Are Gifts and Loans between Households Voluntary?*

Margherita Comola[†] Marcel Fafchamps[‡]
Paris School of Economics Oxford University

October 2011

Abstract

We propose a simple test of unilateral versus bilateral link formation using a maximum likelihood estimator that deals with mis-reporting. This estimator is suitable for pairwise data sources with multiple discordant measures. We illustrate the methodology using dyadic data on inter-household gifts and loans from the village of Nyakatoke in Tanzania. We find reasonably strong evidence in favor of unilateral link formation: if a household wishes to enter in a relationship with someone who is sufficiently close socially and geographically, it can do so unilaterally. Flows of gifts and loans between two households are nevertheless more likely if both households wish to link. We show that not taking mis-reporting into account leads to serious underestimation of the total amount of gifts and loans between villagers.

Keywords: social networks; link formation; reporting bias; informal arrangements
JEL codes: C13; C51; D85

1 Introduction

In market exchange, it is customary to assume that transactions are voluntary in the sense that both sides are willing participants to the exchange. This typically arises because there are mutual gains from trade. There are exceptions, however. For instance, one party may be forced to trade because refusing to do so would contravene a legal obligation not to discriminate. In this example, trade is voluntary only for one side. Given the choice, the other side would prefer not to trade but is compelled to do so by legal or social norms.

Similar issues arise in other exchange processes, such as gifts and transfers. There are norms that compel one person to give to another. These norms may be legally enforced – *e.g.*, alimony or child support – or they may be enforced through a combination of social pressure and guilt – *e.g.*, charitable contributions to religious organizations. Norms may also pressurize people to accept gifts even if doing so implies an obligation to reciprocate – *e.g.*, Christmas cards, lunch invitation. In these examples, one party to the gift exchange (the giver or the recipient) may *ex ante* prefer for the transfer not to take place, but cannot refuse once it is initiated by the other party. We define such gifts as not being voluntary.

Gifts and market transactions are examples of processes that can be given a graph or network representation, but there are many others. According to the economic literature on networks (*e.g.*, Goyal 2007, Jackson 2009), a link formation process is defined as one-sided or unilateral if a link between two agents can be created by one of them without the assent of the other. If both agents must agree for a link to be formed, the process is said to be two-sided or bilateral – *i.e.*, voluntary. Note that these definitions characterize the link formation process, not the link itself: unilateral vs. bilateral refers to the way the link is formed, not to the direction of flows between the nodes. In fact, a link could involve a two-way flow (as in a sales transaction), a one-way flow from the link creator to the other agent (as in junk mail), or a one-way flow from the other agent to the link creator (as when accessing information on the internet). In each of these examples the link can be unilateral or bilateral: the consent of the other agent may or may not be required, depending on the context.[1]

As has been shown theoretically, the architecture of a network depends on whether the link formation process is unilateral or bilateral, and the architecture can have dramatic implications for efficiency and equity (*e.g.*, Bala and Goyal 2000, Jackson and Wolinsky 1996). It is therefore important to have a general method to empirically test whether a link formation process is bilateral or unilateral, *i.e.*, whether link formation is voluntary for both

---

[1]Two-way flow (one-way flow) networks can be also called undirected (directed): the two definitions are indeed equivalent. Market transactions are two-way flow networks which normally result from a two-sided (bilateral) link formation process, but we have already offered an example of unilateral exchange. For junk mail and tele-marketing the law may allow recipients to withdraw their consent to unsollicited communications, and for internet access the target link may provide access to only some individuals (*e.g.*, via a password): if this is the case, we observe a one-way flow network where link formation is two-sided (bilateral).

parties or not.

This paper makes two methodological contributions. First, we propose a general test of unilateral versus bilateral link formation when the researcher has information on individual desire to link. Second, we offer a methodology to correct for under- (or over-) reporting of link flows from survey data. This second methodological contribution is distinct from the first, but it complements it in a critical way. It is increasingly common for surveys to collect information on links and flows – $e.g.$, friendship, gifts, advice, referral. Experience shows that self-reported links and flows are often discordant, $i.e.$, i reports a gift to j but j does not report a gift from i. The accuracy of our test of unilateral versus bilateral link formation is affected by mis-reporting. We propose a maximum likelihood estimator to correct for the mis-reporting of links and flows in pairwise network data. While there is an established literature on measurement error in binary variables ($e.g.$ Hong and Tamer 2003; Schennach 2004), up to our best knowdledge no specific solution for dyadic data has been previously proposed.

We illustrate our methodology with dyadic data from a village in Tanzania. These data contain detailed information on all transfers ($i.e.$ gifts and loans) between all households in the village, as well as a proxy for each household's willingness to link with others in the village. There are massive discrepancies in survey responses about loans and gifts given and received: in many cases household i reports giving something to j but j does not report receiving anything from i. Throughout our analysis, we use the answer to a first-round question on whom people would ask and/or provide help to as proxy for desire to link. Previous work by Comola and Fafchamps (2009) – and additional evidence presented here – suggest that this is a reasonable assumption.

We test whether transfers are voluntary for both parties or not, that is, whether the link formation process is best understood as bilateral or unilateral. We find reasonably strong evidence in support of unilateral link formation: if a household wishes to enter in a gift-exchange relationship with another household that is sufficiently close socially and geographically, our results suggest that it can do so unilaterally. Flows of gifts and loans between two households are nevertheless more likely if both households wish to link: convergence of willingness to link is not required for transfers to take place, but it results in a higher likelihood of gifts and loans.

We interpret these findings as suggesting that surveyed households find it difficult to extricate themselves from social and familial obligations to assists others in need. This stands in contrast with much of the economic literature on risk sharing which emphasizes self-interest and reciprocal obligations as basis for mutual support (Coate and Ravallion 1993, Ligon, Thomas and Worrall 2001). As a by-product of our correction strategy we show that not taking mis-reporting into account leads to serious underestimation of the total amount of gifts and loans between villagers.

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 contains the discussion, and Section 6 concludes. Additional tables are reported in Appendix A, while in Appendix B we replicate the paper's main results under the alternative assumption of over-reporting.

2   Estimation strategy

It increasingly common for researchers studying link formation to obtain information about individual preferences over possible links. For instance, the work on matching processes (*e.g.*, Roth and Sotomayor 1990) typically asks employers and employee to rank all possible matches. Recent examples include: men and women listing potential partners in speed dating experiments (Belot and Francesconi 2006); students listing their preferred schools and schools selecting their preferred applicants (Erdil and Ergin 2007); chat room users sending emails to each other to signal interest (Hitsch, Hortacsu, and Ariely 2011); and relative prices for car part suppliers and automotive assemblers (Fox 2011). Even when willingness to link was not collected, proxies often are available for the objective utility or material gain that individuals derive from different matches. For our test to be usable, it must be possible to summarize the available information into a single index that captures the desire to link.

The method we propose to correct for mis-reporting is of particular interest to researchers studying social networks. Much social network analysis is based on dyadic data reported by survey respondents – *e.g.*, answers to questions such as 'to whom did you lend money', 'who are your friends', or 'are you related to X' (*e.g.*, Christakis and Fowler 2009, Steglich, Snijders, and Pearson 2010, Fafchamps and Lund 2003). In principle answers to these questions should agree: if i reports lending money to j, then j should report receiving money from i. Yet it is common for such data to be discordant, *i.e.*, there often are considerable discrepancies between answers given by i and j.

In many cases it is reasonable to assume that the main reason for these discrepancies is under-reporting: a link exists or a flow took place between i and j but one of them forgot to report it to enumerators. It can also happen that links or flows are over-reported, as when individuals inflate the number of their friends or sexual partners. Researchers rarely deal with these issues explicitly and typically rely on ad hoc assumptions, failing to recognize that the way discordant answers are handled can affect estimation and inference. In particular, simulations indicate that test results on unilateral versus bilateral link formation are sensitive to mis-reporting if the propensity to report is correlated with the variable of interest, *i.e.*, the desire to link.[2]

---

[2]To understand why, consider the following example. Imagine we have data on desire to link and subsequent gifts between households. We want to test whether gifts are the result of unilateral or bilateral link formation.

We propose a maximum likelihood estimator that deals with discordant answers in a systematic and consistent way. It forces the researcher to explicitly assume either under- or over-reporting – but also allows to investigate the sensitivity of the findings to assuming one or the other. This estimator can in principle be extended to other data containing two discordant self-reported measures of the same objective phenomenon.

2.1 Bilateral versus unilateral link formation

We first introduce the conceptual framework that underlies the test of bilateral versus unilateral link formation. In our empirical analysis, a link is defined to exist between i and j if there is an implicit favor exchange agreement between them, and τ_{ij} refers to loans or gifts observed over a given time interval.[3] More generally, a link can be any economic relationship of interest, and τ_{ij} any manifestation of this relationship, typically flows of money, goods, or services. We observe τ_{ij} between pairs of agents i and j.

We also have individual measures d_{ij} and d_{ji} of agents' desire to link, that is, to trade flows with each other. These measures are dichotomous, with $d_{ij} = \{0, 1\}$ and $d_{ji} = \{0, 1\}$. If link formation is unilateral, we are more likely to observe $\tau_{ij} > 0$ between i and j when either of them wishes to link. In this case the likelihood of observing $\tau_{ij} > 0$ increase in both d_{ij} and d_{ji}. If link formation is bilateral, a link between i and j only gets formed if both i and j wish to link, that is, if $d_{ij}d_{ji} = 1$. Furthermore, once we control for $d_{ij}d_{ji}$, variables d_{ij} and d_{ji} should have no additional effect on the probability of observing $\tau_{ij} > 0$.

This suggests the following testing strategy. Estimate a regression model of the form:

$$\Pr(\tau_{ij} > 0) = \lambda(\alpha d_{ij} + \beta d_{ji} + \gamma d_{ij}d_{ji} + \theta X_{ij}) \tag{1}$$

where X_{ij} is a vector of controls and λ is the logit function. If link formation is unilateral, the likelihood of $\tau_{ij} > 0$ is the same whether $\{d_{ij}, d_{ji}\} = \{1, 0\}, \{0, 1\}$, or $\{1, 1\}$. It follows

---

Faced with discordant gift data, researchers typically do one of two things. They may assume that if either i or j report a gift, then a gift between i and j took place; this is equivalent to assuming that when both reports agree they are true statements, and all observed discordances are due to under-reporting. Alternatively, they may assume that a gift between i and j took place if both i and j reported it; this is equivalent to assuming that when both reports agree they are true statements, and observed discordances are due to over-reporting. Now assume that, if a household i has reported wishing to link with household j, it is also more likely to subsequently report gifts to j. If the researcher adopts the first approach, *i.e.*, assumes discordance is due to under-reporting, test results will be biased towards rejecting bilateral link formation: gifts (as measured by the researcher) are more likely if i wants to link with j or if j wants to link with i – but rarely both if there is a lot of misreporting. In contrast, if the researcher adopts the second approach, *i.e.*, assumes discordance is due to over-reporting, gifts (as measured by the researcher) are more likely only if both i and j want to link with each other. This biases results towards rejecting unilateral link formation. Simulations show that more accurate inference on link formation is obtained with the mis-reporting correction.

[3]Because gifts and loans respond to shocks affecting i and j, they need not be observed over a fixed time interval even if a link exists between i and j.

that:

$$\alpha = \beta = \alpha + \beta + \gamma > 0$$

which implies that $\gamma = -\beta = -\alpha$. If link formation is bilateral, $\tau_{ij} > 0$ arise only if $\{d_{ij}, d_{ji}\} = \{1, 1\}$. It follows that

$$\alpha = \beta = 0 \text{ and } \gamma > 0$$

2.2 Mis-reporting

Our objective is to estimate (1). To do so effectively, we must address the issue of mis-reporting. Suppose that the researcher has data in which both i and j were asked about flows (in our case, loans and gifts) between them. In principle, i and j should report the same τ_{ij}. This is not, however, what we observe: when one side reports $\tau_{ij} > 0$, the other typically reports $\tau_{ij} = 0$. This is a common problem when analyzing datasets with multiple discordant measures of the same objective phenomenon, *e.g.*, multiple measurements of schooling levels in twins (Ashenfelter and Krueger, 1994), discrepancies over earnings reported by workers and companies (Duncan and Hill, 1985), estimates of time spent on housework by the spouse (Lee and Waite 2005).

We have no reason to suspect that respondents report flows that did not take place, since reporting a loan or gift to an enumerator takes time and effort. This is typical of survey data on network flows. It follows that discrepancies between reports made by i and j most likely correspond to under-reporting due to recall error.[4] It is therefore reasonable to assume that if a transfer is reported by either i or j, a transfer took place. It is also possible that a transfer took place but was not reported by either i or j. Although under-reporting is the most reasonable scenario in our context, we show in Appendix B how the methodology can be modified to correct for over-reporting.

Dropping the ij subscripts to improve readability, let τ denote the true binary flow or transfer from i to j, *i.e.*, $\tau = 1$ if i made a transfer to j.[5] Further let G be the report that the giver i made on this transfer and let R be the report that the receiver j made on the same transfer. We have $G = 1$ if i reported making a transfer and 0 otherwise. Similarly, $R = 1$ if j reported receiving a transfer, and 0 otherwise. We do not observe τ, only G and

---

[4]There is some evidence of this in the data itself. Transfers reported by both sides are on average much larger than transfers reported by one side only. For instance, the average value of a gift declared by the receiver is 2044 Tanzanian shillings (*tzs*) when the giver also declares a non-zero amount, and 1260 *tzs* when the giver does not declare any transfer. The gap is smaller for what concerns loans, but still significant. This is in line with the hypothesis of recall mistakes that decrease in the amount transferred.

[5]We prefer to model transfers as binary because of the major discrepancies between the amounts declared by giver and receiver (see Section 3). However, the analysis could be extended to a framework where transfers are continuous.

R. Under-reporting implies that $G = 1$ only if $\tau = 1$, and that $R = 1$ only if $\tau = 1$. Given these assumptions, the data generation process takes the following form:

$$
\begin{aligned}
\Pr(G = 1, R = 0) &= \Pr(\tau = 1, G = 1, R = 0) \\
&= \Pr(\tau = 1) * \Pr(G = 1 | \tau = 1) * \Pr(R = 0 | G = 1, \tau = 1) \\
\Pr(G = 0, R = 1) &= \Pr(\tau = 1, G = 0, R = 1) \\
&= \Pr(\tau = 1) * \Pr(G = 0 | \tau = 1) * \Pr(R = 1 | G = 0, \tau = 1) \\
\Pr(G = 1, R = 1) &= \Pr(\tau = 1, G = 1, R = 1) \\
&= \Pr(\tau = 1) * \Pr(G = 1 | \tau = 1) * \Pr(R = 1 | G = 1, \tau = 1) \\
\Pr(G = 0, R = 0) &= 1 - \Pr(G = 1, R = 0) - \Pr(G = 0, R = 1) - \Pr(G = 1, R = 1)
\end{aligned}
$$

If we further assume that under-reporting by i is independent of under-reporting by j, then $\Pr(R | G, \tau) = \Pr(R | \tau)$. This assumption, which is required for identification, is reasonable if under-reporting results primarily from mistakes and omissions. With this assumption, we can rewrite the system as:

$$
\begin{aligned}
\Pr(G = 1, R = 0) &= \Pr(\tau = 1) * \Pr(G = 1 | \tau = 1) * \Pr(R = 0 | \tau = 1) & (2) \\
\Pr(G = 0, R = 1) &= \Pr(\tau = 1) * \Pr(G = 0 | \tau = 1) * \Pr(R = 1 | \tau = 1) & (3) \\
\Pr(G = 1, R = 1) &= \Pr(\tau = 1) * \Pr(G = 1 | \tau = 1) * \Pr(R = 1 | \tau = 1) & (4) \\
\Pr(G = 0, R = 0) &= 1 - \Pr(G = 1, R = 0) - \Pr(G = 0, R = 1) - \Pr(G = 1, R = 1) & (5)
\end{aligned}
$$

Equations (2) to (5) express the data generating process in terms of three probabilities: $P(\tau = 1)$, $P(G = 1 | \tau = 1)$ and $P(R = 1 | \tau = 1)$. To obtain the likelihood function, we assume that these three probabilities can be represented by three distinct logit functions $\lambda(.)$ as follows:

$$
\begin{aligned}
\Pr(\tau = 1) &= \lambda_T(\beta_\tau X_\tau) & (6) \\
\Pr(G = 1 | \tau = 1) &= \lambda_G(\beta_G X_G) & (7) \\
\Pr(R = 1 | \tau = 1) &= \lambda_R(\beta_R X_R) & (8)
\end{aligned}
$$

Together with (2) to (5), equations (6) to (8) fully characterize the likelihood of observing the data. The main equation of interest is $\lambda_T(\beta_\tau X_\tau)$ which corresponds to equation (1): it is on this equation that we wish to test the restrictions imposed by our testing strategy. Conditioning on X_G and X_R in $\Pr(G = 1 | \tau = 1)$ and $\Pr(R = 1 | \tau = 1)$ allows for correlation on observables in reporting probabilities between the giving and receiving households.

2.3 Simulation analysis

In the literature to date, mis-reporting has typically been ignored and estimation has proceeded using transfers τ_{ij} reported by i, j, or a combination of the two (e.g., Glaeser, Sacerdote, and Scheinkman 1996, Liu, Patacchini, Zenou and Lee 2011, Snijders, Koskinen, and Schweinberger 2010). For instance, Fafchamps and Lund (2003) and De Weerdt and Fafchamps (2011) use transfers information obtained from one of the two households only – i for transfers given, and j for transfers received. Fafchamps and Gubert (2007) combine answers given by i and j to construct a unique measure of τ_{ij}.

Whether or not mis-reporting affects inference depends on the hypothesis being tested. Our ultimate objective is to test whether gifts and loans are unilateral or bilateral. Hence we are primarily interested in the coefficients of d_{ij}, d_{ji} and $d_{ij}d_{ji}$ in equation (6). We expect the propensities to report a transfer (equations 7 and 8) to vary systematically with d_{ij} and d_{ji}: household i is more likely to report transfers to households from whom he wishes to seek help, *i.e.*, households for which $d_{ij} = 1$. Similarly, j may be more likely to report gifts received from households for which $d_{ji} = 1$.

If the data generating process has these properties, using probit or logit to estimate (1) is likely to yield incorrect inference. To see this, let τ_{ij}^i denote i's report about a transfer from i to j, and let τ_{ij}^j be j's report on the same actual transfer τ_{ij}. Faced with discordant gift data, researchers may assume that $\tau_{ij} \neq 0$ if *either* i or j report a gift – *i.e.* if $\tau_{ij}^i \neq 0$ or $\tau_{ij}^j \neq 0$ – or *both* i and j report a gift – *i.e.*, if $\tau_{ij}^i \neq 0$ and $\tau_{ij}^j \neq 0$. Suppose that we apply the first approach and that the true underlying model is bilateral link formation. If τ_{ij}^i (τ_{ij}^j) is correlated with d_{ij} (d_{ji}), then $\tau_{ij} \neq 0$ is more likely if $d_{ij} = 1$ *or* $d_{ji} = 1$ even though, under bilateral link formation, only the coefficient of $d_{ij}d_{ji}$ should be different from 0. This shows that magnitude of the coefficients of d_{ij} and d_{ji} in (1) will be biased. This, in turn, will affect the sign and significance of the cross term $d_{ij}d_{ji}$. By a similar reasoning, a bias also arises if we adopt the second approach.

To formally illustrate how mis-reporting affects inference regarding the coefficients of d_{ij}, d_{ji} and $d_{ij}d_{ji}$ in equation (1), we conduct a simulation analysis of the data generating process defined by equations (2) to (8) under different assumptions regarding mis-reporting and link formation. Results, not shown here to save space, show that mis-reporting can dramatically affect our inference regarding d_{ij}, d_{ji} and $d_{ij}d_{ji}$ in equation (1). We report the main findings in what follows.

If we observe the actual gifts τ_{ij} without mis-reporting, equation (1) can be estimated directly through logit or probit. Results are as anticipated: if link formation is bilateral, $\alpha = \beta = 0$ while $\gamma > 0$; if link formation is unilateral, $\alpha = \beta > 0$ and $\gamma = -\beta = -\alpha$ holds. If we do not observe the actual gifts τ_{ij}, we can choose to ignore mis-reporting and estimate equation (1) by assuming that a transfer took place if *either* i or j reported it. Simulation

results indicate that, in this case, coefficient estimates (1) are reasonable if mis-reporting does not depend on desire to link d_{ij} and d_{ji}. However, if it does, they are severely biased.[6]

Next we estimate equation (1) by maximum likelihood using the likelihood function defined by equations (2) to (8). We first assume that mis-reporting is present but does not depend on d_{ij} and d_{ji}. In this case, ML estimates are consistent whether or not we include d_{ij} in X_G and d_{ji} in X_R. We then assume that d_{ij} is in X_G and d_{ji} is in X_R. This is equivalent to assuming that respondents are more likely to remember a transfer to (or from) individuals with whom they wish to link. In this case, ML estimates are consistent only if we include d_{ij} in X_G and d_{ji} in X_R. If we do not, the coefficient of $d_{ij}d_{ji}$ – which is essential to our testing strategy – is severely biased, often with the incorrect sign. These findings motivate the specifications presented in Section 4, where we estimate model (2) to (8) with d_{ij} in X_G and d_{ji} in X_R. Simulations also show that, if mis-reporting does not depend on desire to link, consistent ML estimates obtain even if X_G and X_R only contain an intercept term. This indicates that identification does not require that X_G and X_R contain a variable absent from X_τ.

2.4 Standard errors

Dyadic observations such as those on τ_{ij} are typically not independent. This does not invalidate the application of standard maximum likelihood techniques to estimate β_τ, β_G and β_R in equations (6) to (8). But standard errors must be adjusted to correct for dyadic dependence across observations, otherwise inference will be inconsistent. If we had data from a sufficient number of distinct sub-populations we could cluster the standard errors to correct for correlation across observations from the same sub-population (Arcand and Fafchamps 2012). Unfortunately, we only have data from a single village. Given this, we apply the formula developed by Fafchamps and Gubert (2007), using the scores in place of X. This approach corrects for arbitrary correlation across all τ_{ij} and τ_{ji} observations involving either i or j. The simulation analysis reported earlier was conducted using dyadic standard errors. Results indicate that t-values obtained via this method are a good basis for correct inference about α, β and γ.

3 The data

We illustrate our methodology using a unique dataset on transfers between all the households in an African village, Nyakatoke, in the Buboka Rural District of Tanzania, at the west of

---

[6]We also estimated equation (1) assuming that a transfer took place if *both* i and j report it. In this case, results are inconsistent irrespective of the form of response bias. This is hardly surprising given the assumed data generation process precludes over-reporting.

Lake Victoria. These data were collected in multiple rounds over an entire year and have been the object of numerous articles (e.g., De Weerdt and Dercon 2006, De Weerdt and Fafchamps 2011, Comola 2008, Comola and Fafchamps 2009, Vandenbossche and Demyunck 2010).

The village's main livelihood is the farming of bananas, sweet potatoes and cassava for food; 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.[7] 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.

From February to December 2000 each adult household member was repeatedly asked whether they had received or given any loans or gifts. If they said yes, information was collected on the name of the partner, the value of what was given or received, whether in cash or kind. Loan repayment and gifts in labor are not included. Aggregating at the household level across rounds, we obtain a picture of transfers of funds between all households in the village.[8] We aggregate across rounds to reduce discrepancies in answers due to difference in interview dates across households, *i.e.,* if household i declares a transfer in round t while household j declares the same transfer in round $t + 1$. We also aggregate at the household level to reduce discrepancies if i mentioned giving to member a of household j but member b is the one who mentions receiving a gift from household i.

For each household dyad ij we have four variables: gifts τ_{ij}^i that i stated giving to j; gifts τ_{ij}^j that j stated receiving from i; gifts τ_{ji}^j that j declared giving to i; and gifts τ_{ji}^i that i stated receiving from j. Similar data is available for loans. The literature on informal risk sharing has noted that informal loans often serve to smooth consumption against shocks (Udry 1994) and can be a way of reducing self-enforcement constraints (Foster and Rosenzweig 2001, Kocherlakota 1996, Ligon and Thomas and Worrall 2001). In Nyakatoke, gifts are more frequent than loans but smaller in magnitude (De Weerdt and Dercon 2006, De Weerdt and Fafchamps 2011). This is in line with findings reported by Fafchamps and Lund (2003) for the Philippines. Gifts in Nyakatoke have been shown to serve an insurance purpose against health shocks (De Weerdt and Fafchamps 2011).

There are major discrepancies between τ_{ij}^i and τ_{ij}^j. In fact, $\tau_{ij}^i \neq \tau_{ij}^j$ in nearly all cases, especially for loans. There are 1420 dyads (*i.e.,* 10% of the dyads) for which either τ_{ij}^i or τ_{ij}^j is not zero for gifts. Of those, in 42% of cases the report comes from the giver only, in 30% from the receiver only, and in 27% from both. For inter-household loans, there are 545 dyads (*i.e.,* 4% of the dyads) for which either i or j reports a loan from i to j. In 56% of these cases,

---

[7]Everyone in the village agreed to participate in the survey, but there are some missing data for 4 households.
[8]When aggregating at the household level, questionnaires were carefully checked by survey supervisors to avoid any double-counting of identical gifts reported by two different members of the same household.

the report comes from the giver only, in 36% from the receiver only, and in 8% from both. Out of 378 dyads in which both i and j report a gift from i to j, only 22 report the same amount. For loans, the corresponding number is 5 out of 37. When the amounts declared differ, they differ by a large margin: the highest of the two declared amounts is on average double the smallest one. This is true for both loans and gifts. Transfers reported by both sides are on average much larger than transfers reported by one side only.[9] The frequency distribution of loan and gift amounts is given in Table A1, Appendix A.

We checked whether discrepancies are due to the fact that respondents mix up loans and gifts. The within-dyad correlation between the difference in reported loans and the difference in reported gifts is indeed negative, as would be the case if, say, i reports giving a loan while j reports receiving a gift. But the correlation is small and not statistically significant: if we restrict the sample to the dyads for which at least one loan or gift was declared, the correlation between the difference in reported loans and the difference in reported gifts is -0.036 with a significance level of 0.13.

In summary, there are massive discrepancies between the responses given by i and j about the same gifts and loans τ_{ij}. These discrepancies are mostly due to the fact that in the the large majority of cases – 93% of the cases for loans and 73% of the cases for gifts – one side reports something while the other reports nothing. Under-reporting by those who receive gifts and loans may not be too surprising: they may have a strategic motive in 'forgetting' the favors they probably have a moral obligation to reciprocate. But we also observe massive under-reporting by those who give. Consequently there may be many transfers which took place but are not observed in the data because they were not mentioned by either sides. When estimating model (1), our main challenge is to address this bias.

3.1 Variables definition

Our unit of observation is the dyad: in Nyakatoke there are 119 households, which gives $119 * 118 = 14042$ possible dyads. We organize the data such that the first listed household refers to the giver and the second to the receiver, $i.e.$, τ_{ij} refers to a transfer from i to j. Note that τ_{ij} defines a directed graph: τ_{ij} represents the transfer from i to j, while τ_{ji} represents the transfer from j to i. For τ_{ij} we have two different measurements: the information provided by the giver τ_{ij}^i, and the information provided by the receiver τ_{ij}^j. Similarly for τ_{ji}.

From equation (1) our main regressors of interest are d_{ij}, d_{ji} and $d_{ij}d_{ji}$. In the first Nyakatoke survey round (February 2000), each adult household member was asked: *"Can you give a list of people from inside or outside of Nyakatoke, who you can personally rely*

---

[9]For instance, the average value of a gift declared by the receiver is 2044 Tanzanian shillings (tzs) when the giver also declares a non-zero amount, and 1260 tzs when the giver does not declare any transfer. This is consistent with the idea that respondents are more likely to recall large transfers than small transfers.

on for help and/or that can rely on you for help in cash, kind or labor?" Answers to this question, aggregated at the household level, are used as proxies for d_{ij} and d_{ji}.[10] This requires some explanation given that the question in principle asks about existing links – not desire to link. We first note that if responses perfectly captured actual links, then we would observe $d_{ij} = d_{ji}$ for all i and j. This is not the case: out of 14042 possible dyads, there are 980 dyads for which d_{ij} or d_{ji} is not 0. Of those, only 280 have $d_{ij} = d_{ji} = 1$ while 700 dyads have $d_{ij} = 1$ but $d_{ji} = 0$ or the reverse.

There remains the possibility that d_{ij} and d_{ji} are about actual links but contain a lot of mis-reporting. Comola and Fafchamps (2009) examine this issue in detail using the same data. They test whether d_{ij} and d_{ji} are best viewed as desire to link or as mis-reported links, and they find that the data are best interpreted as desire to link.[11] In what follow we provide evidence confirming the conclusion of their test. Let z_j be a characteristic of j correlated with i's desire to link with j, and similarly for z_i. Stack observations d_{ij} and d_{ji} and regress them on z_i and z_j in a logit regression of the form $d_{ij} = az_i + bz_j + u_{ij}$ and $d_{ji} = az_j + bz_i + u_{ji}$. Consider what happens if d_{ij} and d_{ji} are measurements of actual links and link formation is bilateral, but i and j sometimes forget to report existing links. In this case, d_{ij} should be 1 only when i knows j wants to link with him. Similarly, d_{ji} should only be 1 when j knows that i wants to link with him. Since both d_{ij} and d_{ji} enter the regression, on average we should have $a \approx b$. A similar prediction arises when link formation is unilateral: i should report a link whenever i or j wishes to link – and thus the likelihood of reporting a link d_{ij} should rise with both the attractiveness of i and that of j.

It is also conceivable that i only mentions those links that he cares about, and j likewise. When this happens, d_{ij} is increasing in the attractiveness of j for i, but not in i's attractiveness to j, *i.e.*, $b > 0$ but $a = 0$. In this case, d_{ij} proxies for i's desire to link with j, not for a link between i and j.

Regression results are reported in Appendix A, Table A2. As predictors of attractiveness z_j we use wealth and popularity: wealthier households are in a better position to assist others in need; and popularity proxies for other attributes correlated with attractiveness.[12] We find

---

[10]This question was first piloted in the Philippines (Fafchamps and Lund 2003) and subsequently adopted in the Tanzania survey. This question was used because respondents understand it and are willing to answer it. Other questions were tried, for instance drawing a distinction between those the respondent would help and those the respondent would seek help from. But respondents were confused by the distinction which they perceived as non-existent, and complained they were asked the same question twice, which undermines the survey. 34% of the mentioned partners live out of the village. They are omitted from the analysis since we have no information on the partner and hence cannot apply our testing methodology.

[11]The intuition behind the identification strategy is that, if d_{ij} and d_{ji} measure desire to link, i lists node j that is attractive to i irrespective of whether i is attractive to j. In contrast, if d_{ij} and d_{ji} are two statements about the same actual link, i should take into account his own attractiveness to j when answering the question.

[12]Wealth is computed as the total value of land and livestock assets in Tanzanian shilling (1 unit = 100000 tzs). Popularity of household j is defined as the number of times j is listed by households other than i in

$b > 0$ but $a = 0$: the wealth and popularity of the partner are strong predictors of d_{ij} but own characteristics are not significant. These results confirm that d_{ij} and d_{ji} can reasonably be regarded as proxying for the desire to link.

Turning to other regressors, the main regression of interest is $\Pr(\tau = 1) = \lambda_T(\beta_\tau X_\tau)$. The regressors entering X_τ are control variables expected to influence the actual flows of funds between households. Since τ_{ij} is directional, regressors for observation ij can differ from regressors for observation ji; this stands in contrast with undirected network data where regressors by construction have to be identical. We expect flows of funds between households to depend on the wealth of the giver and receiver, which we control for. From the work of Fafchamps and Lund (2003), De Weerdt and Dercon (2006) and De Weerdt and Fafchamps (2011), we suspect informal arrangements to be more frequent among households that are geographically and socially proximate. Finally, larger households have more individuals involved in giving and receiving transfers. We therefore control for the wealth of i and j, the number of adult members for i and j, the distance between the two houses, and dummies for whether i and j are blood-related, and share the same religion.[13]

Next we discuss the variables that enter $\Pr(G = 1|\tau = 1) = \lambda_G(\beta_G X_G)$ and $\Pr(R = 1|\tau = 1) = \lambda_R(\beta_R X_R)$. The first measures the propensity for the giver to report a transfer that has taken place; the second measures the receiver's propensity to report a transfer that has taken place. As discussed earlier, based on our simulation results we include d_{ij} in X_G – givers are more likely to remember transfers to individuals whose name they listed in response to first-round interviews. We include d_{ji} in X_R for the same reason. We also include own wealth (wealth of i in X_G and wealth of j in X_G) as regressor given that wealthy people are more likely to forget a transfer. Social and geographical proximity variables are included to allow for the possibility that respondents remember better transfers to and from proximate households.

We also include regressors that can be *a priori* expected to affect mis-reporting but not transfers themselves.[14] For X_G, we use $n_i \equiv \sum_j d_{ij}$, that is, the number of individuals listed in response to the first-round question on who respondents would turn to for help and to whom they would provide help. The logic underneath this choice is that households intending to seek help from (or provide help to) many other households may be more sensitive to the issue of inter-household transfers, and therefore recall transfers better. For X_R we include the number of male and female adult dependents. The idea is that adult dependents who have received transfers from other households may not have reported them to the household

---

response to the first-round question.

[13]We consider households i and j blood-related if an adult member of i is the parent/sibling/child of an adult member of j.

[14]Simulation analysis reported earlier indicates that ML estimates are reliable even without identifying instruments, so including these variables is not necessary for identification.

head – and therefore may be reluctant to report them to enumerators.

To illustrate how our correction for mis-reporting affects inference regarding the link formation process, we estimate two logit regressions comparable with $\Pr(\tau = 1)$. In the first of them, the dependent variable equals one if at least one side has declared a gift. This is equivalent to defining $\tau_{ij}^u \equiv \max\{\tau_{ij}^i, \tau_{ij}^j\}$ and assumes that all discordances are due to under-reporting. In the second regression the dependent variable equals one if both the giver and the receiver have declared a gift, $i.e.$, it is $\tau_{ij}^o \equiv \min\{\tau_{ij}^i, \tau_{ij}^j\}$, which is equivalent to assuming that discordances are due to under-reporting. As argued before, systematic over-reporting is unlikely in our context, but we report the results anyway for the sake of comparison and to illustrate how the methodology works.

In Table 1 we present summary statistics for all variables used in the analysis. The upper section of the table reports different versions of the dependent variable. The first two rows focus on the gifts from i to j, as reported by i and j. Variables τ_{ij}^i takes value 1 if i reported a gift to j, and 0 otherwise. Similarly for τ_{ij}^j. We see that givers are more likely to report a gift than receivers. In the next two rows we report $\tau_{ij}^u \equiv \max\{\tau_{ij}^i, \tau_{ij}^j\}$ and $\tau_{ij}^o \equiv \min\{\tau_{ij}^i, \tau_{ij}^j\}$. They demonstrate the extent of the divergence between the information given by households i and j on the same τ_{ij}. In the next four rows we report the same information for inter-household loans. Variables are constructed in the same way. Here too we see that lenders are more likely to report a loan than borrowers, and that there are considerable discrepancies between loans reported by the giver and loans reported by the receiver.

Table 1. Descriptive statistics (N=14042)

variable	mean	min	max	sd
τ_{ij}^{i} (gifts)	0.071			
τ_{ij}^{j} (gifts)	0.059			
τ_{ij}^{u} (gifts)	0.101			
τ_{ij}^{o} (gifts)	0.028			
τ_{ij}^{i} (loans)	0.025			
τ_{ij}^{j} (loans)	0.017			
τ_{ij}^{u} (loans)	0.039			
τ_{ij}^{o} (loans)	0.003			
d_{ij} and d_{ji}	0.045			
$d_{ij}d_{ji}$	0.020			
$weighted\ d_{ij}$ and $weighted\ d_{ji}$	0.023	0	0.933	0.117
$wealth_{i}$ and $wealth_{j}$	4.546	0	27.970	4.815
$same\ religion$	0.354			
$related$	0.016			
$distance$	0.522	0.014	1.738	0.303
$hh\ members_{i}$ and $hh\ members_{j}$	2.555	1	9	1.314
n_{i}	5.294	0	19	3.063
$female\ dependents_{j}$	1.101	0	6	0.864
$male\ dependents_{j}$	0.437	0	3	0.729

From these figures it is possible to compute a rough estimate of the extent of under-reporting, before introducing covariates in our analysis. We focus on gifts first. Assuming independence in reporting probability between i and j, we wish to estimate three unconditional probabilities: $\Pr(\tau = 1)$, $\Pr(G = 1|\tau = 1)$, and $\Pr(R = 0|\tau = 1)$. We have three equations to do so:

$$\Pr(G = 1, R = 0) = \Pr(\tau = 1) * \Pr(G = 1|\tau = 1) * \Pr(R = 0|\tau = 1) = 0.043 \qquad (9)$$
$$\Pr(G = 0, R = 1) = \Pr(\tau = 1) * \Pr(G = 0|\tau = 1) * \Pr(R = 1|\tau = 1) = 0.031 \qquad (10)$$
$$\Pr(G = 1, R = 1) = \Pr(\tau = 1) * \Pr(G = 1|\tau = 1) * \Pr(R = 1|\tau = 1) = 0.028 \qquad (11)$$

Simple algebra yields the following solution:

$$\Pr(\tau = 1) = 15\%$$
$$\Pr(G = 1|\tau = 1) = 47\%$$
$$\Pr(R = 1|\tau = 1) = 39\%$$

The above calculation shows that there is considerable under-reporting of gifts and that $\tau_{ij}^u = 10.1\%$ underestimates the frequency of gifts by almost 50%. A similar calculation for loans yields:

$$
\begin{aligned}
\Pr(\tau = 1) &= 14\% \\
\Pr(G = 1 | \tau = 1) &= 18\% \\
\Pr(R = 1 | \tau = 1) &= 12\%
\end{aligned}
$$

which suggests massive under-reporting of loans and indicates that $\tau_{ij}^u = 3.9\%$ only captures a quarter of the loans we suspect were made.

The rest of Table 1 focuses on regressors. Variable $d_{ij} = 1$ if someone in household i mentioned someone in household j in response to the first-round question on who respondents would turn to in order to give or receive help. The product $d_{ij}d_{ji} = 1$ if i listed j and j listed i, something that occurs only for 2% of the dyads.

We also report a weighted version of d_{ij} that is constructed as follows. Remember that the first-round question on who respondents turn for help was answered separately by each adult members of the household. For each household member l in household i, we know the order in which they listed various individuals m from other households j. This order may contain information on how seriously l regards m to be a possible source of assistance. To aggregate this information at the level of the household, we construct a weighted link variable $weight_{lm}$ for each lm pair. This variable is defined as:

$$
weight_{lm} = \frac{(\#names_l + 1) - rank_{lm}}{\#names_l + 1}
$$

where $\#names_l$ is the total number of names given by l and $rank_{lm}$ is the order in which m was listed by l. We then average $weight_{lm}$ across all l members of household i and all m members of household j.[15]

Control variables are reported next. Whenever the average is the same for giver and receiver, we only report one of them. Wealth is computed as the total value of land and livestock assets (1 unit $= 100000$ tzs). We see there is considerable variation in wealth levels across Nyakatoke households. There is also significant diversity in religion: only 35% of households heads share the same religion.[16] Around 1.6% of household pairs are closely related, $i.e.$, are siblings or children-parents. Distance between households is measured in Km and is on average 500 meters.[17] Adult members are those aged 15 and above. Male and

---

[15]Whenever l mentions someone who lives outside Nyakatoke, we take this person into account when computing $\#names_l$ and $rank_m$.

[16]Out of 119 households, 24 are Muslim (20%), 46 are Protestant (39%) and 49 are Catholic (41%).

[17]For 3 households the distance to other households is missing, so we have imputed the sample average to avoid losing those observations.

female dependents are defined as adult members of the household who are not the head of household. Wives are included in the dependents, the idea being that they too may seek to dissimulate gifts and loans received from other households.

4 Estimation results

4.1 Main results

We now estimate the model presented in Section 2. In Table 2 transfers τ_{ij} refer to gifts from i to j, in cash or in kind. Columns (1) and (2) report logit regressions where the dependent variables are τ_{ij}^u and τ_{ij}^o, respectively. These regressions are to be compared to the estimates of column (3) which correct for mis-reporting. Columns (3) to (5) of Table 2 report coefficients obtained from estimating the likelihood function combining equations (2) to (8). Column (3) corresponds to our equation of interest (1).

Comparing the two logit models with (1), we see that when we correct for mis-reporting in column (3), the magnitude and significance of the coefficient of the cross-product $d_{ij}d_{ji}$ change. This difference is consistent with simulation results that suggest that without mis-reporting correction, the coefficient of $d_{ij}d_{ji}$ is seriously biased, occasionally leading to sign reversal and incorrect inference. Since inference about unilateral versus bilateral link formation relies heavily on the sign of the $d_{ij}d_{ji}$ coefficient, estimates reported in column (3) should be regarded as the most reliable.

Results reported in Table 2 strongly reject the bilateral link formation model: both α and β are strongly significant, while γ is never significantly positive. Coefficient estimates are at least partly consistent with unilateral link formation: α and β, the coefficients of d_{ij} and d_{ji}, are both significant and of the same order of magnitude. A Wald test cannot reject the hypothesis that $\alpha = \beta$ in column (3), with a p-value of 0.365. As predicted by the unilateral model the coefficient γ of $d_{ij}d_{ji}$ is negative in all three columns (1), (2) and (3) – significantly so in column (1) and (2). However, contrary to the predictions of the unilateral link formation model, $\gamma \neq -\beta$ and $\gamma \neq -\alpha$. For our preferred model (column 3) a Wald test rejects the joint hypotheses $\gamma + \alpha = 0$ and $\gamma + \beta = 0$ with p-value=0.002. This means that the probability of transfer is larger if both $d_{ij} = 1$ and $d_{ji} = 1$ than if only one of them is equal to 1. In other words, when both households list each other as someone they would go to for help, they are much more likely to help each other than if only one lists the other. This suggests that some bilateral dimension is present, even the the results reject the pure bilateral model itself.

Control variables in Column (3) have reasonable coefficients, and occasionally differ from columns (1) and (2) in significance and magnitude. Wealthier households are more likely to give (and more likely to receive but only in column 1). People are more likely to give to relatives, neighbors, and members of the same religion.

Results for the two under-reporting regressions – columns (4) and (5) – show that respondents are more likely to report a transfer from/to those households they with to link with. In the $\Pr(G = 1|\tau = 1)$ regression, d_{ij} is positively significant, indicating that if household i has listed household j in response to the first-round question, then i is more likely to report a gift given to j. Variable n_i, which is the total number of individuals listed in response of the first-round question, is significantly positive, suggesting that large households are more likely to report gifts given. Own wealth is significant and negative: wealthy respondents are more likely to forget reporting the gifts they have made. Analogously, in the $\Pr(R = 1|\tau = 1)$ regression d_{ji} is positively significant, and $wealth_j$ is negatively significant. The numbers of female and male dependents have the anticipated negative sign, but they are not significant.

To get a sense of the relative magnitude of under-reporting coefficients, we calculate marginal effects for the $\Pr(G = 1|\tau = 1)$ and $\Pr(R = 1|\tau = 1)$ regressions. Results, reported in Table A3 in Appendix, confirm that d_{ij} and d_{ji} have quantitatively the largest effect on mis-reporting. Relatedness and geographical distance also have effects that are large in magnitude.

In Table 3 we repeat the same analysis for loans instead of gifts. Coefficient estimates reported in column (3) approximately satisfy $\alpha = \beta = -\gamma$, a finding that is consistent with the unilateral model. A Wald test of the joint hypothesis that $\alpha = \beta = -\gamma$ has p-value of 0.930, implying that we cannot reject the hypothesis of unilateral link formation. But individual coefficients are only statistically significant in column (1) and (2). This may be because the proportion of non-zero observations is small for loans, making ML estimation more demanding for the multi-equation estimator. In terms of the other regressors, few of them are significant, a point already noted by De Weerdt and Fafchamps (2011) in the same dataset. In column (3), we find $wealth_i$ (marginally) significant, indicating that wealthy households are more likely to lend money. In the $\Pr(G = 1|\tau = 1)$ regression only n_i is significantly positive, and in the $\Pr(R = 1|\tau = 1)$ regression only the previously declared desire to link d_{ji} is significantly positive. Marginal effects reported in Table A3 in Appendix A show that the variables with the largest impact are: d_{ij} (which is not significant) and n_i for the giver; and d_{ji} for the receiver.

Table 2. Results for gifts

	(1) τ_{ij}^u	(2) τ_{ij}^o	(3) $\Pr(\tau=1)$	(4) $\Pr(G=1|\tau=1)$	(5) $Pr(R=1|\tau=1)$
d_{ij}	2.477***	2.527***	2.563***	1.492***	
	(0.180)	(0.233)	(0.371)	(0.180)	
d_{ji}	2.794***	3.260***	2.817***		1.920***
	(0.159)	(0.236)	(0.305)		(0.227)
$d_{ij}d_{ji}$	-0.681**	-1.036***	-0.196		
	(0.306)	(0.313)	(0.980)		
$wealth_i$	0.058***	0.024	0.081***	-0.035**	
	(0.012)	(0.015)	(0.016)	(0.016)	
$wealth_j$	0.066**	-0.001	0.105		-0.045***
	(0.030)	(0.024)	(0.066)		(0.015)
$same\,religion$	0.421***	0.353***	0.530**	0.025	0.012
	(0.099)	(0.132)	(0.251)	(0.211)	(0.196)
$related$	1.728***	0.944***	1.961**	0.433	0.614
	(0.284)	(0.294)	(0.762)	(0.505)	(0.377)
$distance$	-1.711***	-1.789***	-1.678**	-0.585	-0.533
	(0.294)	(0.476)	(0.660)	(0.536)	(0.485)
$hh\,members_i$	0.084***	0.069	0.110**		
	(0.032)	(0.060)	(0.043)		
$hh\,members_j$	0.216**	0.169**	0.262		
	(0.098)	(0.086)	(0.168)		
n_i				0.026*	
				(0.013)	
$female\,dependents_j$					-0.149
					(0.143)
$male\,dependents_j$					-0.191
					(0.133)
$constant$	-3.631***	-4.964***	-3.525***	-0.277	-0.209
	(0.302)	(0.419)	(0.540)	(0.590)	(0.359)

*** p<0.01, ** p<0.05, * p<0.1. Dyadic-robust standard errors in parentheses.

Table 3. Results for loans

	(1) τ_{ij}^u	(2) τ_{ij}^o	(3) $\Pr(\tau=1)$	(4) $\Pr(G=1|\tau=1)$	(5) $Pr(R=1|\tau=1)$
d_{ij}	1.966***	1.558	2.639	0.570	
	(0.206)	(1.028)	(5.599)	(0.624)	
d_{ji}	2.018***	3.327***	2.536		1.206**
	(0.200)	(0.402)	(6.437)		(0.558)
$d_{ij}d_{ji}$	-1.601***	-0.982	-2.021		
	(0.208)	(1.056)	(8.388)		
$wealth_i$	0.019	-0.019	0.061*	-0.041	
	(0.017)	(0.028)	(0.036)	(0.051)	
$wealth_j$	0.016	0.014	0.031		-0.012
	(0.013)	(0.022)	(0.051)		(0.031)
$same\,religion$	0.178	-0.255	0.323	-0.058	-0.041
	(0.119)	(0.432)	(2.717)	(1.601)	(1.048)
$related$	0.140	-0.229	0.681	-0.079	0.133
	(0.274)	(0.633)	(18.080)	(1.946)	(1.760)
$distance$	-1.218***	-1.149*	-1.775	-0.083	0.020
	(0.263)	(0.604)	(1.282)	(1.608)	(1.191)
$hh\,members_i$	0.050	0.068	0.013		
	(0.078)	(0.098)	(0.270)		
$hh\,members_j$	0.029	0.028	0.192		
	(0.055)	(0.138)	(0.635)		
n_i				0.113**	
				(0.047)	
$female\,dependents_j$					-0.047
					(0.168)
$male\,dependents_j$					-0.222
					(0.157)
$constant$	-3.509***	-6.498***	-1.991	-2.478	-2.442*
	(0.299)	(0.530)	(2.032)	(2.208)	(1.409)

*** p<0.01, ** p<0.05, * p<0.1. Dyadic-robust standard errors in parentheses.

We worry that what household i reported as a gift was reported as a loan by j. Misclassification would affect estimated reporting propensities and hence may affect inference. To investigate whether misclassification affected our results, we reestimate the baseline model using combined gifts and loans as the dependent variable. Results are not shown here to save space. All the coefficients of interest are similar to those reported in Table 2 for gifts. Misclassification therefore does not seem to explain our results. In the reporting equation for transfer recipients, the number of male dependents is negative and significant at the 10% level. This provides some support to the idea that under-reporting of transfers received is to avoid detection by other household members – a point already made by Anderson and

Baland (2002).

4.2 Robustness analysis

To further investigate the robustness of our results, we reestimate the ML model (2) to (5) only including d_{ij} and d_{ji} and an intercept in the mis-reporting equations. Results, not shown here, are similar to those reported in Tables 2 and 3. In contrast, if we omit d_{ij} and d_{ji} from the mis-reporting equations, the results are dramatically different. In particular, the coefficient of $d_{ij}d_{ji}$ in the $\Pr(\tau = 1)$ equation becomes large and positive, and has a large t-value.[18] These findings are consistent with the discussion and simulation results presented in Section 2. They confirm that our ML estimator represents an improvement over logit only if we include d_{ij} and d_{ji} in the mis-reporting equations, as done in Tables 2 and 3.

Next, we repeat the analysis adding the weighted version of d_{ij} and d_{ji} as additional regressors in the reporting equations. Everything else is unchanged. Results, reported in Appendix A in Table A5 for gifts and Table A6 for loans, are very similar to those reported in Tables 2 and 3, and the new variables are not significant, with the exception of weighted d_{ij} for gifts.

We also re-estimate the model with different sets of regressors. Convergence is generally smooth for a moderately sized set of regressors as the ones of Table 2 and 3, and estimated coefficients for the key regressors, self-declared links and relational attributes, are similar across specifications. A few regressors in columns (4) and (5) are sufficient to get stable estimates for $\Pr(\tau = 1)$ as long as we include those variables that impact both the propensity to declare the transfer and the likelihood of a transfer itself. Including significant regressors in the mis-reporting equations increases the difference between logit results in columns (1) and (2) and the ML results in column (3). However, identification gets more problematic if we include partner's characteristic in the mis-reporting equations (*i.e.*, j's characteristics in $\Pr(G = 1|\tau = 1)$ and i's characteristics in $\Pr(R = 1|\tau = 1)$). The results presented here should thus be interpreted as based on these exclusion assumptions.

4.3 Mis-reported links

So far we have focus our concerns about mis-reporting on the flows of gifts and loans between households. We now consider the possibility that the first-round answers do not reflect respondents' desire to link, as Comola and Fafchamps (2009) have concluded for this dataset and as we have assumed until now. We now explore the possibility that first-round answers reflect mis-reported links instead.

---

[18]Virtually identical results for $\Pr(\tau = 1)$ are obtained if the $\Pr(G = 1|\tau = 1)$ and $Pr(R = 1|\tau = 1)$ only include an intercept (see Table A4 in Appendix A).

The difference is important because it would bias our test results in favor of unilateral link formation. To illustrate, let $g_{ij} = g_i$ be the true (unobserved) link between i and j and let g_{ij}^i and g_{ij}^j be reported links by i and j, respectively. Assume that g_{ij}^i and g_{ij}^j differ because of under-reporting. We have $g_{ij}^i = 1 \Rightarrow g_{ij} = 1$ and $g_{ij}^j = 1 \Rightarrow g_{ij} = 1$. Hence $\max(g_{ij}^i, g_{ij}^j) = g_{ij}^i + g_{ij}^j - g_{ij}^i g_{ij}^j = 1 \Rightarrow g_{ij} = 1$. Let τ_{ij} be a subsequent transfer between i and j. Since $Pr(\tau_{ij} > 0)$ is a strictly increasing function of g_{ij}, we obtain $Pr(\tau_{ij} > 0) = \lambda(g_{ij})$ $= \lambda(\alpha g_{ij}^i + \alpha g_{ij}^j - \alpha g_{ij}^i g_{ij}^j)$. This shows that if g_{ij}^i and g_{ij}^j are erroneously assumed to represent willingness to link and used to estimate (1), results would induce us to conclude in favor of unilateral link formation.

This can be corrected if we have variables z_{ij} and z_{ji} that are systematically correlated with the desire to link of i and j, respectively. We can use z_{ij} to instrument g_{ij}^i, extracting the systematic component of g_{ij}^i that is due to desire to link d_{ij}, and similarly use z_{ji} to instrument g_{ij}^j. Identification in (1) is achieved if there are enough dyads for which the predicted desire to link of i and j are sufficiently different.

We report in Table 4 estimation results for gifts. The possible endogeneity of d_{ij} and d_{ji} is handled using the control function approach, i.e., the residuals v_{ij} and v_{ij} from the first-stage linear probability regressions of d_{ij} and d_{ji} are included as additional regressors.[19] The instruments we use for i and j's desire to link are the popularity of the partner and the overlap in productive activities between the two households. Popularity is defined as the number of times household j get mentioned in the first-round question by households other than i. It proxies for unobserved characteristics of household j (sociability, charitable disposition) that makes other households wish to link with j. In the survey each adult individual mentions the productive activities he or she is involved into, divided in seven categories (casual labor, trade, crops, livestock rearing, assets, processing of agricultural products, and other off-farm work). Overlap in productive activities 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 income generating activity a. Presumably households with similar activities have more in common and find it easier to bond.

The results reported in Table 4 are still consistent with the hypothesis of unilateral link formation: estimated coefficients for instrumented d_{ij} and d_{ji} tend to be larger than in Table 2, a feature commonly observed in IV estimation, but they remain significantly positive and of similar magnitude as each other, in agreement with the unilateral link formation model. The coefficients of the logit equations (columns 2 and 3) follow the same pattern as in Table 2, and the coefficient of the interaction term $d_{ij} d_{ji}$ remains significantly negative in columns (1) and

---

[19]As Wooldridge (2007) reports, the control function approach is simpler and more precise than the standard 2SLS-like procedures when the model is non-linear in endogenous variables, as in our case. In particular he shows how, under certain restrictions on the conditional distribution of the endogenous regressors, it is not necessary to include the residuals of the quadratic term.

(2). We have experimented with different sets of instruments and instrumenting strategies and the results are remarkably consistent in supporting the unilateral link formation hypothesis.

Table 4. Results for gifts, instrumented

	(1)	(2)	(3)	(4)	(5)
	τ_{ij}^u	τ_{ij}^o	$\Pr(\tau = 1)$	$\Pr(G = 1 \mid \tau = 1)$	$Pr(R = 1 \mid \tau = 1)$
d_{ij}	9.932***	4.313**	8.528*	11.033	
	(2.302)	(2.112)	(5.107)	(7.756)	
d_{ji}	9.512***	12.976***	9.649***		8.121***
	(1.524)	(1.719)	(3.182)		(2.682)
$d_{ij}d_{ji}$	-0.568*	-0.910***	0.380		
	(0.321)	(0.341)	(2.548)		
$wealth_i$	0.025	-0.004	0.035	-0.022	
	(0.016)	(0.016)	(0.023)	(0.034)	
$wealth_j$	0.034	-0.024	0.056		-0.045***
	(0.025)	(0.022)	(0.065)		(0.017)
$same\ religion$	0.218**	0.184	0.256	-0.029	0.004
	(0.107)	(0.144)	(0.333)	(0.347)	(0.222)
$related$	-3.708***	-3.458***	-3.119	-3.047	-1.673
	(1.084)	(1.033)	(2.022)	(2.629)	(1.071)
$distance$	-0.579**	-0.966**	-0.484	-0.146	-0.208
	(0.264)	(0.483)	(0.649)	(0.871)	(0.560)
$hh\ members_i$	0.012	-0.046	0.027		
	(0.051)	(0.050)	(0.059)		
$hh\ members_j$	0.135	0.130	0.160		
	(0.085)	(0.088)	(0.136)		
n_i				-0.070*	
				(0.042)	
$female\ dependents_j$					-0.139
					(0.110)
$male\ dependents_j$					-0.161
					(0.140)
v_{ij}	-7.677***	-1.870	-6.281	-9.583	
	(2.340)	(2.223)	(4.670)	(7.390)	
v_{ji}	-6.770***	-9.912***	-6.882**		-6.343**
	(1.505)	(1.766)	(2.875)		(2.705)
$constant$	-4.073***	-5.223***	-3.589***	-0.675	-0.842**
	(0.263)	(0.364)	(0.523)	(0.420)	(0.403)

*** p<0.01, ** p<0.05, * p<0.1. Dyadic-robust standard errors in parentheses.

The usual caveat applies since the instruments are selected by us, based on a priori considerations regarding factors likely to affect the desire to link. It would have been better if data had been collected on desire to link. However, as Belot and Francesconi (2006) and

Hitsch, Hortacsu, and Ariely (2011) have shown, self-reported desire to link is subject to self-censoring: people often refrain from listing people they truly wish to link with but fear being rejected by. It should be possible to design a controlled experiment in which truth-telling is incentivized, or in which the true payoffs are known to the researcher, but experimental data of this kind at the moment do not exist. Given this, the results presented here should be taken as the best suggestive evidence available at this point.

It is also important to realize that, if link formation is indeed unilateral, then the first-round question *will* elicit information about the desire to link: when asked who they would turn to in an emergency, respondents simply list the households they would most wish to go to, even if a link does not already exist, since they know they can unilaterally create such a link. So, in this sense our evidence is internally consistent.

4.4 Estimates of under-reporting

A by-product of the estimation of the maximum likelihood model formed by equations (2) to (8) is that we can estimate the extent of under-reporting. This is achieved by comparing the frequency of giving or lending in data to the average frequency of the fitted $\Pr(\tau = 1)$ from Tables 2 (for gifts) and 3 (for loans). The result of these calculations is reported in Table 5.

Table 5. Estimates of mis-reporting

	gifts	loans
average fitted $\Pr(\tau_{ij} = 1)$	0.1568	0.1942
in data: declared by i	0.0709	0.0249
in data: declared by j	0.0587	0.0169
in data: declared by i or j	0.1011	0.0388
average fitted $\Pr(G = 1 \mid \tau = 1)$	0.3742	0.1138
average fitted $\Pr(R = 1 \mid \tau = 1)$	0.3110	0.0729

The average fitted propensity to give gifts from Table 5 is 16%, which is almost the same figure as the one we obtained in Section 3 without conditioning on regressors. For loans, the average fitted $\Pr(\tau_{ij} = 1)$ of 19% is larger than our earlier estimate of 14%. Based on these results, informal loans between villagers are more frequent than gifts, although much fewer of them are reported in the survey. Comparing these estimates to actually reported gifts and loans, we see that not taking mis-reporting into consideration leads to serious underestimation of the extent of gift giving and, especially, of lending and borrowing between villagers.

Table 5 also reports the average fitted propensity to report giving and receiving respectively. The average propensity for the giver to report a gift is 37% when we condition reporting on individual characteristics, compared to 47% when we do not. For recipients of

a gift, the propensity to report is 31%, compared to 39% when we do not condition on individual characteristics. Estimated reporting probabilities are much lower for loans. Lenders are estimated to report only 11% of loans – compared to 18% when we do not condition. Borrowers estimated to report as little as 7% of loans, versus 12% if we do not condition on household characteristics. If anything, estimated propensities to report gifts and loans fall when we allow them to depend on household characteristics.

The Nyakatoke data were collected with an unusually high level of care, using multiple survey rounds and interviewing each household member separately. Yet results suggests massive under-reporting. This casts some doubt on the reliability of reported gifts and loans in household surveys. For instance, many studies have found that reported gifts and loans are insufficient to insulate households against shocks. But if actual gifts and loans are much larger, these findings might be called into question. For instance, Rosenzweig (1988) reports that loans between households represent only 2% of the value of the shocks they face. If there is as much loan under-reporting in his data as in ours, the corrected figure is 10% – a five-fold increase.

5   Discussion

The proxies for desire to link d_{ij} and d_{ji} are based on the survey question *"Can you give a list of people [...] who you can personally rely on [...] and/or that can rely on you [...]?"* It is unclear whether answers to this question capture desire to provide help or to seek help – or both. If we had separate information on i's desire to give help to j and on i's desire to ask j for help, we could test whether it is one or the other that drives the exchange of gifts and informal loans between Nyakatoke households.

To illustrate this idea, let d_{ij}^g denote i's desire to help j and let d_{ji}^r denote j's desire to solicit help from i. With this information we could construct a more specific test as follows:

$$\tau_{ij} = \lambda(\alpha d_{ij}^g + \beta d_{ji}^r + \theta X_{ij})$$

If it is unilateral desire to give that determines transfers, then we should have $\alpha > 0$ and $\beta = 0$: transfers take place whenever i wishes to give something to j. This could reflect altruism, or perhaps moral norms regarding charitable giving. In contrast, if it is unilateral desire to receive help that determines τ_{ij}, transfers will take place whenever j wishes to receive something from i. Consequently we should obtain $\alpha = 0$ and $\beta > 0$. This could arise, for instance, because of social norms of redistribution, the existence of which has been argued by Hayami and Platteau (1996) for sub-Saharan Africa.[20]

---

[20]If j perfectly internalizes i's altruism towards him/her, then both α and β should in principle be positive. But since $d_{ji}^r = d_{ij}^g$ in this case, the $d_{ij}^g d_{ji}^r$ cross term will capture the effect of *both* d_{ij}^g and d_{ji}^r on transfers – and link formation will appear bilateral.

We do not have separate information about desire to give and desire to receive. But let us imagine for a moment that d_{ij} should in fact be interpreted as desire to give, i.e., $d_{ij} = d_{ij}^g$. If this were the case, then when we regress τ_{ij} on d_{ij} and d_{ji}, it is like estimating a model of the form:

$$\tau_{ij} = \lambda(\alpha d_{ij}^g + \beta d_{ji}^g + \theta X_{ij})$$

If transfers are unilaterally driven by the desire to give of the giver, then we should observe $\alpha > 0$ and $\beta = 0$. This is not what we observe in Tables 2, 3, and 4.

Alternatively, imagine that answers to the undirected question of round 1 measure desire to ask for help, i.e., $d_{ij} = d_{ij}^r$. In this case, when we regress τ_{ij} on d_{ij} and d_{ji}, it is like estimating a model of the form:

$$\tau_{ij} = \lambda(\alpha d_{ij}^r + \beta d_{ji}^r + \theta X_{ij})$$

If transfers are unilaterally driven by the recipient's desire to request assistance, then we should observe $\alpha = 0$ and $\beta > 0$. Once again, this is not what we observe in Tables 2, 3 and 4.

What inference can we draw from the above? First, there is no evidence that answers to the undirected question of round 1 should be interpreted as reflecting only desire to give or only desire to receive. If this had been the case, we should not have found d_{ij} and d_{ji} to be both significant in Tables 2, 3 and 4 with coefficients of equal magnitude. It follows that answers to the undirected question of round 1 were indeed undirected: they capture both desire to give and desire to receive.

Secondly, we cannot a priori tell whether d_{ij} captures desire to give and receive from the same person – as in a reciprocal relationship – or whether some d_{ij}'s capture desire to give and others capture desire to receive. But in the latter case, both types of d_{ij}'s would need to be present in the data in exactly the right proportions for α and β to be of equal magnitude. Since there is no particular reason for this to be the case, we find this possibility unlikely. It follows that d_{ij} most probably represents desire to enter in a reciprocal relationship – as indeed is suggested by the wording of the question, and by the difficulties that Fafchamps and Lund (2003) and De Weerdt and Dercon (2006) encountered when they sought to separately ask who respondents would turn to and who would turn to them.

6   Conclusion

The architecture, efficiency, and equity of social networks ultimately rests on whether individuals can create links unilaterally or whether the consent of both parties is required. In this paper we have proposed a simple test of unilateral versus bilateral link formation that combines dyad-specific information about flows – an indication that a link exists – and about

the desire to link. When implementing the test, it is essential to correct for mis-reporting. Self-reported flows are typically discordant: i may report a transfer to j while j reports no such transfer from i. The propensity to report a transfer is likely to be correlated with the desire to link. This creates a correlation between the desire to link and the propensity to report manifestation of a link that can affect the test, as demonstrated using simulations. We propose a maximum likelihood estimator to deal with mis-reporting.

We illustrate the methodology using detailed dyadic data on inter-household gifts and loans from the village of Nyakatoke in Tanzania. In these data, there are substantial discrepancies between gifts and loans reported by givers and receivers. We see no serious reason to suspect that respondents systematically over-report transfers they did not give or receive. We therefore focus on the results that assume under-reporting – presumably because respondents forget. An alternative version of the methodology that assumes over-reporting is presented in Appendix B.

We find no strong evidence in favor of bilateral link formation. We do, however, find reasonably convincing evidence in favor of the unilateral link formation hypothesis, except that flows are more likely to occur if both households wish to link with each other. Results are robust to different choices of regressors and model specification, including instrumentation of the desire to link variable. Given data limitations, we cannot formally test whether it is desire to give transfers or to request transfers that drives gifts and loans. But taken as a whole, the evidence is most consistent with transfers being driven by the desire to enter in a reciprocal relationship.

If this interpretation is correct, the evidence implies that if one household wishes to enter in a reciprocal relationship with another household, it can unilaterally do so – provided this other household is sufficiently close socially and geographically. This could arise, for instance, because inter-personal norms of reciprocation can be activated unilaterally by Nyakatoke villagers – as when giving to someone is a way of obligating him or her to reciprocate in the future (Platteau 2000). If confirmed by future research, the above interpretation could explain the puzzling findings of Fafchamps and Gubert (2007) and those of De Weerdt and Fafchamps (2011) using the same data. These authors find that, contrary to theoretical predictions, households do not appear more likely to have links with those who face less covariate risk. If households can wait after shocks are realized before deciding who to ask for help, they need not worry about covariate risk *ex ante*.

This interpretation ties with another surprising result of our analysis, namely that loans are less likely to be reported than gifts. It is easy to see why borrowers would fail to report the loans they have received, but why would lenders do so? Much of the theoretical discourse about risk sharing has emphasized repeated games and reputation sanctions (Coate and Ravallion 1993, Kocherlakota 1996, Ligon, Thomas and Worrall 2001). Yet, if lenders hide the loans they make, it is hard to see how group reputational sanctions could be imposed.

There must therefore be a cost to the lender from publicizing loans. One possible explanation is that lenders fear that disclosing loans reveals they have money they do not need, and this could attract additional requests for help, as suggested by the work of Goldberg (2010). A similar point is made by Anderson and Baland (2002) who argue that secrecy within households serves to avoid claims on resources by spouses. If link formation was bilateral, it would be possible to refuse to assist others and secrecy would not be necessary.

References

[1] Anderson, Siwan and Jean-Marie Baland (2002), 'The Economics of Roscas and Intra-household Resource Allocation', *Quarterly Journal of Economics*, 117(3): 963-95

[2] Arcand, Jean-Louis and Marcel Fafchamps (2012), 'Matching in Community-Based Organizations', *Journal of Development Economics*, forthcoming

[3] Ashenfelter, Orley and Alan Krueger (1994), 'Estimates of the Economic Return to Schooling From a New Sample of Twins', *American Economic Review*, 84: 1157-73

[4] Bala, Venkatesh and Sanjeev Goyal (2000), "A Non-Cooperative Model of Network Formation," *Econometrica*, 68(5): 1181-1229, September

[5] Belot, Michele and Marco Francesconi (2006), 'Can Anyone Be The One? Evidence on Mate Selection from Speed Dating', *CEPR Discussion Paper No. 5926*, London

[6] Christakis, Nicholas A. and James H. Fowler (2009), *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*, Little, Brown and Company, London

[7] Coate, Stephen and Martin Ravallion (1993), 'Reciprocity Without Commitment: Characterization and Performance of Informal Insurance Arrangements', *Journal of Development Economics*, 40: 1-24

[8] Comola, Margherita (2008), 'The Network Structure of Mutual Support Links: Evidence from Rural Tanzania', Paris School of Economics Working Paper 2008/74

[9] Comola, Margherita and Marcel Fafchamps (2009), 'Testing Unilateral and Bilateral Link Formation', Department of Economics, Oxford (mimeograph)

[10] De Weerdt, Joachim and Stefan Dercon (2006), 'Risk-Sharing Networks and Insurance Against Illness', *Journal of Development Economics*, 81(2): 337-56

[11] De Weerdt, Joachim and Marcel Fafchamps (2011), 'Social Identity and The Formation of Health Insurance Networks', *Journal of Development Studies*, forthcoming

[12] Duncan, Greg and Daniel Hill (1985), 'An Investigation of the Extent and Consequences of Measurement Error in Labor Economic Survey Data', *Journal of Labor Economics* 3: 508-522

[13] Erdil, Aytek and Haluk Ergin (2007), 'What's the Matter with Tie-breaking? Improving Efficiency in School Choice', Department of Economics, Oxford University

[14] Fafchamps, Marcel and Susan Lund (2003), 'Risk Sharing Networks in Rural Philippines', *Journal of Development Economics*, 71: 261-87

[15] Fafchamps, Marcel and Flore Gubert (2007), 'The Formation of Risk Sharing Networks', *Journal of Development Economics*, 83(2): 326-50

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

[17] Fox, Jeremy (2011), "Estimating Matching Games with Transfers," Department of Economics, University of Michigan (mimeograph)

[18] Glaeser, Edward, Bruce Sacerdote,and Jose Scheinkman (1996), "Crime and Social Interactions", *Quarterly Journal of Economics*, 111: 507-48, 1996

[19] Goldberg, Jessica (2010), "The Lesser of Two Evils: The Roles of Social Pressure and Impatience in Consumption Decisions", Department of Economics, University of Michigan. December (mimeograph)

[20] Goyal, Sanjeev (2007), *Connections: An Introduction to the Economics of Networks*, Princeton University Press, Princeton and Oxford

[21] Hitsch, Gunter J., Ali Hortacsu, Dan Ariely (2011), 'Matching and Sorting in Online Dating', *American Economic Review*, forthcoming

[22] Hong, Han and Elie Tamer (2003), "A simple estimator for nonlinear error in variable models," *Journal of Econometrics*, 117(1): 1-19

[23] Jackson, Matthew O. (2009), *Social and Economic Networks*, Princeton University Press, Princeton

[24] Jackson, Matthew O. and Asher Wolinsky (1996), "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71(1): 44-74

[25] Kocherlakota, Narayana R. (1996), 'Implications of Efficient Risk Sharing Without Commitment', *Rev. Econ. Stud.*, 63(4): 595-609

[26] Lee, Yun-Suk and Linda J. Waite (2005), 'Husbands and Wives Time Spent on Housework: A Comparison of Measures', *Journal of Marriage and Family*, 67: 328-336

[27] Ligon, Ethan, Jonathan P. Thomas, and Tim Worrall (2001), 'Informal Insurance Arrangements in Village Economies', *Review of Economic Studies*, 69(1): 209-44

[28] Liu, Xiaodong, Eleonora Patacchini, Yves Zenou, and Lung-Fei Lee (2011), "Criminal Networks: Who is the Key Player?", Research Papers in Economics 2011:7, Stockholm University, Department of Economics

[29] Platteau, Jean-Philippe and Yujiro Hayami (1996), 'Resource Endowments and Agricultural Development: Africa vs. Asia', University of Namur and Aoyama Gakuin University, Tokyo. Paper presented at the IEA Round Table Conference The Institutional Foundation of Economic Development in East Asia, Tokyo, 16-19

[30] Platteau, Jean-Philippe (2000), *Institutions, Social Norms, and Economic Development*, Harwood Academic Publishers, Amsterdam

[31] Rosenzweig, Mark R. (1988), "Risk, Implicit Contracts and the Family in Rural Areas of Low-Income Countries," *Economic Journal*, 98: 1148-1170, December

[32] Roth, Alvin and Marilda Sotomayor (1990), *Two-Sided Matching*, Cambridge University Press, Cambridge

[33] Schennach, Susanne (2004), "Estimation of Nonlinear Models with Measurement Error," *Econometrica*, 72 (1): 33-75

[34] Snijders, Tom A.B., Johan Koskinen, and Michael Schweinberger (2010), "Maximum Likelihood Estimation for Social Network Dynamics", *Annals of Applied Statistics*, forthcoming

[35] Steglich, Christian E.G., Tom A.B. Snijders, and Michael Pearson (2010), 'Dynamic Networks and Behavior: Separating Selection from Influence', *Sociological Methodology*, (forthcoming)

[36] Udry, Christopher (1994), 'Risk and Insurance in a Rural Credit Market: An Empirical Investigation in Northern Nigeria', *Review of Economic Studies*, 61(3): 495-526

[37] Vandenbossche, Joost and Thomas Demuynck (2010), 'Network Formation with Heterogeneous Agents and Absolute Friction', Working Paper 10/642, Ghent University, Faculty of Economics and Business Administration

[38] Wooldridge, Jeffrey (2007), 'Control Function and Related Methods', lecture for the 2007 NBER Summer Institute *What's New in Econometrics?*, online at *http://www.nber.org/WNE/lect_6_controlfuncs.pdf*

Appendix A

Table A1. Quintiles of declared loans and gifts

Information given by:	Gifts		Loans	
	giver	receiver	giver	receiver
nonzero obs.	996	824	350	237
cut-off values:				
0-20%	240	200	456	400
20-40%	500	450	900	700
40-60%	1000	850	1500	1532
60-80%	1796	1800	3000	3000
80-100%	39400	46800	60000	40000

Note: the total sample size is 14042 dyads, and the quintiles cut-off values are computed on nonzero observations only. Values expressed in tzs.

Table A2. Testing whether desire to link

dependent variable: d_{ij}	
$popularity_i$	0.031
	(0.020)
$wealth_i$	0.019
	(0.015)
$popularity_j$	0.100***
	(0.006)
$wealth_j$	0.012**
	(0.006)
$constant$	-4.032***
	(0.119)
	(0.061)

*** p<0.01, ** p<0.05, * p<0.1.

Estimator is logit. Dyadic-robust standard errors in parentheses.

Table A3. Marginal effects

	$Pr(G=1\|\tau=1)$			
	gifts		loans	
	coeff.	mfx.	coeff.	mfx.
d^*_{ij}	1.4917	0.5388	0.5702	0.0425
$wealth_i$	-0.0354	-0.0123	-0.0411	-0.0018
$same\,religion^*$	0.0246	0.0086	-0.0580	-0.0025
$related^*$	0.4332	0.1634	-0.0795	-0.0032
$distance$	-0.5854	-0.2036	-0.0835	-0.0036
n_i	0.0259	0.0090	0.1126	0.0049
	$Pr(R=1\|\tau=1)$			
	gifts		loans	
	coeff.	mfx.	coeff.	mfx.
d^*_{ji}	1.9196	0.6625	1.2059	0.0707
$wealth_j$	-0.0447	-0.0125	-0.0120	-0.0002
$same\,religion^*$	0.0118	0.0033	-0.0409	-0.0006
$related^*$	0.6136	0.2092	0.1331	0.0022
$distance$	-0.5326	-0.1496	0.0201	0.0003
$female\,dependents_j$	-0.1495	-0.0420	-0.0467	-0.0006
$male\,dependents_j$	-0.1906	-0.0535	-0.2221	-0.0031

*dy/dx is for discrete change of dummy variable from 0 to 1

Table A4. Constant-only model

	gifts	loans
$\Pr(\tau = 1)$		
d_{ij}	3.222***	2.856***
	(0.396)	(1.074)
d_{ji}	3.749***	3.184**
	(0.534)	(1.457)
$d_{ij}d_{ji}$	13.490***	10.787***
	(1.037)	(2.972)
$wealth_i$	0.064***	0.036
	(0.013)	(0.029)
$wealth_j$	0.083**	0.021
	(0.036)	(0.017)
$same\,religion$	0.519***	0.208
	(0.120)	(0.155)
$blood\,link$	2.423***	0.950
	(0.359)	(0.635)
$distance$	-2.049***	-1.799***
	(0.466)	(0.432)
$hh\,members_i$	0.115***	0.065
	(0.040)	(0.109)
$hh\,members_j$	0.235*	0.106
	(0.123)	(0.094)
$constant$	-3.483***	-2.270***
	(0.388)	(0.610)
$\Pr(G = 1 \mid \tau = 1)$		
$constant$	0.143	-1.528***
	(0.225)	(0.403)
$\Pr(R = 1 \mid \tau = 1)$		
$constant$	-0.228*	-1.986***
	(0.133)	(0.320)

*** p<0.01, ** p<0.05, * p<0.1.

Dyadic-robust standard errors in parentheses.

Table A5. Results for gifts with weighted d_{ij}

	(1) $\Pr(\tau = 1)$	(2) $\Pr(G = 1 \mid \tau = 1)$	(3) $Pr(R = 1 \mid \tau = 1)$
d_{ij}	2.578***	1.081***	
	(0.376)	(0.278)	
d_{ji}	2.817***		1.819***
	(0.306)		(0.324)
$d_{ij}d_{ji}$	-0.218		
	(0.987)		
$wealth_i$	0.081***	-0.035**	
	(0.016)	(0.016)	
$wealth_j$	0.105		-0.044***
	(0.067)		(0.015)
$same\,religion$	0.533**	0.024	0.010
	(0.251)	(0.212)	(0.194)
$related$	2.002***	0.415	0.597
	(0.776)	(0.504)	(0.379)
$distance$	-1.675**	-0.585	-0.537
	(0.668)	(0.547)	(0.490)
$hh\,members_i$	0.109**		
	(0.043)		
$hh\,members_j$	0.261		
	(0.169)		
n_i		0.027**	
		(0.013)	
$weighted\,d_{ij}$		0.796*	
		(0.435)	
$female\,dependents_j$			-0.150
			(0.143)
$male\,dependents_j$			-0.189
			(0.133)
$weighted\,d_{ji}$			0.200
			(0.472)
$constant$	-3.526***	-0.283	-0.209
	(0.542)	(0.597)	(0.360)

*** p<0.01, ** p<0.05, * p<0.1. Dyadic-robust standard errors
in parentheses.

Table A6. Results for loans with weighted d_{ij}

	(1) $\Pr(\tau = 1)$	(2) $\Pr(G = 1 \mid \tau = 1)$	(3) $Pr(R = 1 \mid \tau = 1)$
d_{ij}	2.639	0.483	
	(5.653)	(0.548)	
d_{ji}	2.531		1.260**
	(6.476)		(0.598)
$d_{ij}d_{ji}$	-2.027		
	(8.449)		
$wealth_i$	0.061*	-0.041	
	(0.037)	(0.051)	
$wealth_j$	0.031		-0.012
	(0.052)		(0.031)
$same\,religion$	0.324	-0.058	-0.042
	(2.742)	(1.616)	(1.058)
$related$	0.681	-0.082	0.133
	(18.455)	(1.978)	(1.809)
$distance$	-1.775	-0.081	0.020
	(1.284)	(1.619)	(1.195)
$hh\,members_i$	0.013		
	(0.272)		
$hh\,members_j$	0.192		
	(0.644)		
n_i		0.113**	
		(0.048)	
$weighted\,d_{ij}$		0.171	
		(0.466)	
$female\,dependents_j$			-0.047
			(0.167)
$male\,dependents_j$			-0.222
			(0.158)
$weighted\,d_{ji}$			-0.103
			(0.578)
$constant$	-1.991	-2.479	-2.439*
	(2.038)	(2.220)	(1.418)

*** p<0.01, ** p<0.05, * p<0.1. Dyadic-robust standard errors in parentheses.

36

Appendix B

In this appendix we explain how the model can be estimated when flows are over-estimated instead of under-estimated, i.e., when survey respondents reported flows that did not actually take place. In the context of our data, this could arise because people wish they had made these transfers but were ashamed to admit to enumerators that they did not, and so made up some numbers. Whether or not this is a reasonable assumption depends on the context. For our data, it is unlikely, but we wish to investigate the robustness of our results to this assumption. It should be noted that, in our data, only 10% of household pairs both declare a gift and only 3% both declare a loan. This means that, under the assumption of over-reporting, the number of loan observations for which $\tau = 1$ is small, making inference more difficult and possibly creating identification and convergence problems. It is nevertheless instructive to investigate whether we obtain results that do not contradict our earlier conclusions regarding unilateral link formation.

Formally, we now assume that unless both i and j declare a transfer, it did not take place. As before, we assume that response errors are independent between i and j, an assumption that is required for identification. With these assumptions we can write:

$$\Pr(G = 1, R = 0) = \Pr(\tau = 0) * \Pr(G = 1|\tau = 0) * \Pr(R = 0|\tau = 0) \tag{12}$$

$$\Pr(G = 0, R = 1) = \Pr(\tau = 0) * \Pr(G = 0|\tau = 0) * \Pr(R = 1|\tau = 0) \tag{13}$$

$$\Pr(G = 0, R = 0) = \Pr(\tau = 0) * \Pr(G = 0|\tau = 0) * \Pr(R = 0|\tau = 0) \tag{14}$$

$$\Pr(G = 1, R = 1) = 1 - \Pr(G = 1, R = 0) - \Pr(G = 0, R = 1) - \Pr(G = 0, R = 0) \tag{15}$$

Equations (12) to (15) express the data generating process in terms of three probabilities: $P(\tau = 0)$, $P(G = 1|\tau = 0)$ and $P(R = 1|\tau = 0)$. As before, we assume that these three probabilities can be represented by three distinct logit functions $\lambda(.)$ as follows:

$$\Pr(\tau = 0) = \lambda_T(\beta'_\tau X_\tau) \tag{16}$$

$$\Pr(G = 1|\tau = 0) = \lambda_G(\beta'_G X_G) \tag{17}$$

$$\Pr(R = 1|\tau = 0) = \lambda_R(\beta'_R X_R) \tag{18}$$

The main equation of interest now is $\Pr(\tau = 0)$. Our objective remains to test whether transfers are unilateral or bilateral.

A testing strategy in terms of $\tau = 0$ is provided by model (19). Let $h_{ij} = 1$ if $\tau_{ij} = 0$, i.e., h_{ij} is an indicator variable that takes value 1 is i does *not* give something to j. Similarly define $u_{ij} = 1 - d_{ij}$, i's lack of desire to link with j. In the unilateral model of link formation, $h_{ij} = 1$ if both i and j are unwilling to link, i.e., if $\{u_{ij}, u_{ji}\} = \{1, 1\}$. In contrast, in a bilateral model of link formation, $h_{ij} = 1$ if either i or j are unwilling to link, i.e., if

$\{u_{ij}, u_{ji}\} = \{1, 0\}, \{0, 1\}$, or $\{1, 1\}$. We estimate a model of the form:

$$h_{ij} = \lambda(\alpha' u_{ij} + \beta' u_{ji} + \gamma' u_{ij} u_{ji} + \theta' X_{ij}) \tag{19}$$

If risk sharing is unilateral, transfers do not take place only if $\{u_{ij}, u_{ji}\} = \{1, 1\}$. It follows that $\alpha' = \beta' = 0$ and $\gamma' > 0$. In contrast, if risk sharing is bilateral, we have $\gamma' = -\beta' = -\alpha' < 0$.

Estimation results are presented in Table A7 for gifts and Table A8 for loans. Results are less conclusive than those reported in Tables 2 and 3. Coefficients α' and β' are significantly positive in all three gift regressions (Table A7), that is, for the two logit models and for the ML model that corrects for mis-reporting. The magnitude of the estimated coefficients is smaller in the ML model. In the loan regressions (Table A8), α' and β' are positive in all three regressions, although only significantly so in the logit regressions. This evidence is consistent with the unilateral link formation hypothesis. However, γ', the coefficient of $(1 - d_{ij})(1 - d_{ji})$, is also positive and significant in several of the regressions, which is consistent with bilateral link formation. Hence, when we assume that there is no only over-reporting, the evidence is ambiguous in the sense that it supports both models – or a hybrid of the two, where links are formed in a way that is largely unilateral but contains some bilateral element as well. Regarding the reporting equations, we find, as before, that the likelihood of reporting a gift increases in d_{ij} and d_{ji}. It also increases significantly with kinship, geographical proximity, and co-religion. For loans (Table A8), results show that the likelihood of reporting a loan increases with d_{ij} and d_{ij} and with geographical proximity.

Table A7. Results for gifts, over-reporting

	(1) τ_{ij}^u	(2) τ_{ij}^o	(3) $\Pr(\tau=0)$		(4) $\Pr(G=1|\tau=0)$	(5) $\Pr(R=1|\tau=0)$
$1-d_{ij}$	1.795***	1.491***	1.143***	d_{ij}	2.724***	
	(0.275)	(0.206)	(0.323)		(0.251)	
$1-d_{ji}$	2.113***	2.223***	1.889***	d_{ji}		3.036***
	(0.252)	(0.167)	(0.209)			(0.249)
$(1-d_{ij})(1-d_{ji})$	0.681**	1.036***	1.595***	$wealth_i$	0.026***	0.029
	(0.306)	(0.313)	(0.508)		(0.009)	(0.019)
$wealth_i$	-0.058***	-0.024	-0.022	$wealth_j$		
	(0.012)	(0.015)	(0.020)			
$wealth_j$	-0.066**	0.001	0.007	same religion	0.353***	0.371***
	(0.030)	(0.024)	(0.027)		(0.107)	(0.132)
same religion	-0.421***	-0.353***	-0.286	related	1.417	1.914***
	(0.099)	(0.132)	(0.201)		(0.976)	(0.556)
related	-1.728***	-0.944***	0.048	distance	-1.200**	-1.208**
	(0.284)	(0.294)	(1.477)		(0.511)	(0.508)
distance	1.711***	1.789***	1.601	n_i	0.026*	
	(0.294)	(0.476)	(1.110)		(0.015)	
hh members$_i$	-0.084***	-0.069	-0.068	female dependents$_j$		-0.172*
	(0.032)	(0.060)	(0.074)			(0.101)
hh members$_j$	-0.216**	-0.169**	-0.214**	male dependents$_j$		-0.071
	(0.098)	(0.086)	(0.107)			(0.148)
constant	-0.958***	0.214	0.686**	constant	-3.170***	-3.246***
	(0.370)	(0.275)	(0.285)		(0.322)	(0.288)

*** p<0.01, ** p<0.05, * p<0.1. Dyadic-robust standard errors in parentheses.

Table A8. Results for loans, over-reporting

	(1) τ^u_{ij}	(2) τ^o_{ij}	(3) $\Pr(\tau=0)$		(4) $\Pr(G=1|\tau=0)$	(5) $\Pr(R=1|\tau=0)$
$1-d_{ij}$	0.366**	0.576*	0.023	d_{ij}	1.881***	
	(0.185)	(0.321)	(0.572)		(0.323)	
$1-d_{ji}$	0.418**	2.344**	3.037	d_{ji}		2.355***
	(0.181)	(1.096)	(2.577)			(0.321)
$(1-d_{ij})(1-d_{ji})$	1.601***	0.982	0.768			
	(0.208)	(1.056)	(2.541)			
$wealth_i$	-0.019	0.019	0.133*	$wealth_i$	0.003	0.007
	(0.017)	(0.028)	(0.079)		(0.011)	(0.023)
$wealth_j$	-0.016	-0.014	-0.018	$wealth_j$		
	(0.013)	(0.022)	(0.031)			
$same\,religion$	-0.178	0.255	0.266	$same\,religion$	0.192	0.222
	(0.119)	(0.432)	(0.762)		(0.158)	(0.181)
$related$	-0.140	0.229	12.877***	$related$	0.255	0.547
	(0.274)	(0.633)	(0.759)		(0.415)	(0.482)
$distance$	1.218***	1.149*	1.129	$distance$	-1.257***	-1.027**
	(0.263)	(0.604)	(1.269)		(0.456)	(0.457)
				n_i	0.114***	
					(0.016)	
$hh\,members_i$	-0.050	-0.068	0.003			
	(0.078)	(0.098)	(0.207)			
$hh\,members_j$	-0.029	-0.028	-0.107	$female\,dependents_j$		0.030
	(0.055)	(0.138)	(0.256)			(0.165)
				$male\,dependents_j$		-0.114
						(0.164)
$constant$	1.125***	2.595***	2.615*	$constant$	-4.229***	-4.230***
	(0.292)	(0.620)	(1.335)		(0.275)	(0.315)

*** p<0.01, ** p<0.05, * p<0.1. Dyadic-robust standard errors in parentheses.