TESTING UNILATERAL AND BILATERAL LINK FORMATION

MARGHERITA COMOLA
MARCEL FAFCHAMPS

BREAD WORKING PAPER No. 236
JULY, 2009

BREAD

Working Paper

Bureau for Research and Economic Analysis of Development

Testing Unilateral and Bilateral Link Formation*

Margherita Comola † Marcel Fafchamps‡

Paris School of Economics Oxford University

July 2009

Abstract

The literature has shown that network architecture depends crucially on whether links are formed unilaterally or bilaterally, that is, on whether the consent of both nodes is required for a link to be formed. We propose a test of whether network data is best seen as an actual link or willingness to link and, in the latter case, whether this link is generated by an unilateral or bilateral link formation process. We illustrate this test using survey answers to a risk-sharing question in Tanzania. We find that the bilateral link formation model fits the data better than the unilateral model, but the data are best interpreted as willingness to link rather than an actual link. We then expand the model to include self-censoring and find that models with self-censoring fit the data best.

JEL codes: C12; C52; D85

Keywords: network architecture; pairwise stability; risk sharing

*We thank Michael Wooldridge for his useful suggestions and Joachim De Weerdt for making the data available. We have benefitted from useful comments from Yann Bramoullé and from seminar participants at the Paris School of Economics, Oxford University, and the University of Nottingham.

†Paris School of Economics. Email: `comola@pse.ens.fr`

‡Department of Economics, University of Oxford. Email: `marcel.fafchamps@economics.ox.ac.uk`

1. Introduction

It is increasingly recognized that many important economic phenomena, such as goods exchange, information diffusion, and learning, take place within social networks (e.g. Granovetter 1985, Jackson 2008) and that the architecture of these networks can affect the efficiency and equity of the resulting allocation (Vega-Redondo 2006). We also now know that the mechanism through which links are created has a profound influence on the equilibrium architecture of purposely formed networks. In particular, Bala and Goyal (2000) have shown that unilateral and bilateral link formation result in fundamentally different network structures – see also Goyal (2007). This paper proposes a methodology for testing unilateral versus bilateral link formation and presents an empirical illustration of it.

Bilateral link formation refers to situations in which the consent of both nodes is needed for a link to be formed between them; it is a natural assumption for voluntary exchange. Unilateral link formation arises whenever one node can form a link without the express consent of the other; it is a natural assumption for information access networks, e.g., the Internet, and it may also arise in exchange networks when legal or social norms make it unlawful for one party to refuse to trade.[1] We propose a simple methodology for testing whether network data reflect a simple willingness to link or an existing link and, in the latter case, whether this link is generated by an unilateral or bilateral link formation process. Building on the work of Comola (2007), we take pairwise stability as starting point for the estimation process. First introduced by Jackson and Wolinsky (1996), pairwise stability has established itself as a cornerstone equilibrium concept in the study of bilateral link formation processes (Goyal 2007). Comola (2007) has already shown how pairwise stability can be used to construct a bivariate probit estimator of a bilateral link

[1] In many developed countries anti-discrimination laws typically make it unlawful for a retailer to refuse to sell to a specific customer.

formation process. We extend this approach by showing that unilateral link formation can be estimated in a similar fashion: the basic intuition behind our approach is that, in a unilateral link formation framework, the absence of a univariate link is equivalent to a pairwise stable decision by *both* nodes *not* to form a link. Estimates obtained under each regime are compared using a standard non-nested likelihood ratio test first proposed by Vuong (1989), that we adapt in order to take into account the network dependence across residuals.

We illustrate our methodology using data on risk-sharing links from a Tanzanian village named Nyakatoke. In Nyakatoke every individual was asked who in the village would turn to him or he would turn to in times of trouble for help in cash, kind or labor. Aggregating answers at the household level yields a map of reported mutual insurance links among all households in the village. Those reported links are our dependent variable of interest.

A noticeable feature of the data is that, in several cases, household i mentions relying on j for mutual insurance but j does not mention i. This is open to two possible interpretations. The first is that respondents gave the names of households from which they would wish to receive help. In this case, answers are seen as representing not an actual link but a 'willingness to link'. Another possible interpretation is that respondents provided information on actual links but, because of measurement error, answers differ. We test which of these two alternative interpretations best fits the data.

The next issue of interest is whether links are bilateral or unilateral. It would seem natural to expect mutual insurance links to require the agreement of both parties – and this is indeed how the economic literature has modeled informal risk sharing (e.g. Coate and Ravallion 1993, Kocherlakota 1996). It is also possible that social norms make it impossible for villagers to refuse assistance to others. For instance, it may be inconceivable for a son to refuse to assist his father in difficulty. Platteau (1996) argues that many agrarian societies, especially in sub-Saharan Africa, cultivate egalitarian norms, a point that has repeatedly been made by anthropologists and by casual observers alike. Barr and Stein (2008) provide some recent evidence to this effect. In the presence of sharing norms, links would best be seen as unilateral decision. Our contribution is to provide a framework to test whether bilateral or unilateral link formation is most consistent with the responses given by Nyakatoke households. Our empirical findings suggest that, when pegging the bilateral and unilateral against each other, the bilateral link formation model wins. However, both are outperformed by a simple willingness to link model.

We then investigate whether reported willingness to link is affected by self-censoring, an issue that has been raised in the economic literature on dating (e.g. Hitsch, Hortacsu and Ariely 2005, Belot and Francesconi 2006, Fisman, Iyengar, Kamenica and Simonson 2008). Anticipating rejection, respondents may refrain from reporting an intent to link with certain individuals. Respondents may also report links with individuals to whom they would prefer not to link but who they cannot refuse to help. Both cases can be thought of a self-censoring – of willingness to link in the first case, and of unwillingness to link in the second. We test both models against pure willingness to link and find evidence of self-censoring.

The contribution of this paper is primarily methodological. The econometric analysis of social networks is still novel, and there often is a lack of clarity on the implicit assumptions necessary to estimate network models. The ultimate aim of this paper is to shed some light on the way self-declared network data should be interpreted, and how discordant responses should be treated. We find that in our case some models fit the data better than others. Other data may yield different conclusions.

The paper is organized as follows. In Section 2 we provide a conceptual framework and describe our estimating and testing strategy. The data are described in Section 3. Estimation results are discussed in Section 4. Section 5 concludes.

2. Conceptual framework and testing strategy

In this section we begin by presenting the theoretical ground of the different estimation strategies used in the paper. As in Comola (2007) the starting point of our estimation strategy is pairwise stability as defined by Jackson and Wolinsky (1996). We then discuss the important issue of how to draw consistent inference by correcting standard errors for non-independent data. We conclude the section with a discussion of non-nested hypothesis testing with non-independent data.

Formally, for each pair of nodes ("dyad") ij, define $g_{ij}^i = 1$ if i reported a link with j, and 0 otherwise. Similarly define $g_{ij}^j = 1$ if j reported a link with i. Variables g_{ij}^i and g_{ij}^j provide a representation of the data. Their interpretation varies depending on what the data generation process is assumed to be. In subsection (2.1) we consider these data as an indication of willingness to link and we specify the corresponding data generation process. In subsections (2.2) and (2.3) we regard g_{ij}^i and g_{ij}^j as two different measurements of the same actual link g_{ij}. Subsection (2.2) specifies the data generation process if the link formation process is bilateral while subsection (2.3) focuses on the unilateral case.

2.1. Willingness to link

Before introducing the unilateral and bilateral link formation models, it is useful to examine what happens when we interpret the data as indicative of a willingness to link, and not as an existing link. When we do so we cannot draw any inference about the network formation process because the same pattern of willingness to link may result in different equilibrium networks depending on whether the process is bilateral or unilateral.

Here the response variables g_{ij}^i and g_{ij}^j are interpreted as the expression of the willingness of nodes i and j respectively to form the link g_{ij}. Formally, let this network be described by its adjacency matrix $g = [\ g_{ij}]$ with $g_{ij} = 1$ if the link ij exists and $g_{ij} = 0$ otherwise. By a standard abuse of notation, let g_{-ij} denote the network g without the link g_{ij}, that is, with $g_{ij} = 0$. Similarly, let g_{+ij} denote the network with the link g_{ij}, that is, with $g_{ij} = 1$.

The utility that node i derives from network g is written $U_i(g)$. The gain to household i of forming the link g_{ij} is $U_i(g_{+ij}) - U_i(g_{-ij})$. We assume that this gain can be written as a linear function of observables X_{ij} and a zero-mean residual e_{ij}. We thus have:

$$U_i(g_{+ij}) - U_i(g_{-ij}) = X'_{ij}\beta + e_{ij} \tag{2.1}$$

and analogously,

$$U_j(g_{+ij}) - U_j(g_{-ij}) = X'_{ji}\beta + e_{ji} \tag{2.2}$$

Assuming that (e_{ij}, e_{ji}) are jointly normal, equations (2.1) and (2.2) can be estimated as a standard bivariate probit. Since the order in which i and j appear in the data is arbitrary, they must be interchangeable. This implies that the coefficient vector β must be the same in equations (2.1) and (2.2).

2.2. Bilateral link formation

Let us now think of our data as measuring actual links g_{ij}. The set of links define a network g. Given the reciprocal nature of risk-sharing relations (e.g. Coate and Ravallion 1993, Fafchamps and Lund 2003) and the nature of our data, it makes little sense to think of g as a directed network. We therefore assume that $g_{ij} = g_{ji}$: if a risk sharing relationship exists between i and j, by reciprocity it also exists between j and i. Consequently we interpret g^i_{ij} and g^j_{ij} as measures of the actual link g_{ij} – and discrepancies in survey answers g^i_{ij} and g^j_{ij} as due to error of measurement.

In order to specify the data generation process, we have to clarify how links are formed. We first consider the bilateral link formation case. As in Comola (2007) the starting point of our estimation strategy is pairwise stability as defined by Jackson and Wolinsky (1996). Under bilateral link formation, the agreement of both nodes is needed for a link to be formed. This occurs if and only if:

$$\begin{aligned}
\forall g_{ij} &= 1,\, U_i(g_{+ij}) \geq U_i(g_{-ij}) \text{ and } U_j(g_{+ij}) \geq U_j(g_{-ij}) \\
\forall g_{ij} &= 0,\, \text{if } U_i(g_{-ij}) < U_i(g_{+ij}) \text{ then } U_j(g_{-ij}) > U_j(g_{+ij})
\end{aligned}$$

This set of conditions is known as pairwise stability. It implies that:

$$\Pr(g_{ij} = 1) = \Pr\left(U_i(g_{+ij}) \geq U_i(g_{-ij}) \text{ and } U_j(g_{+ij}) \geq U_j(g_{-ij})\right) \tag{2.3}$$

Using (2.1) and (2.2) equation (2.3) is equivalent to:

$$\Pr(g_{ij} = 1) = \Pr\left(e_{ij} \leq X_{ij}'\beta \text{ and } e_{ji} \leq X_{ji}'\beta\right) \tag{2.4}$$

where (e_{ij}, e_{ji}) are jointly normal. Estimating β under the assumption of bilateral link formation thus boils down to maximizing the likelihood function implicitly defined by (2.4).

Model (2.4) has a single dependent variable but two regressing equations. Such model, first proposed by Poirier (1980) and later on used by Comola (2007) to model network formation, is known as a partial observability bivariate probit. This is because the link g_{ij} can be understood as the product of two distinct and unobservable events, i's willingness to form the link ij and j's willingness to form the same link. Let us define these unobservable variables w_{ij}^i and w_{ij}^j such that $w_{ij}^i = 1$ if $e_{ij} \leq X_{ij}'\beta$ and similarly for w_{ij}^j. Under pairwise stability, a link is formed only if both i and j are willing to form it, i.e., $g_{ij} = 1$ iff $w^i_{ij} = 1$ and $w^j_{ij} = 1$ or, more succinctly, iff $w^i_{ij} w^j_{ij} = 1$. The term 'partial observability' comes from the fact that we only observe the product $w^i_{ij} w^j_{ij}$, not each of them separately. That is, whenever a link $g_{ij} = 0$ we can not observe whether one or both nodes are not willing to form it. A partial observability model assumes a smaller amount of knowledge than a standard model in that only uses the information on the equilibrium outcome, which preserves the spirit of a pairwise stable equilibrium.

2.3. Unilateral link formation

An undirected network may also result from a process of unilateral link formation. This corresponds to the situation in which only one side's consent is sufficient for a link to be formed. Put differently, a link does not exist only if both nodes refuse to create it (Goyal 2007). As in the bilateral case, we let w^i_{ij} and w^j_{ij} represent the nodes' unobserved willingness to form link g_{ij}. Under unilateral link formation, $g_{ij} = 1$ whenever either of the two nodes wishes to form a link. It follows that $g_{ij} = 0$ only when both links do not wish to form the link. This simple observation forms the basis of our estimation strategy because it implies that, using a change of variable, the unilateral link formation model can also be estimated as a partial observability model.

To see how this is possible, we begin by noting that:

$$\begin{aligned} \Pr(g_{ij} = 0) &= \Pr\left(U_i(g_{+ij}) < U_i(g_{-ij}) \text{ and } U_j(g_{+ij}) < U_j(g_{-ij})\right) \\ &= \Pr\left(e_{ij} > X'_{ij}\beta \text{ and } e_{ji} > X'_{ji}\beta\right) \end{aligned} \tag{2.5}$$

Let $h_{ij} \equiv 1 - g_{ij}$. We have $h_{ij} = 1$ iff $w^i_{ij} = 0$ and $w^j_{ij} = 0$ or, more succinctly, iff $(1 - w^i_{ij})(1 - w^j_{ij}) = 1$. Estimation can proceed by applying a partial observability bivariate probit to the transformed system:

$$\Pr(h_{ij} = 1) = \Pr\left(-e_{ij} \leq -X'_{ij}\beta \text{ and } -e_{ji} \leq -X'_{ji}\beta\right) \tag{2.6}$$

The dependent variable is still binary, and the partial observability feature ensures that the absence of a link ($h_{ij} = 1$) is interpreted as implying that both nodes do not wish to form that link. As is clear from (2.6), estimated coefficients have the reverse sign compared to (2.4). This is because we are estimating individuals' willingness *not* to form a link.

2.4. Standard errors

Dyadic data can seldom if ever be regarded as made of independent observations; residuals are typically correlated across some observations. This does not invalidate estimation itself: as long as regressors remain uncorrelated with residuals, coefficients can be estimated consistently. But uncorrected standard errors are inconsistent, invalidating inference.

Methods have been proposed to correct standard errors in non-independent data. These methods extend White's formula for robust standard errors to correlation across observations (Conley 1999). For dyadic data, the most pressing concern is the correlation in the residual for observation g_{ij} with those pertaining to all observations involving nodes i and j. This is because i's decision to form a link with j potentially affects his or her decision to form a link with any other node. Fafchamps and Gubert (2007) propose a correction of standard errors that takes care of this form of cross-observation dependence. The formula for the network corrected covariance matrix is of the form:

$$AVar(\widehat{\beta}) = \frac{1}{N-K}(X'X)^{-1}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\frac{m_{ijkl}}{2N}X_{ij}u_{ij}u'_{kl}X_{kl}\right)(X'X)^{-1} \tag{2.7}$$

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. [2] The only structure imposed on the covariance structure is that $E[u_{ij}, u_{ik}] \neq 0$, $E[u_{ij}, u_{kj}] \neq 0$, $E[u_{ij}, u_{jk}] \neq 0$ and $E[u_{ij}, u_{ki}] \neq 0$ for all k but that $E[u_{ij}, u_{km}] = 0$ otherwise. Formula (2.7) was initially developed for linear regressions but it applies to maximum likelihood estimation provided that X_{ij} is everywhere replaced by the corresponding score l_{ij}.

It is conceivable that $E[u_{ij}, u_{km}] \neq 0$ for $i \neq k, m$ and $j \neq k, m$. This would arise, for instance, if i's willingness to form a link with j depends on whether k has a link with m. In this case, formula (2.7) is no longer sufficient to correct standard errors and more cross-terms should be added. Whether this is feasible depends on the data. If the researcher has observations from unlinked sub-populations (e.g., multiple villages), it is possible to allow for arbitrary cross-observation dependence by clustering standard errors at the level of each sub-population (e.g. Arcand and Fafchamps 2008, Barr, Dekker and Fafchamps 2008). In our data, we only have a single village so this option is not available. Bester, Conley and Hansen (2008) has suggested an approach to approximately eliminate bias in standard errors by dividing the data into large blocks and clustering within blocks. Unfortunately this approach requires a large sample, which again is not our case. The standard errors reported in this paper are all based on formula (2.7).

2.5. Non-nested tests

Our aim is to test which one of the models presented above best accounts for the data. To this effect we proceed by pairwise comparisons. Vuong (1989) has proposed a framework for hypothesis testing in non-nested models. Say we want to test which of two alternative, non-

[2] The dependence across g_{ij} and g_{ji} due to the fact that we include both of them in the estimation is automatically corrected for in formula (2.7) since the central term is divided by 2.

nested models k and m fit the data best. Let $M = N(N-1)$ be the total number of dyadic observations. The original form of the Vuong test statistic is

$$V = \frac{M^{-1/2}LR(k,m)}{\hat{\omega}} \xrightarrow{d} N(0,1)$$

where $LR(k,m) \equiv L^k - L^m$ is the log of the likelihood ratio statistic and:

$$\hat{\omega}^2 = \frac{1}{M}\sum_{ij=1}^{M}\left[\log\frac{l_{ij}^k}{l_{ij}^m}\right]^2 - \left[\frac{1}{M}\sum_{ij=1}^{M}\log\frac{l_{ij}^k}{l_{ij}^m}\right]^2$$

where l_{ij}^k and l_{ij}^m are the observation-specific scores for each model k and m. This test can be implemented more simply by regressing the difference between the scores on a constant:

$$l_{ij}^k - l_{ij}^m = \alpha_{km} + v_{ij}^{km}$$

The t-value on the constant α_{km} is the Vuong statistic that tests whether model k outperforms model m. In our case, for inference to be valid we need to correct the standard error of the constant $\widehat{\alpha}_{km}$ for possible cross-dependence across observations. We do this by applying formula (2.7).

3. The data

To illustrate our estimation and testing strategy we use survey data from a village community named Nyakatoke in the Buboka Rural District of Tanzania, at the west of Lake Victoria. The village is mainly dependent on farming of bananas, sweet potatoes and cassava for food, while coffee is the main cash crop. The community is composed by 600 inhabitants, 307 of which are adults, for a total of 119 households interviewed in five regular intervals during 2000. This dataset is ideal for our purpose because it is a census covering all 119 households in the village.[3] The data include information on households' demographics (composition, age, religion, education), wealth and assets (land and livestock ownership, quality of housing and durable goods), income sources and income shocks, transfers and network relations.

Each adult respondent was asked: *"Can you give a list of people from inside or outside of Nyakatoke, who you can personally rely on for help and/or that can rely on you for help in cash, kind or labor?"*. Aggregated at the level of each household, the responses to this question constitute variables g_{ij}^i and g_{ij}^j. In other words, $g_{ij}^i = 1$ if an adult member of household i mentions an adult member of household j in their response to the above question. We explain in detail below how survey responses g_{ij}^i and g_{ij}^j are used to build the dependent variables. Nyakatoke data have been analyzed by De Weerdt and Dercon (2006) and De Weerdt and Fafchamps (2007). These authors have shown that reported mutual insurance links g_{ij}^i and g_{ij}^j are strong predictors of subsequent loans and gifts, and that linked households give and receive much more from each other in times of illness.

Given the cultural context, it is not obvious how to interpret Nyakatoke villagers' responses to the risk sharing link question. One possible interpretation is that responses represent the respondent's desire to establish a link. This interpretation is particularly appealing when the responses are discordant, that is, when $g_{ij}^i \neq g_{ij}^j$. It is nevertheless possible that discordant responses as due to measurement error and that the data describe, albeit with some error, actual links between villagers.[4]

The process by which links between villagers are formed can be bilateral or unilateral. Much

[3] Everyone in the village agreed to participate in the survey, but there are some missing data for 4 households.

[4] Independently of whether the underlying network follows a bilateral or unilateral link formation process, it is necessary to decide how to treat discordant responses in the estimation itself. If respondents forget to mention some of their risk-sharing partners because they are involved in too many links to recall them all, we should treat any discordant pair as an existing link, i.e, as $g_{ij} = 1$. Doing so implicitly assumes that the main form of measurement error is omission, i.e., that respondents do not mention someone as a risk sharing partner unless the expectation of reciprocity is strong. Alternatively, discordant responses may arise because one of the two respondents mistakenly reported a link where none exists, i.e., discordant cases correspond to $g_{ij} = 0$. Without information on individual intent, we cannot disentangle the two.

of the economic literature on informal risk sharing in developing countries has assumed that households willingly enter in such arrangements (e.g. Kimball 1988, Coate and Ravallion 1993). Applied to social networks, this approach implicitly assumes that mutual insurance links follow a bilateral process. In contrast, much of the anthropological literature has emphasized the difficulty for individuals to abstract themselves from the moral and social obligation to assist others in need (e.g. Scott 1976, Platteau 1996). This point has been made by a number of economists as well, notably those studying remittance flows (e.g. Lucas and Stark 1985, Azam and Gubert 2006). Anderson and Baland (2002) provide evidence that individuals living in Kenyan slums put money in rotating savings and credit associations (ROSCAs) to avoid claims on their resources by spouse and relatives. Ambec (1998) and Banerjee and Mullainathan (2007) take these observations as starting point to model the saving behavior of poor households. This line of reasoning implies an unilateral mechanism of link formation. Testing these alternative data generation processes is the objective of this paper.

Because our dataset is small, we are limited in the number of regressors we can credibly include in the analysis. The covariates that appear in the regressions should be seen as illustrative of the kind of variables one may want to include in an analysis of this kind. What matters most for our purpose is whether conclusions regarding bilateral or unilateral link formation are robust to alternative choices of regressors. If we include too few regressors, the alternative models we wish to test will not account for much of the variation in the data, and we will not be able to tell them apart. Ultimately, all we want is a list of regressors that enables us to robustly test the models against each other.

In this section we present our preferred list of regressors. At the end of the paper we discuss whether our results vary with alternative regressors. The covariates X_{ij} used in the regression analysis fall into three categories: variables that reflect the attractiveness of the potential partner

j; variables proxying for homophyly, that is, the desire to link with similar households; and variables controlling for i's need to link.

Two regressors capture attractiveness. The first one, O_{ij}, is the overlap in productive activities between i and j. It is calculated as:

$$O_{ij} = \sum_{a=1}^{7} L_{ai} L_{aj}$$

where L_{ai} is the share of total time spent by adult members of household i in activity a.[5] Each L_{ai} is constructed using information collected on time use in seven broad income generating categories. Households whose productive activities overlap are expected to have more correlated incomes. Since less correlated incomes generate more opportunities for risk pooling, households with less overlap in activities with household i are in principle more attractive risk sharing partners (. Fafchamps and Gubert 2007, De Weerdt and Fafchamps 2007) We therefore expect O_{ij} to have a negative sign.

We also control for the in-degree P_j^i of j, omitting any link between i and j to avoid spurious correlation. We think of P_j^i as a proxy for various unobservable characteristics – e.g., sociability, generosity, moral sense – that make j an attractive partner for many villagers. It is reasonable to assume that, other things being equal, all households in our sample would prefer to be linked to popular households. Of course, popular households may not wish to link to everyone, since this would mean assisting the entire village.[6] They may therefore be unwilling to link with

[5] In the survey each adult individual mentions the productive activities he or she is involved into. These activities are divided in seven categories: casual labor, trade, crops, livestock rearing, assets, processing of agricultural products, and other off-farm work. Individuals can report multiple activities but are not asked about the relative importance of each activity. We have therefore no alternative but to assign equal weight to all listed activities. L_{ai} is calculated as follows. Say household i has n members, m of which report working full time in a and k report a and one other activity. Then $L_{ai} = \frac{1}{n}(m + \frac{k}{2})$. Individuals who do not report any involvement in an income generating activity are omitted from the calculation. Five households in the sample report no active member.

[6] For a formalization of this idea, see for instance Vandenbossche and Demuynck (2009) 's model of risk sharing network formation. See Ellsworth (1989) for a detailed description of mutual assistance flows in a Burkinabe village, and of the role played by one 'holy man' as center of a village-wide redistribution network. There is no

unpopular households, a feature that is captured by pairwise stability and the bilateral link formation model.

A second set of regressors seeks to control for homophyly, that is, the desire to link with similar or proximate households. The literature has shown that social ties depend to a large extent on social and geographical proximity (e.g. Fafchamps and Gubert 2007, De Weerdt and Fafchamps 2007). To control for geographical proximity, we introduce a dummy that takes value one if i and j are neighbors, that is, live less than 100 meters apart.[7] Blood ties are controlled for using a kinship dummy that takes value one if i and j – or members of their household – are related.[8] Constructing this variable is particularly demanding in terms of data collection, a strong point of the Nyakatoke dataset. We also include a religion dummy taking the value of one if i and j have the same religion.[9]

To capture similarity in social status, we include as regressor the absolute difference in total wealth (computed as the sum of land and livestock) $|w_j - w_i|$ between i and j.[10] If i prefers to link with someone of similar wealth, the coefficient of $|w_j - w_i|$ should be negative. To avoid spurious results, we borrow from Fafchamps and Gubert (2007) and include the sum of wealth $(w_i + w_j)$ to control for the possibility that wealthier individuals have, on average, more links.

The third set of regressors includes factors likely to make household i more interested in forming links. Some respondents report more links than others. This may be because they are pro-social or anti-social. To control for i's proclivity for forming – or reporting – mutual insurance links with others, we include i's out-degree as regressor, omitting any link with j. Wealthy households are less in need of mutual insurance. To capture this possibility, we include

such central person in our village, however.

[7] Slight variation in the cutoff distance does not affect our main results.

[8] This includes parents/children, siblings, cousins, uncle/aunt/niece/nephew, grand-parents/grand-children, and other blood ties.

[9] Catholic, or Protestant, or Muslim – 41%, 39% and 20% of the village population respectively.

[10] Data on land was collected in acres, but transformed in monetary equivalent using a conversion rate of 300000 tzs for 1 acre. This reflects the average local price in 2000, the time at which the data were collected.

a dummy which is equal to one if household i in top 25% wealth percentile in the village. For similar reasons, we also include the number of adult members of household i. As De Weerdt and Fafchamps (2007) show, informal transfers in Nyakatoke respond to health shocks. Since they pool labor resources, larger households should find it easier to deal with health shocks than smaller ones – and hence are less in need of forming mutual insurance links with other villagers (Binswanger and McIntire 1987).

Descriptive statistics are reported Table 1. The first and second panels of the table present dichotomous and continuous variables, respectively. In the dataset there are 119 households, which make 119*118=14042 dyads in total. We see from the Table that the proportion of pairs for which g_{ij}^{i} or $g_{ij}^{j} = 1$ is 7%. The proportion of discordant responses is large. Around one third of household pairs share the same religion. Wealth and the other continuous regressors display a healthy amount of variation in the data. Some regressors were rescaled to facilitate estimation.[11]

[11] To minimize convergence problems that arise when using bivariate probit with partial observability.

Table 1: Descriptive statistics (n=14042)

g^i_{ij}, g^j_{ij}	$g^i_{ij} = g^j_{ij} = 1$			280
	$g^i_{ij} \neq g^j_{ij}$			700
	$g^i_{ij} = g^j_{ij} = 0$			13062
Neighbors	distance ij <100 m			4%
Same family	ij have blood ties			6%
Same religion	ij same religion			35%
Rich respondent	i in top 25th %			25%
continuous variables	**mean**	**min**	**max**	**sd**
Overlap in productive activities O_i	0.22	0	1	0.162
In-degree of household j (/10) (**)	0.52	0	2.30	0.445
$\lvert w_j - w_i \rvert^{(*)}$	0.44	0	2.80	0.524
$w_i + w_j^{(*)}$	0.91	0	5.56	0.678
Out-degree of household i (/10) (**)	0.52	0	1.90	0.304
No. adult members in household i (/10)	0.26	0.1	0.90	0.131

(*) 1 unit corresponds to 1 million Tanzanian Shillings.

(**) excluding the ij link

4. Empirical results

4.1. Model estimation

We now estimate and compare the three models presented in (2.1), (2.2), and (2.3). Each model includes the list of X_{ij} regressors presented in Table 1. For each set of results the z-values reported in the last column are based on dyadic standard errors corrected using formula (2.7).

We begin by reporting the estimation results obtained when we assume that responses to the risk sharing question capture willingness to link, as explained in subsection (2.1). Since, by our notation, $g_{ij}^j = g_{ji}^j$, equations (2.1) and (2.2) can be estimated by stacking g_{ij}^i and g_{ji}^j observations and applying probit. Coefficients estimates are reported in Table 2. They suggest that respondents prefer to link with popular households who live nearby, are related, and share a similar level of wealth. The coefficient of $w_i + w_j$ is positive and marginally significant, suggesting that willingness to link is higher among wealthy households. Other regressors are not significant.

Table 2: Willingness to link

Regressor	coefficient	dyadic z
Overlap in activities O_{ij}	-0.194	-0.85
Popularity P_j^i	0.508	7.71***
Neighbor dummy	0.760	5.17***
Blood ties dummy	0.987	5.86***
Same religion dummy	0.169	1.31
$\lvert w_j - w_i \rvert$	-0.250	-2.35**
$w_i + w_j$	0.249	1.74*
Out-degree of i	0.287	1.65*
Rich dummy of i	-0.004	-0.04
Nber adult members of i	0.105	0.26
Intercept	-2.659	-15.99***

We then turn to the bilateral link formation model. We experimented with three versions of the model. All assume the same data generation process (2.4) but are based on different assumptions on the meaning of discordant dyads. In the first version, a link between i and j is assumed not to exist whenever the pair is discordant. The discrepancy between g_{ij}^i and g_{ij}^j is assumed to come from over-reporting. With this assumption, $g_{ij} = g_{ij}^i g_{ij}^j$. In the second version, a link is taken to exist if either i or j mentions it. This is akin to assuming that discrepancies come from omission. Here $g_{ij} = 1$ whenever g^i_{ij} or $g^j_{ij} = 1$. The third version is agnostic with respect to the source of discrepancy. In this version, g^i_{ij} are g^j_{ij} are assumed to be two distinct (but non-independent) observations of the same true link g_{ij}, observed with error. A discordant pair thus corresponds to a 50% chance of over-reporting ($g_{ij} = 0$) and a 50% chance of under-reporting ($g_{ij} = 1$).[12]

We experimented with all three versions of the model. All three versions yield parameter estimates that are by and large comparable. But the Vuong test cannot be used to compare the first two versions to the willingness to link model because they ultimately use different dependent variables. For this reason, we focus on the results from the third version.

Results are presented in Table 3. Several coefficient estimates are similar to those reported in Table 2. Popularity P^i_j remains strongly significant. We continue to find plenty of evidence of homophyly. Overlap in activities O_{ij} is now marginally significant with the anticipated sign, suggesting a desire to link with individuals who have a different income profile.

[12] Since in this case the dependent variables for observations g_{ij} and g_{ji} differ, we constraint the coefficients to be the same for the two individual-level bivariate probit equations.

Table 3: Bilateral link formation

Regressor	coefficient	dyadic z
Overlap in activities O_{ij}	-0.065	-1.92*
Popularity P_j^i	0.136	2.52**
Neighbor dummy	0.213	3.20***
Blood ties dummy	0.316	3.95***
Same religion dummy	0.042	1.95*
$\lvert w_j - w_i \rvert$	-0.057	-1.81*
$w_i + w_j$	0.051	1.33
Out-degree of i	0.037	1.05
Rich dummy of i	0.021	0.71
Nber adult members of i	0.213	2.30**
Intercept	-0.271	-1.74*
arc $\tan(\rho)$	-1.894	-3.59***

Next we present the results assuming that the data were generated by the unilateral link formation model (2.6). As explained in subsection (2.3), we transform household responses g_{ij}^i and g_{ij}^j into the equation-level dependent variables $h_{ij}^i \equiv 1 - g_{ij}^i$ and $h_{ij}^j \equiv 1 - g_{ij}^j$. As for bilateral link formation, we estimate three versions of the model: one in which discordant responses are regarded as over-reporting; one in which they are regarded as under-reporting; and one combining the other two. We focus on the mixed model since the other two are not directly comparable with the willingness to link model.

Results for the unilateral link formation model are reported in Table 4. To facilitate comparison with Table 3, we report estimated coefficients $\widehat{\beta}$ directly, which means inverting the sign of the coefficient estimates obtained from estimating (2.6) with partial observability bivariate probit. In terms of coefficient estimates, results are similar to those reported in Table 3. Popularity P_j^i and activity overlap O_{ij} are both significant with the anticipated sign. Homophyly variables are all strongly significant while i's characteristics are not.

Table 4: Unilateral link formation

Regressor	coefficient	dyadic z
Overlap in activities O_{ij}	-0.213	-2.04**
Popularity P_j^i	0.412	7.83***
Neighbor dummy	0.706	9.88***
Blood ties dummy	0.928	9.54***
Same religion dummy	0.155	2.83***
$\lvert w_j - w_i \rvert$	-0.198	-3.69***
$w_i + w_j$	0.171	1.80*
Out-degree of i	0.161	0.97
Rich dummy of i	0.107	1.17
Nber adult members of i	0.564	1.50
Intercept	-2.862	24.64***
arc tan(ρ)	0.628	3.47***

4.2. Specification tests

We now turn to the main object of the paper, which is to compare the performance of the different models in accounting for the data. As explained in Section 2, we proceed by pairwise comparisons, adapting the non-nested Vuong test to the dyadic structure of the data. To compare two models k and m we calculate, for each observation ij, the log-likelihood contributions (or score) under the two models and we regress the difference $l_{ij}^k - l_{ij}^m$ on a constant, correcting the standard errors using formula (2.7). The t-value of the constant is the corrected Vuong test.

Since the distribution of the Vuong test is asymptotically normal, the relevant critical value for a 5% level of significance is 1.96. Note that the test works in two directions: if $t > 1.96$ this means that model k is to be preferred to model m; in contrast, if $t < -1.96$ this means that model m is to be preferred to model k. For values of t between -1.96 and 1.96 the test is inconclusive – both models fit the data equally.

Table 5 reports the result of the pairwise comparisons between the willingness-to-link model and the other two. When the bilateral and unilateral models are compared to each other, the bilateral model is found superior. But the findings unambiguously shows that the willingness-to-link model fits the data best.

Table 5: Vuong tests

Model k	Model m	Vuong test	Best fit
bilateral	unilateral	2.28**	bilateral
willingness to link	bilateral	2.34**	willingness to link
willingness to link	unilateral	3.34***	willingness to link

4.3. Self-censoring

Our results imply that responses given to the mutual insurance question are more a reflection of willingness to link than evidence of an actual link. Yet De Weerdt and Fafchamps (2007) have shown that these responses are strong predictors of gifts and transfers reported in subsequent survey rounds. Fafchamps and Gubert (2007) report similar findings with data collected in the Philippines using a similarly worded risk sharing question.[13] This makes us suspect that responses to the mutual insurance question may actually be more than just willingness to link.

One possibility is that respondents did not report households with whom they would like to share risk but who are likely to turn them down. Self-censoring has been discussed in the eco-

[13] In fact the Philippines question was used as template for the Tanzania survey.

nomic literature on dating. In that literature, the researcher typically has access to information on willingness to date – e.g., answers to a direct question following speed dating interviews (e.g. Belot and Francesconi 2006, Fisman et al. 2008), or emails sent to prospective partners on an internet dating site (Hitsch et al. 2005). In both cases, the authors worry that respondents may fail to list or contact desirable partners who are unlikely to accept them.[14]

A similar kind of self-censoring may also be at work in our data. In particular, household i may have liked to share risk with household j but expected j to refuse, and so failed to mention j as possible mutual insurance link. This corresponds to an alternative data generating process in which j can veto a link that i wants.

Such data generating process can be represented as follows. Let g_{ij}^i be i's report of whether a link to j exists. This report is now thought of as made of two parts: (1) i's willingness to link with j, which we denote w_{ij}; and (2) i's expectation of whether the link would be accepted by j, which we denote e_{ij}. Expectation e_{ij} is thought of as made of two intermingled parts: j's willingness to link with i and j's inability to refuse a link with i even though j does not want to link with i. We observe $g_{ij}^i = 1$ if both $w_{ij} = 1$ and $e_{ij} = 1$. We observe $g_{ij}^i = 0$ if either $w_{ij} = 0$ or $e_{ij} = 0$ or both.

To illustrate what we have in mind, imagine that unpopular households wish to link to popular households ($w_{ij} = 1$) but popular households never wish to link with unpopular households ($w_{ji} = 0$). Yet popular households cannot refuse to help some of the unpopular ones, e.g., members of their church. In that case, unpopular household i will report $g_{ij}^i = 1$ with popular household j whenever i expects that j will not refuse to help ($e_{ij} = 1$) because of social norms

[14] Self-censoring has also been discussed in the context of matching models in which individuals can only rank a subset of their possible choices (e.g., schools or jobs). In such models, it is optimal for low ranked individuals not to 'waste' limited slots on options they are unlikely to get.

or altruism. Formally we have:

$$\Pr(g^i_{ij} = 1) = \Pr(w_{ij} = 1 \text{ and } e_{ij} = 1) \tag{4.1}$$

with

$$\begin{aligned} \Pr(w_{ij} &= 1) = \beta x_{ij} \\ \Pr(e_{ij} &= 1) = \gamma x_{ji} \end{aligned}$$

Model (4.1) can be estimated using bivariate probit with partial observability. The only difference with model (2.4) is that we no longer impose that coefficients be the same in the two equations. Instead, we now estimate different coefficients β and γ for the two equations. As before, the estimator allows for non-independence between $\Pr(w_{ij} = 1)$ and $\Pr(e_{ij} = 1)$ (for instance because of unobserved individual effects common to both). Model (4.1), which we call the 'vetoed link' model, can be seen as a refined version of willingness to link which incorporates expectations about the potential partner's likely behavior.

Estimation results for the vetoed link model are presented in Table 6. Coefficient estimates for the w_{ij} equation have the same interpretation as before. Coefficient estimates for the e_{ij} equation capture two kinds of effects: j willingness to link with i, and j capacity to veto a link with i. If the data generating process behind g^i_{ij} is bilateral link formation, we should observe $\beta = \gamma$. This corresponds to the case where i perfectly internalizes the rejection behavior of others, in which case g^i_{ij} is a measurement of the true bilateral network g_{ij} (possibly with measurement error). In contrast, if g^i_{ij} only represents i's willingness to link, then we should observe $\gamma = 0$. If $\gamma < 0$ for a given regressor x_{ji}, this implies that x_{ji} is associated with a lower e_{ij} and thus a higher likelihood of 'veto' by j. A $\gamma > 0$ in contrast implies that the corresponding

x_{ji} makes it harder for j to refuse to assist i.

We see that estimated coefficients in the w_{ij} equation are somewhat similar in terms of magnitude and statistical significance to those reported in earlier regressions: popularity P_{ij}^{i} is again strongly significant, and so are geographical proximity and a shared religion. The out-degree of i (omitting the ij link) is also statistically significant. In contrast, coefficients in the e_{ij} regression are quite different from those reported for the w_{ij} equation. This confirms that g_{ij}^{i} reports are unlikely to reflect a bilateral link formation process. Only three coefficients are statistically significant: the kinship dummy, j's out-degree, and the size of j's household. This means that kin are less likely to veto a link but the smaller j's household is and the larger j's out-degree, the more likely j will veto a link with i. This suggests that larger households have a duty to care for others, possibly because their size makes them better able to self-insure – and thus to assist others.

Table 6. Vetoed links model

	w_{ij} equation			e_{ji} equation	
Regressor	coefficient	dyadic z	Regressor	coefficient	dyadic z
Overlap in activities O_{ij}	0.281	1.50	Overlap in activities O_{ij}	-2.389	-1.23
Popularity P_j^i	0.462	7.32***	Popularity P_i^j	-0.034	-0.16
Neighbor dummy$_{ij}$	0.643	7.34***	Neighbor dummy$_{ij}$	0.454	0.29
Blood ties dummy$_{ij}$	0.963	5.73***	Blood ties dummy$_{ij}$	-0.306	-0.42
Same religion dummy$_{ij}$	0.205	2.97***	Same religion dummy$_{ij}$	-0.308	-1.26
$\|w_j - w_i\|$	-3.233	-1.06	$\|w_j - w_i\|$	7.914	0.36
$w_i + w_j$	0.271	1.09	$w_i + w_j$	-0.513	-0.60
Out-degree of i	0.259	3.59***	Out-degree of j	-0.845	-1.99**
Rich dummy of i	0.003	0.03	Rich dummy of j	0.103	0.52
Nber adult members of i	0.148	0.88	Nber adult members of j	4.672	2.25**
Intercept	-2.597	-12.85***	Intercept	2.700	2.54**
arc tan(ρ)	-1.999	-4.00***			

By analogy with Section 2, it is also possible to define the 'dual' analogue of the vetoed link model. In this model, i reports his *un*willingness to link with j, except in cases when j can impose a link with i. This implies that i reports $g_{ij}^i = 1$ whenever i expects j to impose a link on i, even if i is not keen to link with j. In this model, we have:

$$\Pr(g_{ij}^i = 0) = \Pr(w_{ij} = 0 \text{ and } e_{ij} = 0)$$

with

$$\Pr(w_{ij} = 0) = -\beta x_{ij}$$
$$\Pr(e_{ij} = 0) = -\gamma x_{ji}$$

This model is the generalized equivalent of the unilateral link formation with $h_{ij} = 1 - g_{ij}$. It can be estimated in a fashion similar to (4.1), but the interpretation is slightly different. Here i reports a missing link ($g^i_{ij} = 0$) if i does not want to link *and* i expects that j cannot impose a link on i. But i reports a link whenever either i wishes to link with j *or* i expects that j can impose a link. We call this model the 'forced link' model since j can force a link that i does not want.

Regression estimates are shown in Table 7. As we did for Table 4, we report estimated coefficients $\widehat{\beta}$ and $\widehat{\gamma}$ directly, i.e., we invert their sign to facilitate comparison with Table 6. Interpretation of the coefficients of the w_{ij} equation is as before. In the case of the e_{ij} equation, unilateral link formation would imply $\gamma = \beta$. This would arise for instance if i fully internalizes the unilateral link formation equilibrium. If all $\gamma = 0$, g^i_{ij} is consistent with pure willingness to link. A $\gamma > 0$ means that the x_{ji} variable raises the likelihood that, in i's opinion, j's can force a link on i.

Coefficient estimates for the w_{ij} equation are fairly similar to those reported earlier in Table 2, except that $w_i + w_j$ is not marginally significant anymore and i's out-degree and i's rich dummy are now statistically significant. Coefficient estimates for the e_{ij} equation are different in sign and magnitude from those of the w_{ij} equation, a result that is consistent with our earlier finding that g^i_{ij} is not consistent with a unilateral link formation process. Several regressors have a significant coefficients in the e_{ij} equation, indicating factors that make it more (or less) likely that j be willing and able to force a link onto i. Geographical proximity and blood ties appear with a strongly significant positive coefficient, indicating that it is difficult to deny assistance to kin and neighbors. The negative coefficient for i's popularity P_i^j indicates that the more popular i is, the less likely it is that j can impose a link onto i.

Table 6. Forced links model

	w_{ij} equation			e_{ji} equation	
Regressor	coefficient	dyadic z	Regressor	coefficient	dyadic z
Overlap in activities O_{ij}	-0.162	-0.62	Overlap in activities O_{ij}	-0.326	-0.98
Popularity P_j^i	0.707	7.31***	Popularity P_i^j	-0.496	-2.13**
Neighbor dummy$_{ij}$	0.436	2.07**	Neighbor dummy$_{ij}$	1.023	8.11***
Blood ties dummy$_{ij}$	0.788	4.86***	Blood ties dummy$_{ij}$	1.218	7.82***
Same religion dummy$_{ij}$	0.131	1.20	Same religion dummy$_{ij}$	0.219	1.47
$\|w_j - w_i\|$	-1.776	-1.85*	$\|w_j - w_i\|$	-5.052	-3.25***
$w_i + w_j$	0.172	1.48	$w_i + w_j$	0.130	0.72
Out-degree of i	0.534	4.15***	Out-degree of j	-0.509	-1.81*
Rich dummy of i	0.168	2.05**	Rich dummy of j	0.277	1.91*
Nber adult members of i	0.270	0.97	Nber adult members of j	0.925	2.75***
Intercept	-3.112	-10.82***	Intercept	-2.247	-10.12***
arc tan(ρ)	1.081	1.29			

While these results are interesting in their own right, our primary interest is whether either of these models fits the g_{ij}^i data better than the pure willingness to link model. The Vuong test for the vetoed link and forced link models are presented in Table 8. Results show that both significantly dominate the willingness to link model.[15] This is consistent with the idea that reported links g^i_{ij} are best interpreted as self-censored willingness to link. The last row of the Table also shows that we cannot distinguish between the vetoed link and forced link model: although the vetoed link provides a slightly better fit, the difference is not statistically significant. This is not entirely surprising given that the two models are fairly similar in terms of the underlying data generation process.

Table 8: Vuong test – vetoed links and forced links

model k	model m	Vuong test	best fit
vetoed links	willingness to link	3.47***	vetoed links
vetoed links	bilateral	3.58***	vetoed links
vetoed links	unilateral	4.05***	vetoed links
forced links	willingness to link	2.65**	forced links
forced links	bilateral	3.27***	forced links
forced links	unilateral	3.94***	forced links
vetoed links	forced links	0.70	both

5. Robustness analysis

To ascertain whether our findings are sensitive to the choice of regressors, we reestimate all models using different sets of explanatory variables. Results, not shown here to save space, indicate that when the included regressors have little predictive power – e.g., when the number of regressors is small – the comparison between models tends to be less conclusive. This is hardly

[15] For comparison purposes, we also computed a standard likelihood ratio test to compare the vetoed link and bilateral link formation models since the latter is nested in/is a restricted form of the former. The value of the test is 87, which is well above the 1% critical value of 20.1 for a χ^2 distribution with 8 degrees of freedom. This confirms that the vetoed link regression dominates the bilateral link formation model. A similar comparison between the forced link and the unilateral link formation model yields a test statistic of 124, which clearly shows that the forced link model dominates. Neither of these test statistics corrects for dyadic correlation across observations, however.

surprising as the problem is common to all non-nested tests. The models are compared in terms of their ability to account for the data. When regressors have little predictive power, all models do rather poorly in predicting observed g_{ij}^i and hence cannot be distinguished.

In most situations eliminating one or more regressors leaves the models' ranking unchanged but turns some pairwise comparison inconclusive. Dropping some regressors can nevertheless change the models' ranking. In particular, if we drop the in-degree P_j^i of j and/or the out-degree of i, non-nested comparisons indicate that willingness to link ranks lower than bilateral or unilateral link formation. Both self-censored models continue to dominate, however.

Finally, it worth mentioning that we have encountered the convergence difficulties that partial observability models are known for. Using a stepping algorithms for non-concave regions of the likelihood function alleviates part of the problem, but occasionally convergence may not be achieved. Also, in our experience the partial observability bivariate probit model is particularly sensitive to the choice of ad-hoc initial values and to collinearity, which in some extreme cases may result in the impossibility of computing standard errors.

6. Conclusion

The theoretical literature on networks has shown that the nature of the link formation process – e.g., whether unilateral or bilateral – has a strong effect on the resulting network architecture. In this paper we develop a methodology to test whether network data reflect a simple willingness to link or an existing link and, in the latter case, whether this link is generated by an unilateral or bilateral link formation process. Taking the equilibrium concept of pairwise stability as starting point, we propose a methodology to compare bilateral and unilateral processes. Central to this methodology is the observation that unilateral link formation requires that both nodes wish *not* to form a link for the link not to exist. This formal similarity between the bilateral and unilateral link formation processes allows us to model them both as partial observability models and to compare them with the appropriate non-nested likelihood test.

We illustrate this methodology with data on informal risk-sharing networks in a Tanzanian village. The data is particularly well suited for our purpose because it covers all households in the community, and because the respondents are asked to enumerate all their network partners. The information provided by respondents is nevertheless open to several interpretations.

One possible interpretation is that responses capture an actual link. This interpretation is consistent with the observation made by De Weerdt and Fafchamps (2007) and Fafchamps and Lund (2003) who have shown that risk sharing links reported by survey respondents strongly predict subsequent inter-household transfers. It however remains unclear what process generated these links. The development literature is uncertain as to whether risk sharing networks should be seen as entirely voluntary, or whether social norms impose an element of moral or social pressure making it difficult for households to refuse helping others. If risk sharing is voluntary, link formation can be modelled as bilateral; if risk sharing is imposed by social norms, unilateral link formation is a more appropriate representation of the data generating process. Using a Vuong non-nested test, we find that the bilateral link formation model fits the data better than a unilateral one.

Another possible interpretation is that responses to a question about mutual insurance links capture the respondent's willingness to link, not an actual link. This may explain the large proportion of discordant answers whereby i reports a link with j although j does not report a link with i. We test a willingness-to-link model against the bilateral and unilateral link formation models and find that willingness to link fits the data best. This finding, however, is reversed if we drop the in-degree of j or the out-degree of i as regressors.

We then expand the data generating process to allow for self-censoring by respondents. We investigate two forms of self-censoring. In the first one, which we call the vetoed link model, we allow respondents to form expectations about the other party's ability to refuse a link. In the second, which we call the forced link model, respondents anticipate that they may be unable to refuse certain links. We find that both models dominate the other three models, suggesting that self-censoring is present. But we are unable to distinguish between the vetoed and forced link models – both fit the data equally well.

While promising, the approach presented here suffers from a number of shortcomings. Test results are ultimately predicated on the assumption that the regressors used in the estimation are reasonable predictors of willingness to link. In the case of the self-censoring models, identification rests on exclusion restrictions that cannot be tested without additional data. The contribution of this paper should therefore be seen as primarily methodological. Stronger inference could be achieved if, in addition to information about links, the survey contained more direct evidence on respondents' willingness to link (or de-link) with other households. Should such data become available together with objective information on social links, the methodology presented here can yield a stronger test of bilateral versus unilateral link formation.

The methodology used here can potentially be expanded to deal with more complex equilibrium concepts, such as the coalition-proof equilibria discussed in Genicot and Ray (2003). To test whether coalition-proofness constraints are binding, one would need to expand the likelihood function to include other voluntary participation constraints. How this could be implemented in practice remains unclear. This is left for future research.

References

Ambec, S. (1998), A Theory of African's Income-Sharing Norm with Enlarged Families. (mimeograph).

Anderson, S. and Baland, J.-M. (2002). “The Economics of Roscas and Intrahousehold Resource Allocation.”, *Quarterly Journal of Economics*, 117(3):963–95.

Arcand, J.-L. and Fafchamps, M. (2008), Matching in Community-Based Organizations. (mimeograph).

Azam, J. P. and Gubert, F. (2006). “Migrants' Remittances and the Household in Africa: A Review of Evidence.”, *Journal of African Economies*, 15(Supplement 2):426–62.

Bala, V. and Goyal, S. (2000). “A Non-Cooperative Model of Network Formation.”, *Econometrica*, 68(5):1181–1229.

Banerjee, A. V. and Mullainathan, S. (2007), Climbing Out of Poverty: Long Term Decisions under Income Stress. (mimeograph).

Barr, A., Dekker, M. and Fafchamps, M. (2008), Risk Sharing Relations and Enforcement Mechanisms. (mimeograph).

Barr, A. and Stein, M. (2008), Status and Egalitarianism in Traditional Communities: An Analysis of Funeral Attendance in Six Zimbabwean Villages. (mimeograph).

Belot, M. and Francesconi, M. (2006), Can Anyone Be The One? Evidence on Mate Selection from Speed Dating., *Technical report*, CEPR Discussion Paper No. 5926, London

Bester, C. A., Conley, T. G. and Hansen, C. B. (2008), Inference with Dependent Data Using Cluster Covariance Estimators. (mimeograph).

Binswanger, H. P. and McIntire, J. (1987). “Behavioral and Material Determinants of Production Relations in Land-Abundant Tropical Agriculture.”, *Econ. Dev. Cult. Change*, 36(1):73–99.

Coate, S. and Ravallion, M. (1993). “Reciprocity Without Commitment: Characterization and Performance of Informal Insurance Arrangements.”, *J. Dev. Econ.*, 40:1–24.

Comola, M. (2007), The Network Structure of Informal Arrangements: Evidence from Rural Tanzania., *Technical report*, INRA-LEA Working Paper 708, Paris.

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

De Weerdt, J. and Dercon, S. (2006). "Risk-Sharing Networks and Insurance Against Illness.", *Journal of Development Economics*, 81(2):337–56.

De Weerdt, J. and Fafchamps, M. (2007), Social Networks and Insurance against Transitory and Persistent Health Shocks. (mimeograph).

Ellsworth, L. (1989), *Mutual Insurance and Non-Market Transactions Among Farmers in Burkina Faso*, University of Wisconsin. Unpublished Ph.D. thesis.

Fafchamps, M. and Gubert, F. (2007). "The Formation of Risk Sharing Networks.", *Journal of Development Economics*, 83(2):326–50.

Fafchamps, M. and Lund, S. (2003). "Risk Sharing Networks in Rural Philippines.", *Journal of Development Economics*, 71:261–87.

Fisman, R., Iyengar, S., Kamenica, E. and Simonson, I. (2008). "Racial Preferences in Dating.", *Review of Economic Studies*, 75:117–32.

Genicot, G. and Ray, D. (2003). "Group Formation in Risk-Sharing Arrangements.", *Review of Economic Studies*, 70(1):87–113.

Goyal, S. (2007), *Connections: An Introduction to the Economics of Networks*, Princeton University Press, Princeton and Oxford.

Granovetter, M. (1985). "Economic Action and Social Structure: The Problem of Embeddedness.", *Amer. J. Sociology*, 91(3):481–510.

Hitsch, G. J., Hortacsu, A. and Ariely, D. (2005), *What Makes You Click: An Empirical Analysis of Online Dating*, University of Chicago and MIT, Chicago and Cambridge.

Jackson, M. O. (2008), *Social and Economic Networks*, Princeton University Press, Princeton.

Jackson, M. O. and Wolinsky, A. (1996). "A Strategic Model of Social and Economic Networks.", *Journal of Economic Theory*, 71(1):44–74.

Kimball, M. S. (1988). "Farmers' Cooperatives as Behavior Toward Risk.", *Amer. Econ. Rev.*, 78 (1):224–232.

Kocherlakota, N. R. (1996). "Implications of Efficient Risk Sharing Without Commitment.", *Rev. Econ. Stud.*, 63(4):595–609.

Lucas, R. E. and Stark, O. (1985). "Motivations to Remit: Evidence from Botswana.", *J. Polit. Econ.*, 93 (5):901–918.

Platteau, J.-P. (1995). "An Indian Model of Aristocratic Patronage.", *Oxford Econ. Papers*, 47(4):636–662.

Platteau, J.-P. (1996), Traditional Sharing Norms as an Obstacle to Economic Growth in Tribal Societies., *Technical report*, CRED, Facultés Universitaires Note-Dame de la Paix, Namur, Belgium. Cahiers No. 173.

Poirier, D. (1980). "Partial Observability in Bivariate Probit Models.", *Journal of Econometrics*, 12:209–17.

Scott, J. C. (1976), *The Moral Economy of Peasant: Rebellion and Subsistence in South-East Asia*, Yale University Press, New Haven.

Vandenbossche, J. and Demuynck, T. (2009), Network Formation with Absolute Friction and Heterogeneous Agents. (mimeograph).

Vega-Redondo, F. (2006), *Complex Social Networks*, Cambridge University Press, Cambridge.

Vuong, Q. H. (1989). “Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses.”, *Econometrica*, 57, No. 2:307–333.