Adverse selection and moral hazard in group-based lending

Mehrteab, H.T.

Publication date:
2005

Chapter 8 Group-based Lending and Adverse Selection: A Study on Risk Behavior and Group Formation[1]

8.1 Introduction

This chapter deals with group formation and the adverse selection problem. In several theoretical papers (e.g. Ghatak, 2000) it has been shown that a debt contract with a joint liability component will lead to assortative matching (homogeneous matching), implying that safe borrowers will group with safe borrowers. The risky borrowers will not be allowed to form a group with safe borrowers and hence have to rely on stand-alone debt contracts. It can be shown that the assortative matching may imply that the choice for a joint liability debt contract or a stand-alone debt contract gives the bank information on the type of borrower. This may, under certain conditions, solve the adverse-selection problem in situations of asymmetric information.

In theory therefore, joint liability debt contracts may help to solve the adverse selection problem in situations of asymmetric information. However, as is so often the case, empirical evidence on this is lagging behind. This chapter takes up the challenge to empirically contribute to the question whether joint liability lending may reduce adverse selection problems. Since it is not possible to directly test whether joint liability lending reduces adverse selection problems, an indirect approach is used. More specifically, this chapter examines whether joint liability lending leads to assortative matching in group formation. Since joint liability debt contracts only help to solve the adverse selection problem if assortative matching holds, testing for assortative matching provides an indirect test of the possibility that joint liability lending may help to solve the adverse selection problem.

There is some literature that argues that homogeneous matching only holds in a frictionless world (see Sadoulet and Carpenter, 2001, and the

[1] This chapter is based on Lensink and Mehrteab (2003).

references therein). However, the real world is characterized by frictions due to e.g. imperfect information, the unavailability of partners with the same risk characteristics, the inability to enforce contracts, and the inability to fully screen and monitor group members. The advocates of the matching frictions theory argue that heterogeneous matching might take place, but that the heterogeneity is entirely due to so-called matching frictions. In other words, the matching frictions theory suggests that there will be homogeneous matching in the case where the analysis controls for matching frictions. If there are matching frictions that lead to some heterogeneity, the matching is still essentially homogeneous; heterogeneity is simply due to frictions and therefore generates deviations from optimality. Yet, empirical evidence on the homogeneous matching hypothesis in general and the matching frictions theory in particular is scarce. One of the few exceptions is Sadoulet and Carpenter (2001) on microcredit groups in Guatemala.

This chapter examines the empirical relevance of the homogeneous matching hypothesis for two MFIs in Eritrea. The data we use has been described in chapter 5. Our data provides information that can be used to test the matching frictions hypothesis.

This chapter is organized as follows. Section 8.2 provides a survey of the literature on risk behavior and group formation; section 8.3 explains the methodology we use to test the matching frictions hypothesis; and section 8.4 explains how risk is measured, a variable that we need to test for homogeneous matching. Section 8.5 presents two groups of independent variables that are assumed to affect risk behaviour. Here factor analysis is applied to regroup these variables in a smaller number of factors. Risk behavior is estimated in section 8.6. The results of this equation are used in section 8.7 to test whether homogeneous matching holds if matching frictions are accounted for. Finally, section 8.8 concludes.

8.2 Group formation and homogeneous matching: a literature review

Most of the matching literature draws heavily from the work of Becker (1993), who has worked extensively on marriage matching theory. Ghatak and Guinnane (1999) present a model indicating that the group self-selection process leads to homogeneous matching. Their theoretical framework will be presented below – albeit in a shorter version.

It is assumed that output R takes two values: high (R^h) or low (R^L), where $R^h > R^L \geq 0$. For simplicity, we normalize R^L to be equal to zero. We have two types of borrowers and both borrowers are risk-neutral; one is safe while the other is risky. Output is high with probability $p \in (0, 1)$. p_s and p_r indicate the likeliness of success by safe and risky borrowers, respectively. The risky type fails more often than the safe one ($p_s > p_r$). Yet, the risky borrower receives higher returns if he succeeds. For simplicity, it is assumed that the expected net returns are equal for both safe and risky types:

$$E(R_i) = p_s R_s = p_r R_r = \overline{R} \qquad (1)$$

The best way for banks to separate the safe borrower from the risky borrower is to ask borrowers to pledge collateral. Risky borrowers are likely to fail more often and lose their collateral. If a bank offers two sets of contracts, one with high interest rate and lower collateral and the other with lower interest rate and higher collateral, the risky borrower will choose the former and the safe borrower the later contract. On the basis of interest rate and collateral the bank may be able to distinguish the risky borrower from the safe one. But since the poor in developing countries do not have any serious assets to offer as collateral, banks have no effective way to distinguish between safe and risky borrowers.

With the help of the joint liability mechanism the bank will be able to distinguish the risky from the safe borrower. In the end, the expected result is that the safe types select their likes and form a group in order to acquire loans from the bank, while the risky ones prefer individual loans. As a consequence, the bank can screen borrowers by varying the degree of joint liability. Under a joint liability contract each borrower pays nothing if his/her project fails, and an amount r if his/her project is successful. Here r is assumed to be exogenously determined. In addition, the successful borrower pays a joint-liability payment c if the other member of the group fails. The expected net return of a safe type teamed with a risky type is then:

$$E(\Pi)_{sr}(r,c) = p_s p_r (R^h - r) + p_s (1 - p_r)(R^h - r - c) \quad (2)$$

$$= p_s R^h - \{r p_s + p_s (1 - p_r) c\} \quad (3)$$

Similar calculations could be performed for exclusively safe and risky groups. Since safe types are always preferred as partners, the question becomes: will the risky types be willing to make a large enough transfer to the safe types so that both risky and safe types do better together? The logic behind this transfer is that in case the risky borrower defaults the safe member is going to pay c to the bank. Therefore, the transfer of money from the risky to the safe borrower is to compensate the safe borrower.

The net expected gain of risky borrower from having a safe partner is

$$E(\Pi)_{sr}(r,c) - E(\Pi)_{rr}(r,c) = p_r (p_s - p_r) c \quad (4)$$

Similarly, the net expected loss for a safe borrower of having a risk partner is

$$E(\Pi)_{ss}(r,c) - E(\Pi)_{sr}(r,c) = p_s (p_s - p_r) c \quad (5)$$

By comparing expected returns under alternative scenarios, we can calculate that a safe type will require a transfer of at least $p_s(p_s - p_r)\, c$ to agree to form a partnership with a risky type. Will the risky type be willing to pay that much? His expected net gain from joining a safe type is as much as $p_r(p_s - p_r)\, c$. If $c > 0$ and since $p_r < p_s$, the expected gains to a risky type are always smaller than the expected losses to a safe type. Hence, a risky borrower will not find it profitable to have a safe partner. While every borrower wants to have a safe partner, safe borrowers value safe partners more than risky borrowers value safe borrowers. Hence, a risky borrower cannot cross-subsidize a safe borrower in order to be accepted as a partner, leading to groups containing partners with similar risk behavior. Group formation will display positive assortative matching under joint liability contract.

Ghatak (2000) goes one step further. He shows that if banks can offer a continuum of joint liability and stand-alone debt contracts, incentive compatible separating equilibria may be the result. The safe types prefer a combination of a high joint liability component and a low lending rate, whereas the opposite will hold for a risky borrower. By choosing a joint liability debt contract a borrower signals that he/she is a safe borrower. Ghatak (2000) shows that a joint liability debt contract by solving the adverse selection problem can lead to a Pareto optimal solution.[2]

Xinhau Gu (2000) also deals with the formation of borrowing groups through the exploitation of local information and joint liability. He states that static models implicitly assume a borrower to always be endowed with acceptable (viable) projects. However, entrepreneurs usually have difficulties finding investment opportunities, and dynamic search models are useful tools to address such problems. He examines the impact of uncertainty about investment opportunities on the borrowers' project search decision and on the rate of loan repayment. He shows that safe borrowers prefer to group with safe borrowers since the effective cost of borrowing is positively related to risk taking by group members.

Laffont (2000) shows the role of group lending in differentiating between borrowers of different types (adverse selection). He states that group-lending contracts offer a subtle method of distinguishing between borrowers. When collusion between borrowers under complete information is allowed for, group lending as an instrument improves discrimination between entrepreneurs of different types. So, similar types group together.

[2] For an extension of this model and other related models see Gangopadhyay and Lensink (2001), Xinhau Gu (2000), and Laffont (2000). See also Gangopadhyay, Ghatak, and Lensink (2005).

Sadoulet (1999) presents a model that challenges the commonly assumed homogeneous matching hypothesis. In his model, group membership is endogenous, and group performance depends on both members' types and on the distribution of those types. According to Sadoulet, group members choose partners in a context of missing insurance markets. The point he wants to make is that if insurance markets are missing, homogeneity is not optimal anymore. Heterogeneity emerges as a constrained first-best choice. Sadoulet suggests that members set up insurance arrangements within their group, in which partners will cover each other's loans in case a project fails. The reason for insurance is that borrowers live and work in risky environments and hence need insurance. If a member who is able to insure a partner in need, refuses to pay for him, together with the other member he will lose access to future loans from the program because of the joint liability principle. Alongside these insurance arrangements there are transfer payments between members when both members are successful to remunerate the safe one for covering for the risky one in times of need. Thus, this insurance arrangement is taken to be an important part of the group formation process.

To this end, Sadoulet's model suggests a non-monotonic matching pattern in which safer borrowers will always form groups heterogeneously with partners riskier than themselves. Middle-type borrowers match either heterogeneously with safer borrowers or homogeneously with borrowers of their type depending on whether these are available. Finally, the riskier borrowers match homogeneously.

Note that the models by Ghatak (1999) and Sadoulet (1999) are similar. Ghatak gets homogeneous matching since his model is static, whereas Sadoulet gets heterogeneous matching since his model is repeated. Moreover, in the model by Ghatak the benefit of homogeneous matching is that it improves repayment rates and thus leads to lower interest rates. The problem is that the decrease in the interest rate cannot compensate the safe borrowers for having to cover the risky borrowers' loans if they fail. So, safe and risky borrowers will not form groups. In the model by Sadoulet the benefit is not lower interest rates but access to future loans, which has a much higher direct value.

Armendáriz De Aghion and Gollier (2000) state that in urban economies with heterogeneous, anonymous, and relatively mobile borrowers, random (rather than assortative) matching is incentive compatible for all types of borrowers. A particular feature of their paper is that they assume that borrowers do not know each other. They show that cross-subsidisation among members provides a kind of a collateral that reduces the negative externalities from risky to safe borrowers. The main implication of their work is that, as we move away from village economies by allowing for imperfect information, assortative matching no longer leads to an equilibrium situation, and yet group lending can improve efficiency and enhance welfare.

There are few empirical studies available that have rigorously tested the homogeneous matching hypothesis. Most empirical studies have simply assumed that homogeneous matching takes place. Some studies, however, do provide some insights. For instance, for groups belonging to BancoSol, Bolivia, Van Tassel (2000) found that groups match heterogeneously in unobservable business characteristics.

The only empirical paper available that has rigorously tried to investigate the matching of group members is the one by Sadoulet and Carpenter (2001). For credit groups in Guatemala they estimated the relationship between risk and the level of risk heterogeneity in the individual groups, explicitly accounting for the endogeneity of group formation and of borrowers' choice of project risk. Their results show that borrowers in Guatemala group heterogeneously, and that the heterogeneity cannot be explained by matching frictions. In line with the theoretical paper by Sadoulet (1999) they suggest that borrowers might want to form heterogeneous groups in order to set up insurance arrangements.

8.3 The methodology: the role of matching frictions

We follow the methodology set out by Sadoulet and Carpenter (2001). The reader is referred to their paper for a detailed explanation of the methodology. The main problem we have to deal with is as follows. In a frictionless world, the assortative matching theory implies that all borrowers will choose their first-best risk level, and will match together with partners with the same (first-best) risk level. However, if there are matching frictions, borrowers may be forced to match with partners of a different risk level, even if they prefer to match with borrowers of the same risk type.

The matching frictions theory states that homogeneous matching only holds in a frictionless world and that all heterogeneity comes from matching frictions. This implies that there should be no statistically significant relationship between first-best risk (risk in a frictionless world) and heterogeneity. In order to test this theory, we need indicators for first-best risk and matching frictions. The problem is that these variables are not observable. Sadoulet and Carpenter (2001) solve this problem as follows. If there are matching frictions, the level of risk heterogeneity of borrower *i* (h_i) depends on her first-best risk choice (r_i^*) and on matching frictions (f_i)[3]

$$h_i = H(r_i^*, f_i) \tag{6}$$

Since with matching frictions a borrower may not be able to match with his/her preferred partners, he/she may decide to adjust his/her own risk choice. This implies that the risk level a borrower chooses (the observed risk, r_i) is a function of characteristics that affect the first-best risk choice (X_i) and the heterogeneity he/she is faced with, i.e.

$$r_i = R(X_i, h_i) \tag{7}$$

If equation (6) is substituted in equation (7) a reduced form expression for the observed risk level can be obtained

$$r_i = k(X_i, f_i) \tag{8}$$

[3] In fact *f* refers to a matrix of variables determining the friction level f_i.

The full system of equations (the structural model) can now be specified as:

$$h_i = H(r_i^*, f_i) \tag{9}$$

$$r_i = k(X_i, f_i) \tag{10}$$

$$r_i^* = k(X_i, 0) \tag{11}$$

If the matching frictions hypothesis holds, then $\frac{\partial h_i}{\partial r_i^*} = 0$. It may be useful to compare this condition with the condition for homogeneous matching in a frictionless world. In a frictionless world with homogeneous matching h_i should be zero in all groups. So, heterogeneity between all members within a group and hence group heterogeneity will then be zero. In this situation, the test for homogeneous matching could be based on a measure of group heterogeneity. Simply testing whether this measure is equal to zero would do. However, with matching frictions h_i does not need to be equal to zero. The correct test is $\frac{\partial h_i}{\partial r_i^*} = 0$, which implies that homogeneous matching cannot be confirmed or rejected by testing whether a measure for group heterogeneity differs from zero. In order to correctly test the matching frictions theory, we need variables for risk and heterogeneity at the individual level. The risk level a borrower chooses depends on the heterogeneity he/she is faced with. Since the heterogeneity an individual borrower is faced with is greater, the more his/her risk level is away from the group mean, a group heterogeneity measure, which gives the same heterogeneity level for all members in a group, cannot be used. In the next section we will explain how we measure risk. Section 8.7 will explain how we measure heterogeneity. But first we need to complete the discussion of the empirical methodology.

The trick is to first estimate the actual risk equation, for which we take, for reasons of convenience, a linear specification:

$$r_i = X_i\alpha + f_i\beta + \varepsilon_i \tag{12}$$

From this regression, estimated values for first-best risk and matching frictions can be obtained:

$$\overline{r_i^*} = X_i \overline{\alpha} \qquad (13)$$

$$\overline{\beta f_i} = f_i \overline{\beta} \qquad (14)$$

These estimated values are then substituted in the equation for heterogeneity:

$$h_i = \alpha + \gamma \overline{r}_i^* + \delta \overline{\beta f}_i + \varepsilon_i \qquad (15)$$

Homogeneous matching will be empirically confirmed if $\gamma = 0$. It is expected that δ≥0.

8.4 How to measure risk

The first step in the analysis is to develop a measure for risk, which is needed to estimate the risk equation (equation 12). Note that in the theoretical models it is assumed that there is only one project available per individual, which implies that projects and borrowers are interchangeable. This also implies that the theoretical measure for risk refers to both the riskiness of the borrower and the project. However, empirically there is no perfect measure for this theoretical risk concept available. We proxy the theoretical concept of risk by developing a measure for the risk of a borrower's repayment strategy. Even this is not directly measurable and therefore has to be proxied by an (admittedly imperfect) indicator. In line with Sadoulet and Carpenter (2001), we proxy risk (r) by:

$$r_i = \frac{P_i - S_i}{P_i}, \text{ for } P_i \geq S_i$$

and $r_i = 0$ for $P_i < S$

where P_i is the loan payment due per month and S_i is the amount the borrower reports to have saved one week before the due date to cover the loan payments.[4] The risk indicator varies between 0 and 1. The higher the percentage amount saved a week before the repayment date, the lower the risk of a borrower's repayment strategy.

We consider loan payments due per month, since for the two microfinance programs in Eritrea the install payments members are supposed to make are monthly. In the questionnaire we asked the borrowers to specify the agreed install payment per month (P_i). We also asked borrowers to specify the average cumulative savings until one week before due date (S_i).

Table 8.1 gives information on the risk measure and on the variables used to construct this measure. The table also provides data on the credit amount.

Table 8-1 Information on credit and risk

	CREDIT SIZE	P	S	r
Mean	3961	422	356	0.17
Median	3500	380	300	0.09
Maximum	8500	2320	2080	1.00
Minimum	750	71.25	0.00	0.00
Std. Dev.	1802	315	272	0.213
Skewness	0.468	2.714	2.440	1.967
Kurtosis	2.406	13.008	12.257	7.761
Jarque-Bera	17.97	1895.87	1601.76	557.80
Observations	351	351	351	351

Note: all values (except for *r*) are in Nakfas. The Jarque-Bera statistic is a test for normality. The statistic has a χ^2 distribution with 2 degrees of freedom under the null hypothesis of normally distributed errors.

[4] Note that Sadoulet and Carpenter use the sum of expected sales in the last three days before the due date as the scaling factor, instead of P_i. Our questionnaire also contains a question on the expected sales in the last days (one week in our case) before the due date. However, since the answers to this question were totally unreliable we decided to scale by P_i .

The value of loans ranges from 750 Nakfas to 8500 Nakfas, with mean and median loan sizes of 3961 and 3500 Nakfas, respectively. Loan terms vary from 3 to 24 months. The mean of our risk indicator is about 0.17, with an even lower median (0.09). Of the 351 borrowers 105 are left censored on the risk measure ($r = 0$), 10 are right censored ($r = 1$) and 236 are uncensored ($0 < r < 1$). Since $r = 0$ for a relatively high percentage of the group of borrowers, many borrowers show a tendency to save enough to repay the full monthly amount by the third week. Thus, borrowers seem to show a high degree of punctuality and a great readiness to save ahead of time in order to be sure of future access to credit from the program. Note that none of the variables is distributed normally.

It should be noted that a possible caveat of our risk measure is that a person who gets a fixed payment (more than P_i) in the week before the payment can be very safe despite the fact that $S_i = 0$. However, we do not think that this will substantially affect our results since in practice this does not seem to happen that often. Related to this problem, the validity of our measure may depend on the time profile for the different projects. Our measure may incorrectly give a higher risk ranking to a borrower with a project that yields an uncertain amount of income in the first week as compared to another borrower with a project that yields a certain amount of income in the fourth week. Since it is impossible to obtain detailed information about the time profile of returns for the different projects the loans are used for, this problem cannot be solved. But given the fact that the bulk of loans by the two programs in Eritrea are forwarded for the same purposes (trading) so that the time profile of most of the projects the loans are used for are probably similar, we are reasonably convinced about the validity of our risk proxy.

8.5 Variables proxying for first-best risk and matching frictions

The next step in the analysis is to determine which variables possibly affect risk, which of those variables are related to first-best risk and which of them are related to matching frictions. Hence, we need to determine a vector of variables X (first-best) and f (matching frictions).

8.5.1 Matching frictions (f)

Sadoulet and Carpenter (2001) argue that variables proxying for matching frictions include indicators of the degree of asymmetric information among different members of a group, proxies for the ability to monitor and screen the activities of the different members in a group, and variables on the available borrowing options. In line with Sadoulet and Carpenter (2001) we select from our data set the following list of variables related to monitoring, screening, the available information on other members, and the possibility to obtain credit.

- BOGROUP = a dummy variable with a 1 if the borrower is born in the village or town where the survey is conducted;
- CHGRDUM = a dummy variable with a 1 if the group member has been a member of another group;
- KNMEMDUM = a dummy variable with a 1 if the borrower knew the members well before they formed the group;
- INTEGRITY = a dummy variable with a 1 if the borrower knew about the behavioral integrity of all his/her fellow group members before the formation of the group;
- KNACTDUM = a dummy variable with a 1 if the group member knows the economic activities of the other group members;
- KNPURPDUM = a dummy variable with a 1 if the borrower knows for what purpose the other group members acquired their latest loans;
- KNSELDUM = a dummy variable with a 1 if the borrower knows the monthly sales of the other group members;
- LDIST = the logarithm of the average distance of the business of the group member from that of the other group members;
- VISTDUM = a dummy variable with a 1 if the group member visits other group members;
- ARREAR = a dummy variable with a 1 if the borrower has had problems repaying his/her debt in the current loan cycle;
- OTHCREDIT = a dummy variable with a 1 if the borrower has other sources of credit;
- ACORDUM = a dummy variable with a 1 if the group belongs to the SZSCS;

- NOMEM = the number of members in a group.

From this list of variables, BOGROUP, CHGRDUM, KNMEMDUM, INTEGRITY primarily refer to social ties and peer screening variables. These are variables that indicate the amount of information members have on each other. These variables, with the exception of CHGRDUM, deal in particular with the available information before forming the group. An increase in the value of one of these indicators implies more information about each other that might increase the probability of better peer screening and stronger social ties. KNACTDUM, KNPURPDUM and KNSELDUM, LDIST and VISTDUM have to do with the (possibility of) peer monitoring. More visits among members and a shorter distance between members increase peer monitoring. More group members tend to increase monitoring efforts, but there is also more scope for free riding. ARREAR and OTHCREDIT refer to possibilities to obtain credit from other sources; OTHCREDIT directly measures whether a borrower has been able to raise funds from other sources than the microfinance institution, and ARREAR measures repayment problems and may indicate future possibilities to raise credit. ACORDUM and NOMEM are not directly related to the issues discussed so far but – as will become clear later on – they have been included since they are highly correlated to each other.

8.5.2 First-best risk

We assume that first-best risk can be picked up by variables that are directly related to the socio-economic situation of the borrower. We consider the following variables:

- LINC: the logarithm of total monthly income;
- AGE: the age of a borrower;
- GENDUM: a dummy with a 1 for a male, and a 0 for a female;
- ILLIT: a dummy with a 1 if the borrower is illiterate;
- PRIM: a dummy with a 1 if the borrower has had any primary education;

- SEC: a dummy with a 1 if the borrower has had any secondary education;
- GLEADER: a dummy with a 1 if the borrower is a group leader;
- MOSLDUM: a dummy with a 1 if the borrower is a Muslim.

The concepts matching frictions and first-best risk are latent variables, which cannot be observed directly. Above, we have selected a group of variables that is assumed to be related to matching frictions, and a group of variables that is assumed to be related to first-best risk. In order to better account for the high collinearity between some of the variables within the two groups, and in order to test whether we can reduce the number of independent variables by constructing a smaller amount of new composite variables, we performed a multiple factor analysis (MFA).

We started by applying a factor analysis to the indicators of the group of variables related to matching frictions. The analysis suggests that eleven indicators in this group can be divided into three underlying factors. The two remaining indicators (ARREAR and OTHCREDIT) are left out of this analysis since they have very low factor loadings, even if more underlying factors are allowed for. The factor loadings of the analysis are given in table 8-2.

The first factor mainly has to do with KNMEMDUM and INTEGRITY, suggesting that the underlying factor in this case relates to information members have about each other before they formed a group. ACORDUM and NOMEM mainly determine the second factor. NOMEM has a negative factor loading, which suggests that with respect to our sample the average number of members in credit groups from the SCSZS is lower than in groups from the SMCP. A closer look at the data set confirms this: the average number of members in credit groups from the SMCP is 5.2, whereas it is 3.6 for the SCSZS. The positive factor loading on VISTDUM suggests that members of credit groups from the SCSZS visit each other more regularly than those of the SMCP system. The third factor mainly has to do with KNPURPDUM and to a lesser extent with KNACTDUM. This may suggest that in this case the underlying factor relates to information members have about each other's business, after the group has been formed.

Table 8-2 Factor loadings for factor analysis on matching frictions variables

	FACTOR1	FACTOR2	FACTOR3
ACORDUM	-0.146	0.916	0.129
BOGROUP	0.275	-0.227	-0.021
CHGRDUM	0.018	0.236	-0.019
KNMEMDUM	0.923	0.038	0.208
INTEGRITY	0.935	0.050	0.202
LDIST	-0.176		-0.025
KNACTDUM	0.226	-0.093	0.376
KNPURPDUM	0.058	0.120	0.733
KNSELDUM	0.102	0.185	0.048
VISTDUM	0.152	0.323	0.306
NOMEM	0.077	-0.632	0.019
Chi square Statistic: 24. 7; 25 Df; p-value: 0.479; CUMVAR=0.394			

Note: factor loadings smaller than 0.01 are not reported. Df denotes the degrees of freedom. CUMVAR gives the cumulative variance explained by the factors taken into account. The factor analysis is done on 323 observations (the common sample of all indicators; observations refer to members of both MFIs). The Chi square Statistic is a test of the hypothesis that three factors are sufficient versus the alternative that more are required. The P-value is the probability of being wrong when the null hypothesis is rejected (the plausibility of the null hypothesis. So, the smaller is the P-value, the less plausible is the null hypothesis).

In the remainder of the analysis we will use the three factors, instead of the eleven original indicators. We interpret FACTOR1 and FACTOR3 as factors that primarily have to do with the asymmetry of information among group members. FACTOR1 picks up information before forming the group; FACTOR3 picks up information after the group has been formed; and FACTOR2 primarily relates to being a member of a credit group within the SCSZS and the number of members within a group. The latter variable is important for risk taking since it gives information on a possible peer monitoring effort. Armendáriz De Aghion (1999, proposition 3, p.95) states that "a larger group size tends to increase peer monitoring effort, due to a joint-responsibility, a cost-sharing, and a commitment effect. However, a larger group size (also) increases the scope for free riding in debt-repayment decisions".[5]

Next, we perform a factor analysis on the indicators for first-best risk. However, here the factor analysis showed that it is not possible to combine the indicators into a smaller group of underlying factors. The number of factors that has to be taken into account to accept the null hypothesis of sufficient factors is almost equal to the original amount of indicators. Therefore, we decided to proceed with the individual first-best indicators in the remainder of the analysis.

8.6 Estimating risk

The next step in the analysis is to examine the possible empirical relevance of our matching frictions and first-best risk variables for explaining risk of a borrower's liquidity strategy. In other words, the next step is the estimation of equation (12).

The dependent variable is the proxy for risk, r, which we have constructed. The independent variables are the eight first-best risk indicators, the three factors related to matching frictions, and the remaining two variables (ARREAR and OTHCREDIT), which are also related to matching frictions. To examine non-linear effects we also tried quadratic terms, but except for the quadratic term of LINC (LINC2) none of these appeared to be significant. Hence, they were left out of the analysis.

The constructed dependent variable is censored between 0 and 1. Therefore, we estimate with the Tobit estimation technique with left and right censoring (using normal distribution of error terms). We also present ordinary least squares (OLS) estimates to test for differences in outcome

[5] Note that in Armendáriz De Aghion (1999) groups are exogenously given. In practice, there is a tradeoff between the cost of group size (monitoring effort) and benefits of size (diversification, easier to cover one defaulting partner). Group size is thus endogenous. We ignore this problem in our analysis.

due to different estimation techniques. The estimation results are presented in table 8-3.

Equations 1A and 1B in table 8-3 show that LINC, LINC2, GLEADER, SEC, ARREAR and FACTOR2 significantly affect risk behavior. Since LINC has a significantly negative coefficient and LINC2 a significantly positive coefficient, there seems to be a non-linear relationship between the income of a borrower and his/her risk behavior. For low-income levels, an increase in income reduces risk, whereas it increases risk after a certain threshold level of income has been passed. Positive significant coefficients for GLEADER, SEC and ARREAR suggest that a group leader takes more chances than a normal group member, that members who are more educated take more risks, and that members who have had repayment problems in the past also take more chances. The negative coefficient for FACTOR2 implies that borrowers in a borrowing group belonging to the SZSCS take less risks. The underlying reason probably is that the numbers of members in credit groups belonging to the SZSCS are lower. Larger groups may lead to more risk taking of the individual members, possibly due to a better scope for freeriding. These results hold for both the OLS and Tobit estimates.

In equations 2A and 2B the regressions are repeated by ignoring the insignificant terms. These regressions confirm the results suggested by equations 1A and 1B. Finally, we re-estimate the equations by replacing ARREAR with AMARREAR (equations 3A and 3B). AMARREAR measures the amount of money that was involved when the borrower had problems repaying the debt, as a percentage of the loan size in the previous loan cycle. This indicator serves as an alternative indicator for ARREAR. The results of these regressions again confirm the basic message of equations 1A and 1B.

Table 8-3 Estimating results on risk

	1A	1B	2A	2B	3A	3B
Method	OLS	Tobit	OLS	Tobit	OLS	Tobit
LINC	-0.866***	-1.224***	-0.880***	-1.260***	-0.487***	-0.790***
	(-2.93)	(-3.48)	(-3.05)	(-3.63)	(-2.19)	(-2.77)
LINC2	0.055***	0.078****	0.056***	0.080***	0.029***	0.048***
	(2.73)	(3.31)	(2.86)	(3.48)	(1.93)	(2.51)
AGE	0.0002	0.0003				
	(0.22)	(0.21)				
GENDUM	-0.016	-0.029				
	(-0.63)	(-0.84)				
ILLIT	-0.029	-0.037				
	(-0.96)	(-0.91)				
PRIM	0.004	0.0020				
	(0.16)	(0.06)				
SEC	0.111***	0.149***	0.116***	0.157***	0.116***	0.148***
	(2.40)	(2.39)	(2.78)	(2.72)	(2.85)	(2.59)
GLEADER	0.0585***	0.073***	0.060***	0.074***	0.042***	0.049**
	(2.70)	(2.46)	(3.00)	(2.62)	(2.25)	(1.91)
MOSDUM	0.012	0.019				
	(0.40)	(0.47)				
ARREAR	0.320***	0.386***	0.321***	0.386***		
	(8.35)	(8.38)	(8.53)	(8.47)		
AMARREAR					0.399***	0.540***
					(6.72)	(7.61)
OTHCREDIT	0.0028	-0.0049				
	(0.06)	(-0.08)				
FACTOR1	-0.00076	0.0078				
	(-0.07)	(0.50)				
FACTOR2	-0.022***	-0.049***	-0.022***	-0.050***	-0.016***	-0.037***
	(-2.07)	(-3.16)	(-2.13)	(-3.25)	(-1.73)	(-2.74)
FACTOR3	-0.006	-0.011				
	(-0.47)	(-0.68)				
CONSTANT	3.443***	4.734***	3.480***	4.846***	2.092***	3.188***
	(3.18)	(3.64)	(3.28)	(3.78)	(2.54)	(3.03)
adj. R^2	0.39	0.40	0.40	0.41	0.49	0.53

Note: the amount of observations is 323 for all regressions. That is, although we have 351 members in our data sets, the number of members who belong to the 102 groups is just 323. t-values (z-values for Tobit) based on white Heteroskedasticity-consistent standard errors (for the OLS regressions) and QML (Huber/White) standard errors between parentheses. The Tobit estimates are done with left (0) and right (1) censoring; there are 94 left censured observations and 10 right censured observations.

Since FACTOR2 mainly has to do with three indicators – ACORDUM, VISTDUM and NOMEM – we also perform OLS and Tobit regressions in which FACTOR2 is replaced by one of these individual indicators. The regression results show that each of these individual terms, with the exception of the OLS estimate for NOMEM, is significant. Being a borrower from a credit group associated with the SZSCS has a negative effect on risk taking. The same holds for more visits among members of a credit group. An increase in the number of members of a credit group enhances risk-taking behavior of an individual borrower. The results are presented in table 8-4.

We are now able to come up with an estimate of $\overline{r_i^*} = X_i \overline{\alpha}$ and $\overline{\beta f_i} = f_i \beta$ (equations 13 and 14, section 8.3). To this end we use the estimation results of equation 2B (the Tobit estimates) presented in table 8-3. As we explained before, we argue that the variables that are related to the socioeconomic situation (i.e. LINC, LINC2, SEC and GLEADER) determine the risk choice in a frictionless world. The other variables (ARREAR and FACTOR2) are primarily related to matching frictions. By using the estimated coefficients of equation 2B (table 8-3) we can now come up with an estimate of $\overline{r}_i^*$, which we name FIRSTBEST, and $\overline{\beta f_i}$, which we name FRICTION.[6]

[6] We assume that the conditional mean (*E[y_i]*) of the Tobit regression equation $y_i = \beta x_i + \varepsilon_i$ equals K_i x_i. If all independent variables are taken into account, this predicts the so-called expected *latent variable.*

Table 8-4 Estimating risk by replacing FACTOR2 with ACORDUM, VISTDUM and NOMEM

	1A	1B	2A	2B	3A	3B
Method	OLS	Tobit	OLS	Tobit	OLS	Tobit
LINC	-0.833***	-1.179***	-0.800***	-1.074***	-0.840***	-1.166***
	(-2.95)	(-3.49)	(-2.87)	(-3.25)	(-2.89)	(-3.35)
LINC2	0.053***	0.076***	0.051***	0.068***	0.053***	0.074***
	(2.77)	(3.35)	(2.68)	(3.07)	(2.69)	(3.18)
SEC	0.085***	0.114***	0.078***	0.092***	0.109***	0.139***
	(2.25)	(2.04)	(2.12)	(1.70)	(2.64)	(2.47)
GLEADER	0.060***	0.075***	0.057***	0.071***	0.060***	0.074***
	(3.10)	(2.74)	(3.02)	(2.65)	(3.00)	(2.64)
ARREAR	0.324***	0.392***	0.316***	0.373***	0.318***	0.379***
	(8.74)	(8.73)	(8.66)	(8.54)	(8.50)	(8.39)
ACORDUM	-0.042***	-0.097***				
	(-2.31)	(-3.48)				
VISTDUM			-0.049***	-0.076***		
			(-2.37)	(-2.79)		
NOMEM					0.010	0.023***
					(1.43)	(2.40)
CONSTANT	3.310***	4.578***	3.224***	4.246***	3.301***	4.422***
	(3.21)	(3.68)	(3.17)	(3.48)	(3.08)	(3.43)
adj. R^2	0.39	0.41	0.39	0.41	0.39	0.41

See the note to table 8.3.

8.7 Heterogeneity

The final step in the analysis is to estimate the heterogeneity equation. Therefore, we first need to develop a measure of risk heterogeneity.

8.7.1 The measure for risk heterogeneity

In line with Sadoulet and Carpenter (2001) we measure risk heterogeneity (h_i) by:

$$h_i = \left[\sum_{r_j \in G_i} \frac{(r_i - r_j)^2}{(N_i - 1)} \right]^{0.5} sign(r_i - \overline{r_i})$$, where $\overline{r_i}$ is the mean risk in i's group G_i.

This proxy measures the average Euclidean distance between the risk of a borrower and all of his/her group partners. Note that our measure for heterogeneity is individual rather than group specific, as it should be. Moreover, the heterogeneity proxy gives higher degrees of heterogeneity for borrowers with a risk level that is further away from the mean risk level in the group – this is also in line with theory. We sign the average Euclidean distance in order to allow for possible differences in behaviour for the relative safe and relative risky borrowers in a group (which is in line with theories that examine group formation in the context of missing insurance markets). However, we also used a measure for heterogeneity that is not adjusted for having a risk above or below the mean risk. This gave qualitatively the same results. Since our space is limited, we have not presented these results.

Table 8.5 gives descriptive statistics of *h*.

Table 8.5 Heterogeneity

	h
Mean	-0.005
Median	-2.78E-17
Maximum	1.00
Minimum	-1.00
Std. Dev.	0.265
Skewness	0.115
Kurtosis	5.227
Jarque-Bera	72.65

Next, we will examine whether or not heterogeneity is caused by matching frictions. We will do this by estimating equation 15.

8.7.2 Estimation results

The estimates of the heterogeneity equation are presented in table 8.6.[7] Again we use the OLS as well as the Tobit estimation technique. The dependent variable in the regressions is our proxy for heterogeneity (*h*). It seems that the coefficient for FIRSTBEST is significantly different from zero at the 99 per cent level, which strongly suggests that homogeneous matching will not occur, even if the estimates are controlled for matching frictions.

Table 8.6 Estimating heterogeneity

	1	2
METHOD	Tobit	OLS
FIRSTBEST ($\overline{r}_i^*$)	0.663	0.660
	(3.20)	(3.19)
FRICTION ($\overline{\beta f_i}$)	0.623	0.620
	(5.54)	(5.52)
CONSTANT	3.129	3.115
	(3.13)	(3.13)
adj R^2	0.15	0.16

Note: the amount of observations is 323 for all regressions. t-values (z-values) for OLS *(*for Tobit*)* between parenthesis (based on White Heteroskedasticity-Consistent Standard Errors and Covariances and Huber/White robust standard errors and covariances, respectively). In equation 1 there is one right and one left censored observation.

A possible caveat of our analysis may be that we have not included all relevant variables in the equation for risk and that this affects our estimates. There may be some relevant matching frictions or first-best variables missing. This is for example suggested by the fact that there are

[7] It should be noted that the variables FIRSTBEST and FRICTIONS are measured with errors. OLS (and Tobit) estimates of the heterogeneity equation may therefore be biased. A possible solution, used by Sadoulet and Carpenter (2001), is to estimate the heterogeneity equation with instrumental variables. However, due to a lack of candidates for instruments in our sample we decided to rely on the OLS estimates.

only a few variables for matching frictions significantly. This may then lead to an omitted variable bias. The problem with omitted variable bias is that it may lead to a correlation between the disturbance term and one of the right-hand side variables. This may bias the estimates of the coefficient and standard error of X_i and f_i , which may consequently affect our estimates of the heterogeneity equation. However, note that we are not directly interested in the parameters in equation (12). We only use the fitted variables for first-best risk and matching frictions. It is easy to show that the predicted values are consistent estimators, so that we can be reasonably sure that our result with respect to FIRSTBEST and hence our conclusion that homogeneous matching does not hold is not affected by omitted variable bias because of a misspecification of the risk equation. Our approach of obtaining FIRSTBEST and FRICTION is somewhat comparable to an instrumental variable (IV) technique. As is well-known IV estimation is designed to overcome problems caused by correlation between the disturbance term and the right-hand side variables.

8.8 Conclusions

This chapter aims to provide new insights into the empirical relevance of the homogeneous matching hypothesis for microfinance groups in Eritrea. A better insight into group formation and whether these groups are homogeneous is extremely important for our understanding of the working of microfinance programs. The result of our analysis can be used as an input for the analysis of repayment performance of joint liability schemes versus individual liability debt contracts. It also provides indirect evidence on the reliability of the hypothesis that group lending by means of joint liability lending can reduce adverse selection problems.

The estimates with respect to risk behavior suggest that among the borrowers from the microfinance programs in Eritrea, there is a non-linear relationship between the income and risk taking. Below a certain threshold level of income, an increase in income will lead to less risk taking, whereas an increase in income above a certain level will increase risk taking. We also find that group leaders take more risk than regular group members, that better educated borrowers take more risk, and that borrowers who have had repayment problems in the past will take more risk. Moreover, we find some evidence that borrowers in larger groups will take more risk than borrowers in smaller groups.

Concerning the homogeneous matching hypothesis, our results strongly indicate that groups are formed heterogeneously. Most importantly, we do not find support for the matching frictions hypothesis, in the sense that even if we control for matching frictions, credit groups in Eritrea do not seem to consist of borrowers of a similar risk type. The implication of this finding for repayment behavior is not clear beforehand. However, our result seems to be bad news for those who argue that group-based lending may reduce problems of adverse selection. In some theoretical papers it has been argued that incentive compatible separating equilibria will result if a lender offers different types of debt contracts, with varying components for joint liability. By choosing a particular debt contract, the borrower will reveal his/her type and hence the asymmetric information and consequently the adverse selection problem will be solved. However, this result is based on the homogeneous matching hypothesis.

Of course, some reservations with respect to our main conclusions can be made. For instance, the classification of variables in a group that primarily deals with matching frictions, and a group of variables dealing with first-best risk determinants may be criticized. In addition, our variables FIRSTBEST and FRICTION are constructed variables, and therefore are measured with error. This may bias the estimates of the coefficients. Moreover, the measure of risk we use may not be the most accurate measure for risk taking. There may exist other measures of risk that are better proxies. It may then be the case that using another measure for risk will lead to homogeneous matching, instead of the heterogeneous matching we found by using our measure for risk. More research on these issues is needed. Nevertheless, given the data we have, and taking into account all possible drawbacks of the methodology used, we think that our analysis, at the least, suggests that the commonly held assumption of homogeneous matching can not be confirmed for the case of Eritrea. If one accepts that groups are formed heterogeneously, an important issue is then to examine why this is so. A possible reason brought forward in some recent papers is the insurance that risky and safe borrowers may provide. The models behind the homogeneous matching hypothesis assume that borrowers are risk neutral and that project returns do not covary. This implies that in these models there is no possibility to gain from economies of risk pooling. However, if borrowers are risk averse and project returns are not independent, a borrower may profit from grouping with another borrower if the project returns of the two borrowers are negatively correlated. This may then imply that heterogeneous matching is the optimal outcome.

APPENDIX: List of variables used in the analysis with expected signs

Table 8-A1 List of variables with their expected signs

INDEPENDENT VARIABLES	EXPECTED SIGNS
PEER MONITORING	
KNACTDUM	-
KNPURPDUM	-
KNSELDUM	-
LDIST	+
VISITDUM	-
SOCIAL TIES	
BOGROUP	-
KNMEMDUM	-
INTEGRITY	-
CHGRDUM	+
PERSONAL CHARACTERISTICS	
AGE	+/-
GENDUM	+/-
ILLIT	+/-
PRIM	+/-
SEC	+/-
MOSLDUM	+/-
LINC	+/-
OTHER VARIABLES	
OTHERCREDIT	-
GLEADER	+/-
AREAR	+
NOMEM	+/-
ACORDUM	+/-