the model, the standard instrumental variable method cannot be applied to deal with the endogeneity problem.

As discussed in the model section, we assume that common unobservable ξ_g is uncorrelated with observed variables z_g, that $\xi_g = \sigma v_g$, and that $v_g \sim N(0, 1)$ is independent and identically distributed across loans. For a given set of parameters $(\gamma, \beta, \eta, \sigma)$ and a random draw of v_g for each group loan g, a fixed point algorithm based on equation (7) can be used to recover the choice probabilities for all the members in the group. These probabilities can then be used to form the likelihood function.

The second empirical challenge arises from the possibility of multiple equilibria, which are more likely to occur when peer effects are positive and strong.[9] With multiple equilibria, the probability of an observed outcome is undefined without a specification of the equilibrium mechanism. In principle, one can compute all (and finite) fixed points to the system of equations defined by equation (7) via an all-solution homotopy method (Bajari *et al.*, 2010). The likelihood function can then be formed based on all the recovered equilibria and a specified equilibrium selection mechanism. However, this method is very computationally intensive, if not practically infeasible for our data, given the significant time needed to find all equilibria. Instead, we follow one of the approaches in the literature and assume that only one equilibrium is observed in the data if multiple equilibria do arise (Seim, 2006; Ellickson and Misra, 2008; Zhu and Singh, 2009). This approach specifies which type of equilibrium (such as random or extremal equilibria) is picked in the empirical model. Robustness analysis with respect to this assumption is discussed below.

To illustrate our estimation strategy, let o_{gi} denote the repayment outcome of gi and $\xi_g = \sigma * v_g$, where v has an independent and identically distributed standard normal distribution (across g) and σ is the standard deviation of a normal distribution. The joint probability of the observed outcome for group g conditional on a realization of v is

$$\begin{aligned} & P\left(o_{g1}, o_{g2}, \ldots, o_{gN_g} \middle| x_g, z_g, v_g\right) \\ &= P\left(a_{g1} = o_{g1} \middle| z_g, v_g\right) P\left(a_{g2} = o_{g2} \middle| x_g, z_g, v_g\right) \ldots P\left(a_{gN_g} = o_{gN_g} \middle| x_g, z_g, v_g\right) \end{aligned} \tag{8}$$

where $P(a_{gi} = o_{gi} | x_g, z_g, v_g)$ can be obtained based on the fixed points recovered from a system of N_g equations defined in (7). The averages of the joint probabilities over many draws of v_g are then used to form the likelihood function. The estimation method is simulated maximum likelihood with a nested fixed-point algorithm. For a given set of parameters and a random draw of v_g for each g, the fixed-point algorithm recovers the equilibrium repayment probabilities of all members in a group loan based on equation (7). These probabilities are then used to evaluate the likelihood function.

This empirical strategy is computationally demanding since the fixed-point algorithm has to be carried out as many times as the number of draws for v_g for each group. To reduce computational burden, we use Gauss–Hermite quadrature to approximate the joint probabilities. Because the choice probabilities can be highly nonlinear, we use as many as 64 points for the approximation of the joint probabilities, recognizing the trade-off between approximation accuracy and computational burden. Robustness checks are performed with respect to approximation and are discussed in the next section. We also check its performance against Halton draws and find quadrature to be favorable.

[9] See Brock and Durlauf (2001) and Bayer and Timmins (2007) for the equilibrium property in the context of social interactions where the reference group is large.

DOI: 10.1002/jae

5.2. Monte Carlo Evidence

In this section, we conduct Monte Carlo simulations to investigate model identification and the estimation method. To generate data, we randomly choose 504 (out of 1007) group loans from our data. We arbitrarily pick four member attributes and two loan characteristics to reduce computational burden. Based on these observed member characteristics, we generate leave-me-out average member characteristics to include as additional controls. Adding group size and a constant term, we have 12 observed variables to generate repayment outcomes.

We perform Monte Carlo simulation for two sets of parameter values as shown in Table II. We assume that the group-level unobservable has a normal distribution with mean zero in both cases. The standard deviation is assumed to be six in the first case and three in the second. In each simulation replication, we first draw group-level unobservables from the normal distribution; we use equation (7) to compute expected repayment probabilities m^e based on parameter values and observed variables. With random draws of the error term for each member, we obtain repayment decisions following equation (4). We then proceed to estimate model parameters using the simulated MLE method proposed above. We conduct 100 simulation replications for either parameter set.

The first three columns in Table II presents Monte Carlo results for the first set of parameters. The average repayment rate based on generated data (across 100 replications) is 86%. Among the 504 group loans, about 74% are fully repaid by all members and roughly 8% were defaulted by all members. These repayment outcomes are similar to what we observe in the data. All the parameters are recovered reasonably precisely: none of the estimates is statistically different from the true value at the 5% significance level and none except one is statistically different from the true value at the 10% significant level.

The last three columns in in Table II provide results for the second set of parameters, whose values are half of those in the first set. The average repayment rate from the generated data is

Table II. Monte Carlo results

	Specification 1			Specification 2		
	True	Est.	SE	True	Est.	SE
	(1)	(2)	(3)	(4)	(5)	(6)
Member characteristics						
If belonging to scheduled tribe/caste	1.2	1.276	0.056	0.6	0.630	0.036
If being poorest family	−1.4	−1.451	0.041	−0.7	−0.679	0.023
If not being poor	1.0	1.087	0.048	0.5	0.491	0.025
If owning land	−0.8	−0.791	0.032	−0.4	−0.377	0.020
Loan characteristics						
Amount of loan	−3.6	−3.581	0.139	−1.8	−1.815	0.082
Annual rate of interest	0.6	0.647	0.029	0.3	0.320	0.013
Group characteristics						
Average share of scheduled tribe	−10.0	−10.093	0.213	−5.0	−4.968	0.094
Average share of being poorest	0.8	0.867	0.109	0.4	0.350	0.065
Average share of not being poor	−5.0	−5.046	0.152	−2.5	−2.495	0.088
Average share of owning land	5.2	5.264	0.149	2.6	2.639	0.072
Group size	−3.2	−3.160	0.146	−1.6	−1.641	0.087
Constant	5.0	4.469	0.352	2.5	2.306	0.199
SD of ξ	6.0	6.108	0.151	3.0	2.950	0.055
$E(P_{-i})$	2.0	2.139	0.104	1.0	0.921	0.096

Note: The number of observations is 5436 from 504 groups randomly chosen from the full sample. Parameter estimates are the average across 100 simulation replications and standard errors are based on standard deviation of the parameter estimates from the 100 runs.

DOI: 10.1002/jae

83%. About 63% of the 504 group loans are fully repaid by all members and 3% defaulted by all members. The simulation results show that none of the parameter estimates is statistically different from the true value at any conventional significance level. These Monte Carlo simulations suggest that the empirical model is identified and that the estimation method performs well.

6. RESULTS AND ROBUSTNESS ANALYSIS

We first present estimation results for several specifications with different explanatory variables. We then show results from robustness checks and discuss two caveats of our analysis.

6.1. Results

Table III presents parameter estimates and standard errors for five specifications, adding more control variables sequentially. Because the parameter estimates cannot be compared directly across the specifications, we present the sample averages of the partial effects for selected specifications in supporting information Table A.2. The first model is a logit model with member characteristics and loan characteristics only. The coefficient estimates largely have expected signs from the standpoint of individual loans. For example, the probability of full repayment is negatively correlated with being in the scheduled caste, being the poorest family, and living in a kacha house, all of which could reflect poor economic status of a family. On the other hand, not being poor and being a self-employed agricultural worker, representing favorable economic status, are associated with higher repayment probability. In terms of loan characteristics, larger loan amount, higher interest rate, longer duration, and lower frequency are all associated with poorer repayment performance.

The second model adds 24 group-level variables including the (leave-me-out) average characteristics of the members participating in the loan and their quadratic terms (shown in Table I). These variables are used to capture contextual effects and many of them have statistically significant coefficient estimates (not shown in the table to save space). The estimated effects of most member characteristics disappear in this specification, likely due to the fact that the group characteristics are averages of member characteristics. Nonetheless, the effects of loan characteristics are largely intact. The third model controls for unobserved heterogeneity and is estimated using simulated MLE. The unobserved heterogeneity is assumed to have a normal distribution with zero mean. The standard deviation of the distribution is estimated to be quite large and statistically significant. As a result, the model fit is improved dramatically. This perhaps is not surprising given that in about 92.5% of group loans members make the same repayment choices.

The fourth and fifth specifications are the model described by equation (4) where the expected repayment rate of other members in the loan is used to capture peer effects. The estimation is done using the procedure outlined in Section 5. The fourth specification includes group size and 11 average member characteristics but not their quadratic terms, while the fifth specification, our preferred model, includes the averages and their quadratic terms. The coefficient estimate on the peer effect variable in both specifications is positive and statistically significant at any conventional level. Similar to the third specification, the coefficients on member characteristics are not statistically significant. The comparison across the five specifications suggests that the choice of modeling unobserved heterogeneity and peer effects could have important impacts on parameter estimates of member and loan characteristics. The causal interpretation of member and loan characteristics on the repayment rate hinges on the assumption that the group-level common unobservable are uncorrelated with these observed variables. A cautious approach is to interpret the estimated effects of observed household and loan characteristics on repayment decisions as correlation rather than causality, noting that our main interest in this paper is in peer effects.

DOI: 10.1002/jae

Table III. Parameter estimates from five specifications

Variables	Model 1 Est.	Model 1 SE	Model 2 Est.	Model 2 SE	Model 3 Est.	Model 3 SE	Model 4 Est.	Model 4 SE	Model 5 Est.	Model 5 SE
Member characteristics (11)										
If belonging to scheduled tribe	−0.137	0.020	−0.020	0.501	−0.365	0.918	−0.074	0.642	0.125	0.920
If literate	0.017	0.037	−0.001	0.261	−0.358	0.375	−0.464	0.332	−0.383	0.439
If member disabled	0.087	0.074	−0.086	0.448	−0.721	0.804	0.073	0.807	−0.591	0.907
If being poorest family	−0.312	0.026	−0.070	0.420	−0.282	0.545	−0.206	0.519	−0.249	0.594
If not being poor	0.314	0.034	0.101	0.300	0.208	0.531	0.369	0.461	0.340	0.538
If owning land	0.015	0.027	0.002	0.316	−0.013	0.466	−0.266	0.393	−0.252	0.504
If owning livestock	−0.023	0.031	−0.007	0.255	−0.164	0.472	0.088	0.354	−0.011	0.440
If living in pucca house	−0.043	0.029	0.017	0.312	0.561	0.683	0.448	0.370	0.504	0.520
If living in kacha house	−0.264	0.030	−0.071	0.304	−0.628	0.551	−0.518	0.422	−0.397	0.455
If self-employed agric. Worker	0.212	0.041	0.056	0.461	0.270	0.715	0.081	0.701	0.379	0.647
If agric. Labor	0.118	0.027	0.031	0.394	0.119	0.690	−0.210	0.446	0.106	0.591
Loan characteristics (8)										
Amount of loan	−1.188	0.031	−1.142	0.033	−1.597	0.595	−0.234	0.317	−0.744	0.518
Annual rate of interest	−0.064	0.003	−0.069	0.003	0.694	0.079	0.081	0.028	0.358	0.087
Length of loan (year)	−0.272	0.018	−0.261	0.020	−0.226	0.631	−1.202	0.356	−0.318	0.386
Frequency > monthly	−0.424	0.062	−0.334	0.068	0.601	1.315	1.576	0.774	0.000	0.721
If due in 2005	−0.303	0.037	−0.370	0.039	0.037	1.224	1.779	0.712	1.041	0.686
If due in 2006	−1.086	0.037	−1.161	0.040	−3.542	1.245	−1.658	0.777	−0.853	0.800
Located in Telangana	0.150	0.023	0.196	0.025	−2.743	0.766	1.746	0.449	−1.503	0.518
Located in Rayalaseema	−0.106	0.020	−0.179	0.021	−2.144	0.771	−1.358	0.342	−0.998	0.454
Constant	4.072	0.086	4.486	0.159	7.295	2.534	1.639	1.723	1.637	1.965
Group characteristics (24)	No		Yes		Yes		Half		Yes	
SD of ξ	No		No		17.824	1.335	7.716	1.408	9.179	1.924
$E(P_{-i})$	No		No		No		4.001	0.539	4.063	0.724
Log-likelihood	4996.37		4886.20		1034.50		1042.86		1024.46	
Pseudo-R^2	0.074		0.095		0.808		0.807		0.810	

Note: The number of observation is 11,037. The first specification is a logit model without group (i.e. group loan) characteristics, while the second one includes 24 group characteristics, described in Table I. The third one adds group-level unobservables and the fourth one includes peer effects. The fourth one includes group size and 11 average member characteristics but not their quadratic terms. The last one is our preferred model, with most control variables.

In supporting information Table A.2, we present two types of partial effects: direct partial effects and total partial effects averaged over the sample for specifications 4 and 5. The difference between the two is that the total partial effect incorporates the feedback/indirect effect transmitted through peer effects. That is, a change in the characteristics of household i will not only have a direct effect on its repayment propensity but also an indirect effect through its influence on other households in the loan. For instance, the direct and total partial effects of being literate are estimated at -0.006 and -0.018, respectively in the fourth specification. This means a change of household i's status from being illiterate (the base group) to being literate would be associated with an decrease in the repayment probability by 0.6%, holding everything else (including other households' repayment probabilities) constant. However, a change in household i's repayment decision would change others' repayment decisions in the same group loan due to the presence of peer effects, which would in turn affect the repayment decision of household i. The total partial effect is the partial effect when the new equilibrium has been achieved and should be larger than the direct partial effect in the presence of positive peer effects. Based on our model estimation, the total partial effect of the literacy status changing from illiterate to literate is -0.018—three times as large as the direct partial effect in magnitude. Interestingly, although the direct partial effect is not statistically different from zero, like the coefficient estimate on the literate dummy itself, the total partial effect is at the 10% level. The total

DOI: 10.1002/jae

partial effects for other variables are also about three times as large as the direct partial effects in models 4 and 5. The ratio between these two types of partial effects suggests the multiplier effect to be around three, which is dictated by the coefficient on peer effects (γ).

Table A.2 also reports the partial effect of group characteristics that capture contextual effects. All except group size are leave-me-out average member characteristics. From model 5, a larger share of other families in the group with disabled members or living in a kacha house, both representing adverse situations, is associated with a negative effect on one's repayment probability. On the other hand, a larger share of landowners in the group is associated with a positive effect on one's repayment outcome. In contrast to the free-riding concern, our results show that a bigger group is associated with higher repayment rates. As discussed in Section 2, the free-riding problem could be mitigated by the peer selection process and offset by stronger peer effects that come with larger groups, as suggested by one of the additional specifications (model s7 in supporting information Table A.3).

The (sample average) total partial effect of the peer effects variable, $E(P_{-i})$, on repayment probability is estimated to be 0.138, compared to the direct partial effect of 0.040 in specification 5, our preferred specification. To understand these estimates, imagine a group with 11 members: Members 1–10 receive positive shocks (e.g. an increase in ε_{gi}) such that each one's repayment probability increases by 0.1, holding other factors constant; therefore, for member 11, $E(P_{-11})$ increases by 0.1. The direct effect of this increase of 0.1 on member 11's repayment probability is 0.4%. This will in turn change other members' repayment decisions. Under the new equilibrium, the repayment probability of member 11 will increase by 1.38%.

In the following, we focus on treatment effects (TE) as the measure to quantify the importance of peer effects. We define TE on a given member as the difference in repayment probability between one scenario where all other members make full repayments and the other scenario where no other members repay their loan, holding other factors the same across the two scenarios. Based on the parameter estimates in the preferred model (model 5), the average repayment rate across all observations would be 74% if $E(P_{-i})=0$, compared with 85.7% if $E(P_{-i})=1$. Therefore, the treatment effect across all observations is about 11.7 percentage points, which provides a similar quantification of the importance of peer effects from total partial effects. We also examine the potential importance of peer effects in non-repaying loans where at least one member defaults. The average repayment rate among these observations would be 72.5% if $E(P_{-i})=0$, compared with 84.7% if $E(P_{-i})=1$. The TE among non-paying groups is about 12.2 percentage points—only slightly higher than the treatment effect from the full sample.

6.2. Robustness Analysis and Caveats

To check the sensitivity of our findings to model specifications and assumptions, we conduct a variety of robustness checks. Table IV presents results for five different specifications and supporting information Table A.3 provides five additional robustness checks.

The five specifications in Table IV investigate the implications of modeling contextual effects and unobserved heterogeneity. Models S4 and S5 in Table IV correspond to models 4 and 5 in Table III, while models S1 and S3 are new. Model S1 does not control for group-level characteristics, both observed and unobserved. The average treatment effect is estimated to be 0.383 across the full sample and 0.406 among non-paying groups, compared to the estimates of 0.117 and 0.122 from our preferred model S5. This finding suggests that the impact of peer effects on repayment rate would be largely overestimated if contextual effects and unobserved heterogeneity are ignored. Model S2 adds 24 group-level variables, which increased the pseudo-R^2 slightly from 0.076 to 0.095. However, the sample average treatment effect decreases substantially from 0.383 to 0.218. Model S3 controls for unobserved heterogeneity but does not include group-level characteristics. The model fit is improved

DOI: 10.1002/jae

Table IV. Sensitivity analysis and treatment effects

	Model S1		Model S2		Model S3		Model S4		Model S5	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Member characteristics (11)	Yes		Yes		Yes		Yes		Yes	
Loan characteristics (8)	Yes		Yes		Yes		Yes		Yes	
Group characteristics (24)	No		Yes		No		Half		Yes	
SD of ξ	No		No		6.623	0.774	9.541	1.662	9.146	1.713
E(P *i*)	2.024	0.176	1.279	0.199	3.662	0.486	4.001	0.539	4.063	0.724
Treatment e_ects (TE)	0.383	0.042	0.218	0.037	0.142	0.047	0.140	0.046	0.117	0.051
TE among non-repaying groups	0.406	0.041	0.218	0.037	0.145	0.048	0.150	0.048	0.122	0.053
Log-likelihood	4986.09		4884.20		1073.99		1042.86		1024.46	
Pseudo-R^2	0.076		0.095		0.801		0.807		0.810	

Note: Models S1–S5 each use a different set of explanatory variables, with models S4 and S5 corresponding to models 4 and 5 in Table III. Standard errors of treatment effects are from parametric bootstrapping.

dramatically over the first two models. The average treatment effect is estimated at 0.142. Model S4 adds 12 average member characteristics but not their quadratic terms. The average treatment effect is estimated at 0.14. These cross-specification comparisons highlight the importance of controlling for both contextual effects and unobserved heterogeneity.

The additional robustness checks presented discussed in supporting information Table A.3 confirm that our findings of peer effects from model S5 are robust to the following modeling assumptions: nonlinearity of peer effects (model S6 in Table A.3), heterogeneity of peer effects (model S7), distribution assumption on unobserved heterogeneity (model S8), quadrature approximation of repayment probabilities (model S9), group-level unobservables (model S10), and sample selection.

Two caveats are worth mentioning regarding our empirical analysis. The first one, a common challenge in empirical studies of games, concerns multiple equilibria. As discussed above, in the presence of multiple equilibria the likelihood is not well defined without an equilibrium selection mechanism. Theoretically, one can incorporate the algorithm for finding all possible equilibria in the estimation, but it is likely to be computationally prohibitive in our context. We follow the literature and assume that only one equilibrium is played in the data. In practice, we draw the starting values for the fixed-point algorithm randomly from the uniform distribution. Nevertheless, the starting values are fixed across parameter iterations. The algorithm stops once it reaches a fixed point, which is assumed to be the equilibrium played in the data. To check the robustness of the solution to the starting value (hence the equilibrium selected in the case of multiple equilibria), we re-estimate the baseline model twice with the starting value being a vector of ones as well as a vector of zeros (Zhu and Singh, 2009) and the results are almost identical to those reported in Table III.

The second caveat of our analysis is related to the choice of modeling the repayment decision in a static game. The loans are usually repaid in multiple installments where members' decisions in previous installments are observed by all. The repayment decisions should be better thought as coming out of an inherently dynamic process. However, the lack of data on each installment precludes us from estimating a dynamic game where the repayment decision of each installment could be modeled explicitly. Instead, our static model abstracts away from multiple repayment decisions made by a member for a given loan and assumes that each member makes a single repayment decision. Therefore, our model could miss potentially important elements in group loans such as peer effects in the temporal dimension (temporal peer effects). Nonetheless, to the extent members can make up their missed payments from earlier installments before the final one, repayment decisions in earlier installments might matter to a lesser extent than the decision in the final installment.

In addition, our static model is not able to examine learning effects that could exist in the process of multiple installments. One member's decisions in previous installments could reveal additional information that other members do not already know, such as current and future financial conditions of that member. The information could then affect other members' expectations on this member's next payment decisions. These two issues are different in nature and should receive different modeling treatment. Temporal peer effects imply that one's payoff (or penalty) from a certain repayment decision is affected by previous decisions of others. On the other hand, to model learning, one would need to incorporate a member's past decisions into peers' expectations of her future decisions. We believe that when appropriate data are available these should be interesting issues to examine.

7. CONCLUSION

Despite the common belief that peer effects play a significant role in group lending, for example by mitigating the moral hazard problem, how quantitatively important those effects are has remained an unanswered question. We examine the importance of endogenous peer effects by modeling members' repayment decisions as a static game of incomplete information. In our model, group members make

DOI: 10.1002/jae

their repayment decisions simultaneously based on individual characteristics, loan characteristics, group characteristics, and the expectation of other members' repayment decisions.

Using a rich dataset from a group lending program in India, we find large and positive endogenous peer effects: everything else being equal, the probability of a member's making full repayment would be 12 percentage points higher on average if all the other members in the group repay in full compared with a scenario in which none of the other members makes full repayment. The empirical finding supports the conjecture highlighted in the theoretical literature that peer effects may be an important factor behind the success of group lending programs. Unlike contextual peer effects focused by the previous literature, the strong endogenous peer effects identified in our analysis suggest a multiplier effect of about three. That is, a policy that could encourage individual repayment would have large positive spillover effects on other members. This in turn points to the importance of group formation in the early stage of a microfinance program and group development and enhancement thereafter in order to cultivate and maintain strong positive peer effects.

In addition to being able to quantify peer effects directly, our analysis demonstrates the importance of explicit modeling of strategic interactions inherent in group lending programs. We find that significant differences could emerge between our approach and traditional ones in understanding the determinants of repayment decisions. This research represents a first attempt to use a game-theoretical framework to empirically investigate the performance of group lending programs. There are many interesting questions yet to be answered that necessitate either richer data or more flexible modeling strategy. These questions include peer selection, learning, heterogeneity in peer effects, and the implication of such heterogeneity on group survival as well as on the design of group lending programs.

ACKNOWLEDGEMENTS

The authors thank Arie Beresteanu, Manuel Hernandez, LungFei Lee, Thierry Magnac (Co-Editor), two anonymous referees, and especially Carlos Martins-Filho for helpful comments and suggestions. All remaining errors are our own.

REFERENCES

Ahlin C, Townsend R. 2007. Using repayment data to test across models of joint liability lending. *The Economic Journal* **117**: F11–F51.

Ammermueller A, Pischke JS. 2009. Peer effects in European primary schools: evidence from PIRLS. *Journal of Labor Economics* **27**: 315–348.

Bajari P, Hong H, Krainer J, Nekipelov D. 2010. Estimating static models of strategic interactions. Forthcoming. *Journal of Business and Economic Statistics* **28**: 469–482.

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

Bayer P, Timmins C. 2007. Estimating equilibrium models of sorting across locations. *The Economic Journal* **117**: 353–374.

Besley T, Coate S. 1995. Group lending, repayment incentives and social collateral. *Journal of Development Economics* **46**: 1–18.

Brock WA, Durlauf SN. 2001. Discrete choice with social interactions. *Review of Economic Studies* **68**: 235–260.

Brock W, Durlauf S. 2007. Identification of binary choice models with social interactions. *Journal of Econometrics* **140**: 52–74.

Chamberlain G. 1980. Analysis of covariance with qualitative data. *Review of Economic Studies* **1**: 225–238.

de Aghion B. 1999. On the design of a credit agreement with peer monitoring. *Journal of Development Economics* **60**: 79–104.

Ellickson P, Misra S. 2008. Supermarket pricing strategies. *Marketing Science* **27**(5): 811–828.

Garikipati S. 2008. The impact of lending to women on household vulnerability and women's empowerment: evidence from India. *World Development* **36**(12): 2620–2642.

DOI: 10.1002/jae

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

Heckman J. 1981. Statistical models for discrete panel data. In *Structural Analysis of Discrete Data*, Manski C, McFadden D (eds). MIT Press: Cambridge, MA; 114–178.

Hemnes N, Lensink R. 2007. The empirics of microfinance: what do we know? *The Economic Journal* **117**: F1–F10.

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

Karlan D. 2007. Social connections and group banking. *The Economic Journal* **117**: F52–84.

Manski C. 1993. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* **60**: 531–542.

Manski C. 2000. Economic analysis of social interactions. *Journal of Economic Perspective* **14**: 115–136.

Morduch J. 1999. The microfinance promise. *Journal of Economic Literature* **37**: 1569–1614.

Newey W. 1994. The asymptotic variance of semiparametric estimators. *Econometrica* **62**: 1349–1382.

Olson M. 1965. In *The Logic of Collective Action: Public Goods and The Theory of Groups*. Harvard University Press: Cambridge, MA.

Papke L, Wooldridge J. 2008. Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics* **145**: 121–133.

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

Seim K. 2006. An empirical model of firm entry with endogenous product-type choices. *The Rand Journal of Economics* **37**(3): 619–640.

Sharma M, Zeller M. 1997. Repayment performance in group-based credit programs in Bangladesh: an empirical analysis. *World Development* **25**(10): 1731–1742.

Sojourner A. 2009. *Inference on peer effects with missing peer data: evidence from project star*. Working paper, University of Minnesota.

Stiglitz J. 1990. Peer monitoring and credit markets. *World Bank Economic Review* **4**: 351–366.

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

Zhu T, Singh V. 2009. Spatial competition with endogenous location choices: an application to discount retailing. *Quantitative Marketing and Economics* **7**: 1–35.

DOI: 10.1002/jae