Risk Sharing and Network Formation

Marcel Fafchamps   Flore Gubert[*]

JEL: O1, I3

Keywords: risk-sharing; networks; dyadic regressions

Interpersonal relationships have long been suspected of shaping agrarian institutions, probably because weak formal institutions must be supplemented by interpersonal trust. This is particularly true for informal risk sharing, a fundamental risk coping mechanism for the rural poor (e.g. Mark R. Rosenzweig, 1988, Marcel Fafchamps, 1992, Stephen Coate and Martin Ravallion, 1993, Robert Townsend, 1994, Ethan Ligon et al., 2001, Andrew Foster and Mark R. Rosenzweig, 2001).

In this paper we examine whether risk sharing networks are formed so as to maximize the mutual gains from pooling income risk. The benefit from risk pooling is largest when households have different income profiles – e.g., different occupations – and are subjected to different sources of risk – e.g., live far apart. Gains from risk sharing therefore increase with social and geographical distance. But distance also raises the cost of interpersonal links. The net effect on link formation is a priori indeterminate.

We investigate this issue empirically using survey data from the rural Philippines. Marcel Fafchamps and Susan Lund (2003) have shown that informal gifts and loans serve a risk sharing purpose but follow interpersonal networks. Here we show that occupation is not a major determinant of risk sharing links. In contrast, geographic proximity – possibly correlated with kinship ties – is a strong correlate of risk sharing networks. This may be because it facilitates monitoring and enforcement.

This paper also makes a methodological contribution to the estimation of network regressions. We clarify the identification issues raised by dyadic regressions – that is, regressions in which each observation expresses a relationship between pairs of nodes. We also extend the concept of robust standard errors to dyadic regressions, thereby providing an easy alternative to network inference methods based on permutations or generalized least squares.

I. Econometric issues

We estimate a reduced-form regression of the form:

$$L_{ij} = 1 \text{ if } B(d_{ij}) - C(d_{ij}) + e_{ij} > 0$$

(I.1)
$$= 0 \text{ otherwise}$$

where L_{ij} denotes the existence of a link between individuals i and j and d_{ij} represents the distance between i and j. The benefits and costs of the link are denoted $B(d_{ij})$ and $C(d_{ij})$, respectively; e_{ij} is a residual effect. A link is established if benefits exceed costs.

Benefits from pooling income risk fall as incomes are more correlated. We expect $B(d_{ij})$ to rise with d_{ij} if income shocks are less correlated the more different individuals are. Because of information asymmetries and enforcement issues, $C(d_{ij})$ is likely to increase with distance. People may also find it difficult to ascertain the income shocks faced by those with different income streams. It is thus unclear whether individuals establish links with people who are in the best position to pool income risk. Estimating equation (I.1) is the main objective of this paper.

Equation (I.1) is a dyadic regression model. The estimation of dyadic regressions raises two types of difficulties: identification, and inference. The first problem relates to the form in which regressors enter the regression. The second relates to the estimation of standard errors. We discuss these in turn.

Dyadic regressors come in two forms: attributes w_{ij} of the link between i and j, such as the geographical distance between them, and attributes z_i and z_j of individuals i and j. Regressors must enter a dyadic regression in a symmetric fashion so that the effect of (z_i, z_j) on Y_{ij} is the same as the effect of

(z_j, z_i) on Y_{ji}. Dyadic regressions must therefore be written in a way that preserves this symmetry. How this is accomplished depends on whether the dyadic relationship is directional or not.

A dyadic relationship is undirectional if $Y_{ji} = Y_{ij}$ for all i, j. In this case, symmetry requires that regressors satisfy $\beta X_{ij} = \beta X_{ji}$. One easy way of satisfying this requirement is to specify the regression as:

$$(I.2) \qquad Y_{ij} = \alpha + \beta_1 |z_i - z_j| + \beta_2 (z_i + z_j) + \gamma |w_{ij}| + u_{ij}$$

where z_i and z_j are characteristics of individual i and j thought to influence the likelihood of a link Y_{ij} between them. Coefficient β_1 measures the effect of differences in attributes on Y_{ij} while β_2 captures the effect of the combined level of z_i and z_j on Y_{ij}. If a dyadic relationship is directional, Y_{ij} needs not equal Y_{ji}. In this case, the symmetry requirement can be satisfied by specifying the model as:

$$(I.3) \qquad Y_{ij} = \alpha + \beta_1 (z_i - z_j) + \beta_2 (z_i + z_j) + \gamma w_{ij} + u_{ij}$$

For β_2 to be identified, individuals must not have the same number of links – see Marcel Fafchamps and Flore Gubert (2006) for details.

Dyadic observations are not independent since $E[u_{ij}, u_{ik}] \neq 0$ for all i and $E[u_{ij}, u_{kj}] \neq 0$ for all j. We also have $E[u_{ij}, u_{jk}] \neq 0$ and $E[u_{ij}, u_{ki}] \neq 0$. Provided that regressors are exogenous, applying OLS to (I.2) and (I.3) yields consistent coefficient estimates but standard errors are inconsistent. To obtain consistent standard errors, we extend the method that Timothy J. Conley developed to deal with spatial correlation of errors:

$$(I.4) \qquad AVar(\widehat{\beta}) = \frac{1}{N-K} (X'X)^{-1} \left(\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} \frac{m_{ijkl}}{2N} X'_{ij} u_{ij} u_{kl} X_{kl} \right) (X'X)^{-1}$$

where β denotes the vector of coefficients, N is the number of dyadic observations, K is the number of regressors, X is the matrix of all regressors, X_{ij} is the vector of regressors for dyadic observation ij,

and $m_{ijkl} = 1$ if $i = k, j = l, i = l$ or $j = k$, and 0 otherwise.[1] Formula (I.4) also corrects for possible heteroskedasticity.

II. Empirical results

The empirical analysis is based on a survey conducted in northern Philippines specifically to study risk sharing networks (Susan Lund, 1996). Sampled households derive most of their income from non-farm activities (Fafchamps and Lund, 2003).

At the beginning of the survey, each household was asked to identify four individuals on which it could rely in case of need. Most of these individuals are close family members residing in the same village. They constitute the network of insurance partners of each household. Approximately 939 network members were identified by respondents, of which 189 are in households covered by the survey. In subsequent three survey rounds, detailed information was collected on income shocks and all loans and gifts between households. Loan repayment information was also gathered.

Using these data, Fafchamps and Lund (2003) have shown that new loans and gifts play a risk mitigating role. Fafchamps and Gubert (2002) show that loan repayment is contingent on shocks faced by the borrower. Here we examine the structure of risk sharing networks. We begin by estimating equation (I.1). Dependent variable L_{ij} is 1 if household i cites household j as a source of assistance, 0 otherwise. Since i can cite j without j citing i, L_{ij} is directional.

Regressors include income correlation and various measures of geographical and social distance. To control for possible endogeneity, we instrument income correlation using the occupation status of various household members as measured at the beginning of the survey – see Fafchamps and Gubert (2006) for details.[2] Geographical distance is measured by two variables: whether both households i and j reside in the same hamlet or *sitio*; and the difference between i's and j's distance to the nearest road, provided they reside in the same *sitio*. Other dimensions of social distance are included as additional controls, such as occupation, household size, age, health, education, and wealth.[3]

---

[1] By construction, all observations where $j = i$ or $k = l$ are identically zero and hence are omitted. Division of the inner term by 2 corrects for the double counting implied by the simple way we have written the formula.

[2] Income include earnings from jobs held in the last three months, unearned incomes received in the last three months, and earnings from the sale of crops and livestock in the last three months.

[3] Wealth is also instrumented, using birthplace, number of siblings, and inherited wealth.

Results are shown in Table 1.[4] Income correlation has the wrong sign and is not statistically significant. The same conclusion holds for differences in occupation, health, and household composition. Surveyed households do not therefore appear to establish risk sharing links with those in the best position to pool income risk. In contrast, geographical proximity is strongly significant. We also find that respondents cite as source of mutual insurance households that are on average older and richer than themselves.

These results could be misleading if L_{ij} has little or no relationship with actual risk sharing. To eliminate this possibility, we need to show that reported links play a role in the sharing of risk. From the work of Fafchamps and Lund (2003) we know that in the study area gifts and loans serve a risk sharing purpose. We thus estimate a model of the form:

$$(\text{II.1}) \qquad G_{ij}^t = \alpha + \beta_0(z_i - z_j)L_{ij}^{t_1} + \beta_1(z_i - z_j) + \gamma w_{ij} + u_{ij}$$

where G_{ij}^t denotes the value of all gifts (or loans) received by i from j in survey rounds $t = 2$ and $t = 3$. Variables z_i and w_{ij} are as in Table 1. Interaction terms $(z_i - z_j)L_{ij}^{t_1}$ test whether reported links are irrelevant.

Results for gifts and informal loans are reported in Table 2. Geographical proximity variables are strongly significant in both regressions, confirming earlier results. The existence of a link is associated with gifts to unhealthy individuals that are 200 times larger than without a link. Loans from non-farmers to farmers are 27 times larger with a link. These results demonstrate that reported links L_{ij} are not irrelevant.

Results so far emphasize the importance of geographical proximity. Could it be that proximity mitigates moral hazard? In the survey area, loans between friends and relatives are granted without specifying a due date. The borrower is expected to exert diligence but repayment may be delayed in case of shock (Fafchamps and Gubert, 2002). As distance increases, monitoring becomes more difficult. This generates moral hazard and may explain why risk sharing takes place primarily among nearby households.

To investigate this issue, we estimate a duration model in which the dependent variable is the time

---

[4]Since each surveyed households was asked to identify four links, it is not possible to reliably estimate β_2. For this reason, β_2 terms are omitted from the regression. Including them does not affect the result – see Fafchamps and Gubert (2006) for details.

elapsed until a loan is repaid and we test whether informal loans between nearby households are repaid faster. As in Fafchamps and Gubert (2002), we control for shocks, loan size, interest charges, and partial repayment. Household characteristics and village dummies are included as well. Results presented in Table 3 show that informal loans between distant households are paid less rapidly. This is consistent with the existence of moral hazard and may in turn explain why proximity plays such a paramount role in the formation of risk sharing networks.

III. Conclusion

We have examined the determinants of risk sharing links between households. We found that geographic proximity is a major determinant of such links. In contrast, occupation and income correlation are not significant determinants of network links. We also find that reported network links have a strong effect on subsequent gifts and loans and that geographical proximity is associated with faster repayment of informal loans. Taken together our findings suggest that surveyed households do not form links that maximize potential gains from pooling income risk, perhaps because of moral hazard considerations.

Two caveats must be made. First, we were unable to control for family relationships because we do not have the necessary data. But from anthropological accounts, in the study area we know that kinship is correlated with geographical proximity. Results relative to geographical proximity should be interpreted in this light. Second, links with network links with individuals outside the four surveyed villages were not included in the analysis since the focus was on intra-village risk sharing. From the literature on migrations, we nevertheless know that links with migrants play an important role in diversifying income risk across space and occupations. Close kinship with migrants probably serves as a substitute for direct monitoring.

This paper also makes a methodological contribution to the burgeoning empirical literature on economic networks. First we clarified identification issues in dyadic data, especially with respect to directed networks and degree distribution. Second we facilitated inference on network processes by applying the well-known concept of robust standard errors to dyadic data.

References

Coate, Stephen and Martin Ravallion, "Reciprocity Without Commitment: Characterization and Performance of Informal Insurance Arrangements.," *Journal of Development Economics*, 1993, *40*, 1–24.

Conley, Timothy J., "GMM Estimation with Cross-Sectional Dependence.," *Journal of Econometrics*, 1999, *92* (1), 1–45.

Fafchamps, Marcel, "Solidarity Networks in Pre-Industrial Societies: Rational Peasants with a Moral Economy.," *Economic Development and Cultural Change*, 1992, *41* (1), 147–74.

___ **and Flore Gubert**, "Contingent Loan Repayment in the Philippines.," 2002. mimeo.

___ **and** ___ , "The Formation of Risk Sharing Networks.," *Journal of Development Economics*, 2006.

___ **and Susan Lund**, "Risk Sharing Networks in Rural Philippines.," *Journal of Development Economics*, 2003, *71*, 261–87.

Foster, Andrew D. and Mark R. Rosenzweig, "Imperfect Commitment, Altruism and the Family: Evidence from Transfer Behavior in Low-Income Rural Areas.," *Review of Economics and Statistics*, 2001, *83* (3), 389–407.

Ligon, Ethan, Jonathan P. Thomas, and Tim Worrall, "Informal Insurance Arrangements in Village Economies.," *Review of Economic Studies*, 2001, *69* (1), 209–44.

Lund, Susan, "Informal Credit and Risk-Sharing Networks in the Rural Philippines.," 1996. unpublished Ph.D. thesis.

Rosenzweig, Mark R., "Risk, Implicit Contracts and the Family in Rural Areas of Low-Income Countries.," *Economic Journal*, 1988, *98*, 1148–70.

Townsend, Robert, "Risk and Insurance in Village India.," *Econometrica*, 1994, *62* (3), 539–591.

Table 1. Links and income correlation

	coefficient estimate	dyadic t-value
Income correlation		
Correlation of i and j's incomes (a)	1.083	1.44
Geographical proximity		
Same sitio =1	2.647	**8.84**
Difference in distance to road if same sitio	-0.121	**-3.90**
Difference in:		
Dummy=1 if primary occupation of head is farming	0.028	0.23
Number of working members x number of activities	0.003	0.06
Age of household head	-0.010	**-2.52**
Health index 1-4 (1=good health, 4=disabled)	0.027	0.46
Years of education of household head	-0.010	-0.59
Total wealth (a)	-0.113	**-2.37**
Village dummies	included but not shown	
Intercept	-5.995	-15.41
Number of observations	10,264	

Note: the dependent variable=1 if i cites j as source of mutual insurance, 0 otherwise. Estimator is logit. All t-values based on standard errors corrected for dyadic correlation of errors.

(a) instrumented variables -- see text for details.

Table 2. Gifts, loans and networks

	Gifts received (a)		Loans received (a)	
	coefficient estimate	dyadic t-value	coefficient estimate	dyadic t-value
Geographical proximity				
Same sitio =1	0.052	**6.56**	0.038	**4.31**
Difference in distance to road if same sitio	-0.002	**-5.27**	-0.001	**-2.62**
Network dummy x difference in:				
Dummy=1 if primary occupation of head is farming	-0.234	-1.25	-0.256	**-2.32**
Number of working members x number of activities	-0.080	-0.72	-0.004	-0.08
Age of household head	0.005	0.61	0.001	0.18
Health index 1-4 (1=good health, 4=disabled)	0.281	**2.62**	0.011	0.12
Years of education of household head	-0.018	-0.66	-0.009	-0.48
Total wealth (b)	-0.035	-0.28	0.019	0.26
Difference in:				
Dummy=1 if primary occupation of head is farming	0.001	0.51	-0.009	**-2.36**
Number of working members x number of activities	0.000	1.36	0.002	1.09
Age of household head	0.000	-0.44	0.000	-0.52
Health index 1-4 (1=good health, 4=disabled)	0.001	**2.04**	0.000	-0.21
Years of education of household head	0.000	-0.81	0.000	-0.74
Total wealth (b)	0.000	0.51	-0.002	-0.71
Village x time dummies	included but not shown			
Intercept	-0.003	-0.59	-0.007	-1.48
Number of observations	21,184		21,184	

Estimator is least squares. All t-values based on standard errors corrected for dyadic correlation of errors.

(a) Dependent variable in log(value of gift or loan +1); (b) instrumented -- see text for details.

Table 3. Propensity to repay

	Hazard	z-stat.
Difference in distance to road if same sitio between resp. and partner	0.933	**-2.860**
Loan characteristics		
Loan amount	0.718	**-3.560**
Interest factor	0.026	**-4.640**
Share still due	3.308	1.350
Shocks:		
Shock to respondent	0.667	**-1.790**
Household characteristics of respondent:		
Age of household head	1.017	1.300
Last grade completed by head	1.095	1.510
Craft skill dummy	2.080	**1.970**
Permanent wage dummy	0.779	-0.680
Household size	0.817	**-2.950**
Wealth	1.052	0.750
Household characteristics of partner:		
Age of household head	1.006	0.500
Income level	0.898	-0.470
Village dummies	included but not shown	
p-parameter	0.701	**8.070**

Estimator is a duration model.