

Comola, Margherita; Fafchamps, Marcel

**Working Paper**

## Estimating Mis-reporting in Dyadic Data: Are Transfers Mutually Beneficial?

IZA Discussion Papers, No. 8664

**Provided in Cooperation with:**

IZA – Institute of Labor Economics

*Suggested Citation:* Comola, Margherita; Fafchamps, Marcel (2014) : Estimating Mis-reporting in Dyadic Data: Are Transfers Mutually Beneficial?, IZA Discussion Papers, No. 8664, Institute for the Study of Labor (IZA), Bonn

This Version is available at:

<https://hdl.handle.net/10419/106604>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

IZA DP No. 8664

**Estimating Mis-reporting in Dyadic Data:  
Are Transfers Mutually Beneficial?**

Margherita Comola  
Marcel Fafchamps

November 2014

# Estimating Mis-reporting in Dyadic Data: Are Transfers Mutually Beneficial?

**Margherita Comola**

*Paris School of Economics  
and IZA*

**Marcel Fafchamps**

*Stanford University  
and IZA*

Discussion Paper No. 8664  
November 2014

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0  
Fax: +49-228-3894-180  
E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## ABSTRACT

### **Estimating Mis-reporting in Dyadic Data: Are Transfers Mutually Beneficial?\***

Many studies have used self-reported dyadic data without exploiting the pattern of discordant answers. In this paper we propose a maximum likelihood estimator that deals with mis-reporting in a systematic way. We illustrate the methodology using dyadic data on inter-household transfers (gifts and loans) from the village of Nyakatoke in Tanzania, investigating whether observed transfers are mutually beneficial, i.e. in the self-interest of both parties involved. Our results suggest that mutual self-interest is not a necessary condition for transfers between households who are sufficiently close socially and geographically to take place, and we show that not taking reporting bias into account leads to serious underestimation of the total amount of transfers between villagers.

JEL Classification: C13, C51, D85

Keywords: social networks, dyadic data, reporting bias, informal transfers

Corresponding author:

Margherita Comola  
Paris School of Economics  
Université Paris 1 Panthéon-Sorbonne  
106-112 Boulevard de l'Hopital  
75647 Paris Cedex 13  
France  
E-mail: [margherita.comola@psemail.eu](mailto:margherita.comola@psemail.eu)

---

\* We are indebted to Joachim De Weerd for sharing his data and answering our questions. We benefitted from useful comments from seminar participants at the Paris School of Economics, Oxford University, Norwegian School of Economics and Business Administration, Stanford and Yale. All remaining errors are our own.

# 1 Introduction

It is increasingly common for surveys to collect information on social links and interpersonal flows – *e.g.*, friendship, loans and gifts, advice, referral. In particular, 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’, ‘with whom do you exchange information’, or ‘are you related to X’ (*e.g.*, Fafchamps and Lund 2003; Christakis and Fowler 2009; Steglich, Snijders, and Pearson 2010; Banerjee *et al.* 2013). In principle answers to these questions should agree: if for instance  $i$  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  $j$  (Ball and Newman 2013). Until now mis-reporting has typically been ignored and estimation has proceeded using information reported by  $i$ ,  $j$ , or a combination of the two (*e.g.*, Glaeser, Sacerdote and Scheinkman 1996; Snijders, Koskinen and Schweinberger 2010; Liu, Patacchini, Zenou and Lee 2011). However, failing to properly account for mis-reporting may bias the estimation results to a large extent. This paper investigates how mis-reporting affects estimation and inference in self-reported dyadic data, and proposes an estimator which deals with it.

We illustrate our methodology using data on informal transfers (loans and gifts) from rural Tanzania. Informal transfers have been recognized a great importance for development, since they represent the major source of insurance against idiosyncratic shocks when formal institutions are weak or absent. In particular, many studies have investigated informal transfers by using self-reported dyadic transfer data. For instance, Fafchamps and Lund (2003) and De Weerd and Fafchamps (2011) use transfers information obtained from one of the two households only, while Fafchamps and Gubert (2007) combine the two answers to construct a unique measure of transfers. All these studies neither exploit the systematic pattern of discordant answers in the data, nor investigate the consequences of mis-reporting. In this paper we show that these studies may underestimate the amount of informal transfers occurred at the village level by a very large extent, which has in turn important policy implications.

Faced with discordant transfer data, researchers typically rely on *ad hoc* assumptions. They may assume that if either  $i$  or  $j$  report a transfer, then a transfer 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 transfer between  $i$  and  $j$  took place only if both  $i$  and  $j$  reported it; this is equivalent to assuming that when both reports agree they

are true statements, and all observed discordances are due to over-reporting. Both assumptions rule out the possibility that a transfer occurred but was not declared by anyone, or that a transfer declared by both parts in reality did not occur. We propose a maximum likelihood estimator that deals with discordant answers in a systematic way. Our estimator accounts separately for the propensity of  $i$  and  $j$  to report a transfer, which may depend on respondent observables. It forces the researcher to assume either under- or over-reporting in the underlying data generation process – but also allows to investigate the sensitivity of the findings to assuming one or the other.<sup>1</sup> While there is an established literature on measurement error in binary variables (*e.g.* Hong and Tamer 2003; Schennach 2004), to the best of our knowledge no specific solution for dyadic data has been offered. The method we propose to correct for mis-reporting is of particular interest to researchers studying social networks but it is also suitable for any pairwise data with two discordant self-reported 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).

Simulations suggest that more accurate inference is obtained with our mis-reporting correction. In particular, we show that estimation results are sensitive to mis-reporting if the propensity to report is correlated with the variables of interest. To understand why, consider the following example. Imagine we have data on households’ wealth and transfers between them. Assume that wealthy households tend to give and receive more transfers, but they are also less likely to subsequently report the transfer. If the researcher assumes that a transfer took place only if at least one side  $i$  or  $j$  reported it, the estimated coefficient of the impact of  $i$ ’s wealth on the probability of a transfer from  $i$  to  $j$  will be biased downwards. This is because the researcher observes transfers more frequently when giver and receiver households are less wealthy – but this is due to the reporting pattern.

Our empirical illustration uses dyadic data from the village of Nyakatoke in Tanzania. These data contain detailed information on all transfers between all households in the village, and display massive discrepancies in survey responses about transfers given

---

<sup>1</sup>The appropriateness of assuming under- versus over-reporting depends on the context. In many cases it is reasonable to assume that the main reason for discrepancies is under-reporting: a transfer took place but one of the parts involved forgot to report it to enumerators. It can also happen that links or flows are suspected to be over-reported, as when individuals inflate the number of their friends or sexual partners.

and received. Specifically, we aim at investigating whether observed transfers are what we define as ‘mutually beneficial’, that is, in the self-interest of both parties involved. Since the rule of link formation has dramatic implications for the efficiency and equity of the resulting network (*e.g.*, Jackson and Wolinsky 1996; Bala and Goyal 2000), it is important for the researcher to empirically test whether link formation is grounded in mutual self-interest or not. Building on Comola and Fafchamps (2014) we use the answer to a first-round question on whom people would ask and/or provide help to as proxy for household’s desire to link with others in the village - our previous work and additional evidence presented here suggests that this is a reasonable assumption. We find reasonably strong evidence against the hypothesis of mutually beneficial link formation: our results suggest that if a household wishes to enter in a gift-exchange relationship with another household that is sufficiently close socially and geographically, it can do so. Flows between households are nevertheless more likely if both households wish to link. We interpret these findings as suggesting that surveyed households find it difficult to extricate themselves from social and familial obligations to assist others in need. This stands in contrast with much of the economic literature on risk sharing which emphasizes self-interest as basis for mutual support (Coate and Ravallion 1993, Ligon, Thomas and Worrall 2001). As a byproduct of our estimation, we show that not taking mis-reporting into account leads to serious underestimation of the total amount of gifts and loans between Nyakatoke villagers. These results cast some doubt on the reliability of previous results based on reported transfers in household surveys. In particular, 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 may 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 correct figure is probably closer to 10% – a five-fold increase.

The paper is organized as follows. In Section 2 we describe the estimation strategy and simulation analysis. The data and the test of mutually beneficial links are illustrated in Section 3. Results are discussed in Section 4. Section 5 focuses on the estimates of under-reporting, and Section 6 concludes. Additional tables are reported in Appendix A. Appendix B discusses our proxy variable for desire to link in Nyakatoke data. Appendix C illustrates how to implement our estimator assuming over-reporting.

## 2 Estimation strategy

### 2.1 The estimator

In our empirical analysis,  $\tau_{ij}$  refers to a binary transfer from  $i$  to  $j$  over a given time interval. More generally, we think of  $\tau_{ij}$  as capturing any manifestation of a social link, typically a flow of money, goods, or favors. Our objective is to estimate a regression model of the form:

$$\Pr(\tau_{ij} = 1) = \lambda(\beta_\tau X_\tau^{ij}) \quad (1)$$

where  $X_\tau^{ij}$  is a vector of controls for dyad  $ij$ ,  $\beta_\tau$  is a coefficient vector of interest, and  $\lambda$  is the logit function. We focus on the case where the data contain two reports, *i.e.* both  $i$  and  $j$  were (separately) asked to report  $\tau_{ij}$ . Let  $G_{ij}$  be the report that the giver  $i$  made on the true transfer  $\tau_{ij}$  and let  $R_{ij}$  be the report that the receiver  $j$  made on the same transfer  $\tau_{ij}$ . In principle,  $i$  and  $j$  should report the same thing, *i.e.*,  $G_{ij} = R_{ij}$ . This is not, however, what is typically observed. For instance, in the dataset that we use for illustration purposes, when respondent  $i$  reports  $G_{ij} = 1$  respondent  $j$  typically reports  $R_{ij} = 0$ .

In what follows we assume that the source of mis-reporting in data is under-reporting, for instance driven by poor recall. With under-reporting, if a flow is reported by either  $i$  or  $j$ , then it must have taken place. But a flow may also have taken place even if it was not reported by either  $i$  or  $j$ . We propose a maximum likelihood estimator that corrects for such mis-reporting pattern. Whether under-reporting is a reasonable assumption or not depends on the context. It seems to us the most reasonable for our application on transfers data in Tanzania.<sup>2</sup> Appendix C illustrates how the methodology can be amended to deal with the polar assumption of over-reporting, and confirms that under-reporting is most appropriate for the data at hand.

Dropping the  $ij$  subscripts to improve readability, let  $\tau$  denote the true binary flow

---

<sup>2</sup>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. 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 gift. 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. See also Akee and Kapur (2012) for evidence on reporting bias about transfers.



or transfer from  $i$  to  $j$ , *i.e.*,  $\tau = 1$  if  $i$  made a transfer to  $j$ .<sup>3</sup> 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  $\tau$ , only  $G$  and  $R$ . Under-reporting implies that  $G = 1$  only if  $\tau = 1$ , and that  $R = 1$  only if  $\tau = 1$ . However, it could be the case that  $G = 0$ ,  $R = 0$  and still  $\tau = 1$ . 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,<sup>4</sup> is reasonable if under-reporting results primarily from mistakes and omissions. With this assumption, we can rewrite the system as:

$$\Pr(G = 1, R = 0) = \Pr(\tau = 1) * \Pr(G = 1|\tau = 1) * \Pr(R = 0|\tau = 1) \quad (2)$$

$$\Pr(G = 0, R = 1) = \Pr(\tau = 1) * \Pr(G = 0|\tau = 1) * \Pr(R = 1|\tau = 1) \quad (3)$$

$$\Pr(G = 1, R = 1) = \Pr(\tau = 1) * \Pr(G = 1|\tau = 1) * \Pr(R = 1|\tau = 1) \quad (4)$$

$$\Pr(G = 0, R = 0) = 1 - \Pr(G = 1, R = 0) - \Pr(G = 0, R = 1) - \Pr(G = 1, R = 1) \quad (5)$$

---

<sup>3</sup>We prefer to model transfers as binary because in our data we observe major discrepancies between the amounts declared by giver and receiver (see Section 3). However, the method could be easily extended to a framework where transfers are continuous.

<sup>4</sup>Setting  $\tau_{ij} = \max\{G_{ij}, R_{ij}\}$  as it is common in the social network literature is equivalent to assuming perfect negative correlation between  $G|\tau$  and  $R|\tau$  - *i.e.*,  $i$  remembers when  $j$  does not and *vice versa*. This is an unreasonable assumption in most cases. Assuming perfect positive correlation between  $G|\tau$  and  $R|\tau$  rules out discordant answers, a feature that is trivially rejected in most datasets, including the one we use in our empirical illustration. With only two reports  $R$  and  $G$ , it is not possible to estimate a model that allows for arbitrary correlation between  $G|\tau$  and  $R|\tau$ . This leaves independence as the only realistic option. As we explain below, we can, however, correct for any systematic reporting bias in  $G$  and  $R$  that is correlated with observables.

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  $\lambda(\cdot)$  as follows:

$$\Pr(\tau = 1) = \lambda(\beta_\tau X_\tau) \tag{6}$$

$$\Pr(G = 1|\tau = 1) = \lambda_G(\beta_G X_G) \tag{7}$$

$$\Pr(R = 1|\tau = 1) = \lambda_R(\beta_R X_R) \tag{8}$$

Together with (2) to (5), equations (6) to (8) fully characterize the likelihood of observing the data. The main equation of interest is  $\Pr(\tau = 1) = \lambda(\beta_\tau X_\tau)$ : it is on this equation that we wish to test restrictions on the true parameter vector  $\beta_\tau$ . Conditioning on  $X_G$  and  $X_R$  in  $\Pr(G = 1|\tau = 1)$  and  $\Pr(R = 1|\tau = 1)$  allows for correlation in reporting probabilities between giving and receiving households.<sup>5</sup>

To illustrate how our correction for mis-reporting affects inference we will compare the estimated results from  $\Pr(\tau = 1)$  with two standard logit regressions which are commonly used in the network literature. In the first of them, the dependent variable equals one if at least one side has declared a transfer, which is equivalent to defining  $\tau_{ij}^{max} \equiv \max\{G_{ij}, R_{ij}\}$ . This assumes that when both reports agree they are true statements and 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 transfer, *i.e.*, it is  $\tau_{ij}^{min} \equiv \min\{G_{ij}, R_{ij}\}$ . This is equivalent to assuming that when both reports agree they are true statements and all discordances are due to over-reporting. In both cases, the possibilities that a transfer occurred but was not declared by anyone, or that a transfer declared by both parts involved in reality did not occur, are ruled out.

## 2.2 Standard errors

Dyadic observations are typically not independent. This does not invalidate the application of standard maximum likelihood techniques to estimate  $\beta_\tau, \beta_G$  and  $\beta_R$  in equations (6) to (8). But standard errors must be adjusted to correct for dyadic dependence across observations, otherwise inference will be inconsistent. Since we only

---

<sup>5</sup>For instance, if wealthy households are less likely to report receiving a transfer than poor households and wealth is correlated across giving and receiving households, this can be controlled for by including the wealth of the giver in  $X_G$  and the wealth of the receiver in  $X_R$ .

have data from a single population,<sup>6</sup> we apply the formula developed by Fafchamps and Gubert (2007) which corrects for arbitrary correlation across all  $\tau_{ij}$  and  $\tau_{ji}$  observations involving either  $i$  or  $j$ .

### 2.3 Simulation analysis

Whether or not mis-reporting affects inference depends on the hypothesis being tested, that is, on the regressors of interest in equation (6). To illustrate this point, we conduct an extensive simulation analysis to investigate how our estimator and the standard logit regressions behave when reporting propensities  $\lambda_G(\beta_G X_G)$  and  $\lambda_R(\beta_R X_R)$  vary systematically with the regressors of interest. Results below show that our estimator always delivers satisfactory coefficients, while the results from the standard logit estimates can be severely biased.

We posit a data generating process of the form

$$\Pr(\tau_{ij} = 1) = \lambda(\beta_{\tau 0} + \beta_{\tau 1} x_i + \beta_{\tau 2} x_j + \beta_{\tau 3} d_{ij} + \varepsilon_{\tau ij}) \quad (9)$$

where  $\tau_{ij}$  is the real transfer from  $i$  to  $j$ ,  $x_i$  and  $x_j$  are two uniformly distributed individual attributes (representing wealth),  $d_{ij}$  is a uniformly distributed relational attribute (representing distance), the error term  $\varepsilon_{\tau ij} \sim N(0, 1)$  and  $\lambda$  is the logit function. While  $\tau_{ij}$  stays unobserved, we generate the two individual binary reports  $G_{ij}$ ,  $R_{ij}$  under three different mis-reporting scenarios as follows:

- Under Scenario 1 we impose that mis-reporting is present but it is purely random, *i.e.*,  $\Pr(G_{ij} = 1) = \lambda(\beta_{G0} + \varepsilon_{Gij})$  and  $\Pr(R_{ij} = 1) = \lambda(\beta_{R0} + \varepsilon_{Rij})$  where  $\varepsilon_{Gij}, \varepsilon_{Rij} \sim N(0, 1)$ .
- Under Scenario 2 we generate mis-reporting on the basis of the individual attributes, *i.e.*,  $\Pr(G_{ij} = 1) = \lambda(\beta_{G0} + \beta_{G1} x_i + \varepsilon_{Gij})$  and  $\Pr(R_{ij} = 1) = \lambda(\beta_{R0} + \beta_{R2} x_j + \varepsilon_{Rij})$ . This corresponds to the case where wealthier respondents are more likely to report transfers given and received.
- Under Scenario 3 we generate mis-reporting on the basis of the relational attribute, *i.e.*,  $\Pr(G_{ij} = 1) = \lambda(\beta_{G0} + \beta_{G3} d_{ij} + \varepsilon_{Gij})$  and  $\Pr(R_{ij} = 1) = \lambda(\beta_{R0} + \beta_{R3} d_{ij} +$

---

<sup>6</sup>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).

$\varepsilon_{Rij}$ ). This corresponds to the case where transfers to (geographically or socially) proximate households are easier to recall.

Under all three scenarios we maintain the frequency of the declarations by givers and receivers to be 60% and 40% respectively conditional on  $\tau_{ij} = 1$ , which matches the relative proportions of our observational data. For each of these three scenarios we then draw 250 random networks and we compare the performance of our estimator with standard logit regressions. Simulation results are summarized in Table 1.

**Table 1. Simulation results**

	(1)	(2)	(3)	(4)	(5)
	true model	our estimator	our estimator	standard logit	standard logit
	$\tau_{ij}$	intercept only	with covariates	$\tau_{ij}^{max}$	$\tau_{ij}^{min}$
Scenario 1:					
$\beta_{\tau_1}$	1.73	1.75	1.76	1.48	1.13
$\beta_{\tau_2}$	1.73	1.75	1.75	1.48	1.14
$\beta_{\tau_3}$	-1.73	-1.74	-1.75	-1.45	-1.09
Scenario 2:					
$\beta_{\tau_1}$	1.73	2.3	1.72	1.92	1.83
$\beta_{\tau_2}$	1.74	2.12	1.72	1.77	2.21
$\beta_{\tau_3}$	-1.74	-1.83	-1.73	-1.51	-0.97
Scenario 3:					
$\beta_{\tau_1}$	1.73	1.72	1.76	1.48	1.18
$\beta_{\tau_2}$	1.73	1.73	1.76	1.48	1.19
$\beta_{\tau_3}$	-1.74	-1	-1.75	-0.8	0.52

Column (1) of Table 1 reports the average logit coefficients over the 250 replications when we estimate equation (9) using the actual transfer  $\tau_{ij}$  as dependent variable. Column (2) reports the average estimated coefficients from equation (6) of our estimator when in the regressor sets  $X_G$  and  $X_R$  for the reporting equations we only include the intercept term. Column (3) reports the average estimated coefficients from equation (6) of our estimator when in the regressor sets  $X_G$  and  $X_R$  for the reporting equations we also add own and relational attributes, *i.e.* we include  $x_i$  and  $d_{ij}$  in  $X_G$ , and we include  $x_j$  and  $d_{ij}$  in  $X_R$ . Column (4) reports average logit coefficients when we posit  $\tau_{ij}^{max} \equiv \max\{G_{ij}, R_{ij}\}$  and estimate equation (9) applying standard logit to  $\tau_{ij}^{max}$ .

Column (5) reports average logit coefficients if we instead let  $\tau_{ij}^{min} \equiv \min\{G_{ij}, R_{ij}\}$  and apply standard logit to  $\tau_{ij}^{min}$ .

Results show that our estimator outperforms the standard logit regressions of columns (4) and (5) in all cases. Under Scenario 1 our estimator does equally well whether or not we condition the reporting equations on the observables. When we do not correct for mis-reporting, the magnitude of the estimated coefficients is biased downwards – more severely in column (5) than in column (4). Under Scenarios 2 and 3 where reporting propensities depend on observables, our estimator delivers consistent results only if we include the controls in the reporting equations. In particular, our estimator with covariates (column 3) always delivers satisfactory coefficients. This is not the case for our estimator with intercept only (column 2) or for the standard logit regressions (column 4 and 5). The bias in the estimated coefficient is particularly severe for the variable which affects reporting: in Scenario 2 both  $\beta_1$  and  $\beta_2$  are upward biased in all columns except column (3), and similarly in Scenario 3  $\beta_3$  is always upward biased with the exception of column (3). However, our estimator seems to perform better than the standard logit regressions even when we only include the intercept in  $X_G$  and  $X_R$ , as in column (2). Indeed, for columns (4) and (5) the coefficients of regressors that do not enter the reporting equations (*i.e.*,  $\beta_3$  for Scenario 2 and  $\beta_1, \beta_2$  for Scenario 3) are severely bias, which is not the case in column (2).

Overall, the simulation exercise suggests that, if the self-reporting of transfer data has the very general properties sketched above, using standard logit regressions to estimate equation (1) yields incorrect inference. Results also indicate that identification does not require that the regressor sets  $X_G$  and  $X_R$  contain a variable absent from  $X_\tau$ .

## 3 Informal Transfers in Tanzania

### 3.1 Nyakatoke household survey

We illustrate our methodology using a unique census dataset on transfers between all the households in an African village, Nyakatoke. The village is located in the Buboka Rural District of Tanzania, at the west of Lake Victoria. The data have been the object of numerous articles (*e.g.* De Weerd and Dercon 2006; De Weerd and Fafchamps 2011; Vandenbossche and Demyunck 2013; Comola 2012; Comola and Fafchamps 2014).

The community is composed by 600 inhabitants, 307 of which are adults.<sup>7</sup> A total

---

<sup>7</sup>Individuals aged 16 and above are considered adult.

of 119 households were interviewed in five rounds at regular intervals from February to December 2000. In the first survey round (February 2000), each adult was asked whom he would ask and/or provide help to in case of need. We use the answers to this question, aggregated at the household level, as proxies for desire to link (discussing this assumption is the scope of Appendix B). During each of the subsequent interview rounds, each adult was asked whether they had received or given transfers (loans or gifts). If they said yes, information was collected on the name of the partner and the value of what was given or received, whether in cash or kind.<sup>8</sup> This provides us with a detailed picture of all transfers occurring within the village over one year. In Nyakatoke reported gifts are more frequent than loans but smaller in magnitude (De Weerd and Dercon 2006) and have been shown to serve an insurance purpose against health shocks (De Weerd and Fafchamps 2011).<sup>9</sup> This is in line with the literature on informal risk sharing which has shown how informal transfers can be a way of smoothing consumption against shocks while satisfying self-enforcement constraints (Udry 1994; Kocherlakota 1996; Foster and Rosenzweig 2001; Ligon Thomas and Worrall 2001).

### 3.2 Transfer data

In order to map the transfers between Nyakatoke households we aggregate the individual-level information on transfers at the household level and across rounds. We aggregate at the household level to reduce discrepancies that could arise if  $i$  mentioned giving to member  $a$  of household  $j$  but member  $b$  of household  $j$  is the one who mentions receiving a gift from  $i$ .<sup>10</sup> We also aggregate across rounds to reduce discrepancies that could arise if household  $i$  declares a transfer in round  $t$  while household  $j$  declares the same transfer in round  $t + 1$ .

Our unit of observation is the dyad: in Nyakatoke there are 119 households, which gives  $119 * 118 = 14042$  dyads. For each household dyad  $ij$  we thus have four measurements of the gifts which took place among them: gifts  $G_{ij}$  that  $i$  stated giving to  $j$ ; gifts  $R_{ij}$  that  $j$  stated receiving from  $i$ ; gifts  $G_{ji}$  that  $j$  declared giving to  $i$ ; and gifts  $R_{ji}$  that  $i$  stated receiving from  $j$ . Similar data is available for loans. These four measurements correspond to two actual gross flows: the flow from  $i$  to  $j$ , denoted  $\tau_{ij}$ ,

---

<sup>8</sup>Loan repayment and gifts in labor are not included.

<sup>9</sup>This is consistent with findings reported by Fafchamps and Lund (2003) for the Philippines.

<sup>10</sup>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.

and the flow from  $j$  to  $i$ , denoted  $\tau_{ji}$ . Since we focus on gross flows, the two are not the same. Hence  $\{\tau_{ij}\}$  defines a directed graph.

There are major discrepancies between  $G_{ij}$  and  $R_{ij}$ . In fact,  $G_{ij} \neq R_{ij}$  in nearly all cases, especially for loans. There are 1420 dyads (*i.e.*, 10.1% of the household dyads) for which either  $G_{ij}$  or  $R_{ij}$  is not zero for gifts. In 596 cases the report comes from the giver only (4.2% of the dyads), in 424 cases from the receiver only (3% of the dyads), and in 400 from both (2.8% of the dyads). For inter-household loans, there are 545 dyads (*i.e.*, 3.9% of the dyads) for which either  $i$  or  $j$  reports a loan from  $i$  to  $j$ . In 308 of these cases the report comes from the lender only (2.2% of the dyads), in 195 cases from the borrower only (1.4% of the dyads), and in 42 cases from both (0.3% of the dyads). Out of the dyads in which both  $i$  and  $j$  report a gift from  $i$  to  $j$ , only 22 report the exact same amount. For loans, the corresponding number is 5. When the amounts declared differ, they differ by a large margin: for both loans and gifts the highest of the two declared amounts is on average double the smallest one. Amounts reported by both sides are on average much larger than amounts reported by one side only.<sup>11</sup> Moreover, these discrepancies do not seem to be due to the fact that respondents mix up loans and gifts.<sup>12</sup> The frequency distribution of loan and gift amounts is given in Table A.1, Appendix A.

In summary, there are massive discrepancies between the responses given by  $i$  and  $j$  about the same gifts and loans. These discrepancies are mostly due to the fact that in the large majority of cases 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 sense 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 source of bias.

---

<sup>11</sup>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.

<sup>12</sup>If we restrict the sample to the dyads for which at least one loan or gift was reported the within-dyad correlation between the difference in reported loans and the difference in reported gifts is small and not statistically significant.

### 3.3 Testing self-interest in link formation

For the scope of this paper we define a transaction as ‘mutually beneficial’ if it is in the self-interest of both parts involved. In market exchange, it is customary to assume that transactions are mutually beneficial because there are mutual gains from trade. There are exceptions, however. For instance, one side may be forced to trade because refusing to do so would contravene a legal or social obligation not to discriminate. In this example, trade is not mutually beneficial - given the choice, one of the two sides 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 may *ex ante* prefer for the transfer not to take place, but cannot refuse to be part of it once it is initiated by the other party. In our definition, these transfers aren’t mutually beneficial.<sup>13</sup> Note that this characterizes the link formation process, not the direction of flows between the nodes.<sup>14</sup>

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 desire to link is not directly elicited, proxies often are available for the objective utility or material gain that individuals derive from dif-

---

<sup>13</sup>Our definition of mutually beneficial transfers largely overlaps with the definition of bilateral links common to the network literature (*e.g.*, Goyal 2007, Jackson 2009). Here we prefer to speak of mutually beneficial transfers to underline that, in most examples we discuss, the transfer takes place with the formal consent of both parties (which technically makes it bilateral) even though it is not in the self-interest of one of the two sides.

<sup>14</sup>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 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 mutually beneficial or not depending on the context.



ferent matches. If (direct or indirect) information on desire to link is available to the researcher, it is of economic interest to explore whether the link observed in data are grounded by mutual self-interest.

Assume we have dichotomous measures  $d_{ij} = \{0, 1\}$  and  $d_{ji} = \{0, 1\}$  of households' desire to exchange transfers with each other. If transfers are not necessarily grounded by mutual self-interest, 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 mutual self-interest is required, 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  $\tau_{ij} > 0$ . In order to investigate whether observed transfers are mutually beneficial, we estimate equation (6) as

$$\Pr(\tau_{ij} > 0) = \lambda(\alpha d_{ij} + \beta d_{ji} + \gamma d_{ij}d_{ji} + \theta Z_{ij\tau}) \quad (10)$$

where  $X_{ij\tau} \equiv [d_{ij}, d_{ji}, d_{ij}d_{ji}, Z_{ij\tau}]$ . If transfer aren't mutually beneficial, the likelihood of  $\tau_{ij} > 0$  is the same whether  $\{d_{ij}, d_{ji}\} = \{1, 0\}$ ,  $\{0, 1\}$ , or  $\{1, 1\}$ : it follows that  $\alpha = \beta = \alpha + \beta + \gamma > 0$  which implies that  $\gamma = -\beta = -\alpha$ . If transfers are mutually beneficial,  $\tau_{ij} > 0$  arise only if  $\{d_{ij}, d_{ji}\} = \{1, 1\}$ : it follows that  $\alpha = \beta = 0$  and  $\gamma > 0$ . This test is at the core of our empirical illustration. Indeed, for the purpose of this test the mis-reporting correction is of great value because we suspect that desire to link may affect not only equation (6), but also the reporting equations. If this is the case, only by correcting for mis-reporting we can draw correct inference about whether observed links are mutually beneficial.

However, in order to perform the test we need proxies for desire to link, which is not directly elicited in our data. 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 on for help and/or that can rely on you for help in cash, kind or labor?"*<sup>15</sup> Answers to this question, aggregated at the household level, are used as proxies for  $d_{ij}$  and  $d_{ji}$  in our illustration of Section 4. This requires some explanation given that the question in principle asks about links – not desire to link. For a careful discussion of this assumption we remand to Appendix B.

---

<sup>15</sup>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.

### 3.4 Variables definition

The regressors used in our analysis are illustrative of the type of variables typically included in an analysis of this kind. The main regression of interest is  $\Pr(\tau = 1) = \lambda(\beta_\tau X_\tau)$ . The regressors entering  $X_\tau$  are control variables expected to influence the actual flows of funds between households. Since  $\tau_{ij}$  is directional, regressors for observation  $ij$  can differ from regressors for observation  $ji$ .<sup>16</sup> The regressors of interest for our testing strategy are  $d_{ij}$ ,  $d_{ji}$  and  $d_{ij}d_{ji}$  as defined in Subsection 3.3. We also add the wealth of  $i$  and  $j$ , since we expect flows of funds between households to depend on it.<sup>17</sup> From the work of Fafchamps and Lund (2003), De Weerd and Dercon (2006) and De Weerd and Fafchamps (2011), we also know that informal arrangements are more frequent among households that are socially and geographically proximate. Therefore we include the distance between the two houses, and relational dummies for whether  $i$  and  $j$  share the same religion and are blood related.<sup>18</sup> Finally, larger households have more individuals involved in giving and receiving transfers. We therefore control for the number of adult members for  $i$  and  $j$ .

Next we discuss the variables that enter the reporting equations of giver and receiver  $\Pr(G = 1|\tau = 1) = \lambda_G(\beta_G X_G)$  and  $\Pr(R = 1|\tau = 1) = \lambda_R(\beta_R X_R)$ . We include own desire to link ( $d_{ij}$  in  $X_G$  and  $d_{ji}$  in  $X_R$ ) since respondents may be more likely to remember transfers to/from individuals whose name they listed in response to first-round interviews. Following a similar argument we include wealth (wealth of  $i$  in  $X_G$  and wealth of  $j$  in  $X_G$ ) as regressor given that wealthy people may be more likely to forget a transfer (Akee and Kapur 2012). 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 are *a priori* expected to affect mis-reporting but not transfers themselves.<sup>19</sup> For  $X_G$ , we use *declared friends<sub>i</sub>*, defined as the total 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 underlying this choice is that households that intend to

---

<sup>16</sup>This stands in contrast with undirected network data where  $\tau_{ij} \equiv \tau_{ji}$  and regressors by construction have to be identical such that  $X_\tau^{ij} = X_\tau^{ji}$ .

<sup>17</sup>Wealth is computed as the total value of land and livestock assets in Tanzanian shilling (1 unit = 100000 *tzs*).

<sup>18</sup>Out of 119 households, 24 are Muslim (20%), 46 are Protestant (39%) and 49 are Catholic (41%). We consider households  $i$  and  $j$  blood-related if an adult member of  $i$  is the parent/sibling/child of an adult member of  $j$ .

<sup>19</sup>Simulation analysis reported earlier indicates that the maximum likelihood estimates are reliable even without identifying instruments, so including these variables is not necessary for identification.

seek help from (or provide help to) many other households are probably more sensitive to the issue of inter-household transfers, and therefore recall transfers better. For  $X_R$  we include separately 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 head – and therefore may be reluctant to report them to enumerators. This is in line with the idea that the under-reporting of transfers received is to avoid detection by other household members – a point already made by Anderson and Baland (2002).

In Table 2 we present descriptive statistics for all variables used in the analysis. The upper section of the table reports different versions of the dependent variables. The first two rows focus on the gifts from  $i$  to  $j$ , as reported by  $i$  and  $j$  respectively. In the next two rows we report  $\tau_{ij}^{max} \equiv \max\{G_{ij}, R_{ij}\}$  and  $\tau_{ij}^{min} \equiv \min\{G_{ij}, R_{ij}\}$  that are used as dependent variables in the standard logit regressions. They demonstrate the extent of the divergence between the information given by households  $i$  and  $j$  on the same  $\tau_{ij}$ . In the next four rows we report the same information for inter-household loans. We see that lenders are more likely to report a loan than borrowers, and that there are important discrepancies between loans reported by the lender and borrower. The rest of Table 2 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. There is considerable variation in wealth levels across Nyakatoke households. There is also significant diversity in religion: only 35% of household head pairs share the same religion. 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.<sup>20</sup> The average number of friends declared in the first-round question is 5.2, which is higher than the number of transfers declared. The average number of adult male and female dependents is 0.44 and 1.1 respectively.

---

<sup>20</sup>For 3 households the distance is missing, so we have imputed the sample average.

**Table 2. Descriptive statistics (N=14042)**

variable	dummy	mean	min	max	sd
$\tau_{ij}^i$ (gifts)	yes	0.071			
$\tau_{ij}^j$ (gifts)	yes	0.059			
$\tau_{ij}^{max}$ (gifts)	yes	0.101			
$\tau_{ij}^{min}$ (gifts)	yes	0.028			
$\tau_{ij}^i$ (loans)	yes	0.025			
$\tau_{ij}^j$ (loans)	yes	0.017			
$\tau_{ij}^{max}$ (loans)	yes	0.039			
$\tau_{ij}^{min}$ (loans)	yes	0.003			
$d_{ij}$ and $d_{ji}$	yes	0.045			
$d_{ij}d_{ji}$	yes	0.020			
$wealth_i$ and $wealth_j$	no	4.546	0	27.970	4.815
<i>same religion</i>	yes	0.354			
<i>related</i>	yes	0.016			
<i>distance</i>	no	0.522	0.014	1.738	0.303
$hh\ members_i$ and $hh\ members_j$	no	2.555	1	9	1.314
<i>declared friends<sub>i</sub></i>	no	5.294	0	19	3.063
<i>female dependents<sub>j</sub></i>	no	1.101	0	6	0.864
<i>male dependents<sub>j</sub></i>	no	0.437	0	3	0.729

## 4 Estimation results

### 4.1 Main results

Table 3 presents the estimates for gifts. Columns (1) and (2) report the results from standard logit regressions where the dependent variable is  $\tau_{ij}^{max} \equiv \max\{G_{ij}, R_{ij}\}$  and  $\tau_{ij}^{min} \equiv \min\{G_{ij}, R_{ij}\}$  respectively. Columns (3) to (5) report coefficients obtained from estimating the likelihood function defined by equations (2) to (8). Column (3) corresponds to the equation of interest (1), while columns (4) and (5) correspond to the reporting equations of the giver and receiver respectively. Marginal effects for Table 3 are reported in Table A.2 in Appendix A.

Looking at columns (1) to (3) we notice that desire to link of both giver and receiver is significantly positive, suggesting that a gift is more likely to occur if the

giver and/or the receiver has listed the partner in response to the first round question. 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. Since inference about mutually beneficial 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 3 strongly reject a link formation model grounded in mutual benefit: both  $\alpha$  and  $\beta$  are strongly significant and of the same order of magnitude, while  $\gamma$  is negative in all three columns – significantly so in column (1) and (2).<sup>21</sup> However, we do not observe  $\gamma = -\beta$  and  $\gamma = -\alpha$  either.<sup>22</sup> 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 more likely to help each other than if only one lists the other. This suggests that some dimension of mutual self-interest is present, even though the results reject the pure model of mutually beneficial link formation. The other covariates have the expected sign across all three specifications. Wealthier households are more likely to give in columns (1) and (3), and more likely to receive in column (1) only. People are more likely to give to relatives, neighbors, and members of the same religion. Household members are always positive in sign, and occasionally significant.

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 wish 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  $j$ . *declared friends<sub>i</sub>* is significantly positive as expected, while own wealth is significantly 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<sub>j</sub>* is negatively significant. The numbers of female and male dependents have the anticipated negative sign, but they are not significant.

Marginal effects, reported in Table A.2 in Appendix A, give a full sense of the estimation bias: when we do not account for mis-reporting  $d_{ij}$  and  $d_{ji}$  appear biased downwards in the transfer equation, which is consistent with the fact that desire to link also affects reporting probabilities.

In Table 4 we repeat the same analysis for loans. The corresponding marginal ef-

---

<sup>21</sup>A Wald test cannot reject the hypothesis that  $\alpha = \beta$  in column (3), with a  $p$ -value of 0.365.

<sup>22</sup>For our preferred model of Column 3 a Wald test rejects the joint hypotheses  $\gamma + \alpha = 0$  and  $\gamma + \beta = 0$  with  $p$ -value=0.002.

facts are reported in Table A.3, Appendix A. Coefficient estimates reported in column (3) approximately satisfy  $\alpha = \beta = -\gamma$ , a finding that would be consistent with the mutually beneficial link model, but individual coefficients are only statistically significant in column (1) and (2). This may be because the proportion of non-zero observations is very small for loans, making maximum likelihood 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 Weerd 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 *declared friends<sub>i</sub>* is significantly positive, and in the  $\Pr(R = 1|\tau = 1)$  regression only the desire to link is significant.

**Table 3. Results for gifts**

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

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

Table 4. Results for loans

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

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



## 4.2 Robustness analysis

As a robustness check we have re-estimated the model with different sets of regressors. Convergence is generally smooth for a reasonably sized set of regressors, and estimated coefficients are similar across specifications. Including significant regressors in the mis-reporting equations increases the difference between standard logit results in columns (1) and (2) and the maximum likelihood results in column (3). A few regressors in the reporting equations are sufficient to get stable estimates for  $\Pr(\tau = 1)$  as long as we include the variables that are likely to impact both the propensity to declare the transfer and the likelihood of a transfer itself.<sup>23</sup> However, no improvement in the estimated coefficients for  $\Pr(\tau = 1)$  is obtained if the  $\Pr(G = 1|\tau = 1)$  and  $\Pr(R = 1|\tau = 1)$  only include an intercept. These findings are consistent with the discussion and simulation results confirming that our estimator represents an improvement over logit only if we include relevant variables in the mis-reporting equations. It has to be mentioned that 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.

We also investigate whether our findings are affected by the possibility that what household  $i$  reported as a gift was reported as a loan by  $j$ . To investigate whether such mis-classification affects our results, we re-estimate the model using combined gifts and loans as the dependent variable. Results are not shown here to save space, but all coefficients of interest are similar to those reported in Table 3 for gifts.

## 5 Estimates of under-reporting

From the raw figures reported in Subsection 3.2 is it already possible to compute benchmark estimates of the extent of under-reporting, without any estimation and before introducing covariates. Assuming independence in reporting probability between  $i$  and  $j$ , we can calculate the three unconditional probabilities  $\Pr(\tau = 1)$ ,  $\Pr(G = 1|\tau = 1)$ , and  $\Pr(R = 0|\tau = 1)$  from the following three equations:

---

<sup>23</sup>For instance, if we omit  $d_{ij}$  and  $d_{ji}$  from the mis-reporting equations, the results are dramatically different: the coefficient of  $d_{ij}d_{ji}$  in the  $\Pr(\tau = 1)$  equation becomes large and positive, and has a large  $t$ -value.

$$\Pr(G = 1, R = 0) = \Pr(\tau = 1) * \Pr(G = 1|\tau = 1) * \Pr(R = 0|\tau = 1) = 0.042 \quad (11)$$

$$\Pr(G = 0, R = 1) = \Pr(\tau = 1) * \Pr(G = 0|\tau = 1) * \Pr(R = 1|\tau = 1) = 0.030 \quad (12)$$

$$\Pr(G = 1, R = 1) = \Pr(\tau = 1) * \Pr(G = 1|\tau = 1) * \Pr(R = 1|\tau = 1) = 0.028 \quad (13)$$

Straightforward algebra yields the solutions reported in Table 5 below:

**Table 5. Benchmark estimates of under-reporting**

	gifts	loans
in data: declared by $i$	0.071	0.025
in data: declared by $j$	0.059	0.017
in data: declared by $i$ or $j$ ( $\tau_{ij}^{max}$ )	0.101	0.039
in data: declared by $i$ and $j$ ( $\tau_{ij}^{min}$ )	0.028	0.003
Pr( $\tau_{ij} = 1$ )	0.158	0.120
Pr( $G = 1 \tau = 1$ )	0.444	0.176
Pr( $R = 1 \tau = 1$ )	0.400	0.142

The simple calculation above suggest that there is considerable under-reporting of gifts and that  $\tau_{ij}^{max} = 10.1\%$  underestimates the frequency of gifts by around 50%. The table also suggests massive under-reporting of loans and indicates that  $\tau_{ij}^{max} = 3.9\%$  only captures about a third of the loans suspected to have been made.

We can obtain similar estimates from the maximum likelihood model formed by equations (2) to (8). The only difference is that these estimates are conditional on covariates, a feature that allows for correlation in reporting propensities based on observables. The result of these calculations is reported in Table 6.

**Table 6. Estimates of under-reporting with covariates**

	gifts	loans
average fitted Pr( $\tau_{ij} = 1$ )	0.157	0.194
average fitted Pr( $G = 1 \tau = 1$ )	0.374	0.114
average fitted Pr( $R = 1 \tau = 1$ )	0.311	0.073

The average fitted propensity to give gifts from Table 6 is 15.7%, nearly identical to the figure of obtained without conditioning on covariates. For loans, the average fitted  $\Pr(\tau_{ij} = 1)$  of 19.4% is larger than our earlier estimate of 12%. These results suggest that 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 may lead to serious underestimation of the extent of gift giving and, especially, of lending and borrowing between villagers.

Table 6 also reports the average fitted propensities to report giving and receiving. The average propensity to report a gift is 37.4% for the giver and 31.1% for the receiver, smaller than the figures of Table 5. Estimated reporting probabilities are much lower for loans. Lenders are estimated to report only 11.4% of loans – compared to 17.6% of Table 5. Borrowers are estimated to report as little as 7.3% of loans, versus 14.2% in Table 5. If anything, estimated propensities to report gifts and loans fall when we allow them to depend on household observables.

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 general reliability of transfers of money, goods, or favors reported in household surveys. This matters for our understanding of the importance of favor exchange. In particular, 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 closer to 10% – a five-fold increase.

## 6 Conclusions

Self-reported transfer data are typically discordant:  $i$  may report a transfer to  $j$  while  $j$  reports no such transfer from  $i$ . In this paper we propose a maximum likelihood estimator to deal with mis-reporting of this kind. Using simulations, we show that the consequences of neglecting mis-reporting may be severe when determinants of transfers are correlated with the propensity to report a transfer given or received. Our estimator corrects for this bias by conditioning reporting on such determinants.

We illustrate the methodology using dyadic data on inter-household transfers from the village of Nyakatoke in Tanzania, where we observe substantial discrepancies between gifts and loans reported by givers and receivers. In particular, we combine data about flows and a proxy of desire to link to investigate whether observed transfers are mutually beneficial, *i.e.* in the self-interest of both parts involved. We find reasonably convincing evidence that loans and gifts do not require mutual self-interest to take place (even though flows are more likely to occur if both households wish to link). If this interpretation is correct, the evidence implies that if one household wishes to enter in a reciprocal relationship with another household, it can 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 De Weerd 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. But if households can wait after shocks are realized before deciding who to ask for help, they need not worry about covariate risk *ex ante*.

Our results suggest that both loans and gifts are heavily under-reported, and that loans are even more under-reported than gifts. This latter result is somewhat surprising: it is easy to see why borrowers would fail to report the loans they have received, but why would also lenders do so? Much of the theoretical discourse about risk sharing has emphasized repeated games and reputational 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 works of Goldberg (2010) and Kinnan (2012). 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 grounded in mutual self-interest, it would be possible to refuse to assist others and secrecy would not be necessary.

## References

- [1] Akee, Randall and Devesh Kapur (2012), ‘Remittances and Rashomon’, Center for Global Development, Working Paper 285, January
- [2] Anderson, Siwan and Jean-Marie Baland (2002), ‘The Economics of Roscas and Intra-household Resource Allocation’, *Quarterly Journal of Economics*, 117(3): 963-95
- [3] Arcand, Jean-Louis and Marcel Fafchamps (2012), ‘Matching in Community-Based Organizations’, *Journal of Development Economics*, forthcoming
- [4] 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
- [5] Bala, Venkatesh and Sanjeev Goyal (2000), “A Non-Cooperative Model of Network Formation,” *Econometrica*, 68(5): 1181-1229, September
- [6] Ball, Brian and M.E.J. Newman (2013), ‘Friendship networks and social status’, *Network Science*, 1(01): 16-30.
- [7] Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo and Matthew O. Jackson (2013), ‘The Diffusion of Microfinance’, *Science*, Vol. 341.
- [8] Belot, Michele and Marco Francesconi (2006), ‘Can Anyone Be The One? Evidence on Mate Selection from Speed Dating’, *CEPR Discussion Paper No. 5926*, London
- [9] 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
- [10] Coate, Stephen and Martin Ravallion (1993), ‘Reciprocity Without Commitment: Characterization and Performance of Informal Insurance Arrangements’, *Journal of Development Economics*, 40: 1-24
- [11] Comola, Margherita (2012), ‘Estimating Local Network Externalities’, SSRN Working Paper n. 946093

- [12] Comola, Margherita and Marcel Fafchamps (2014), ‘Testing Unilateral and Bilateral Link Formation’, *The Economic Journal*, 124: 954-976.
- [13] De Weerd, Joachim and Stefan Dercon (2006), ‘Risk-Sharing Networks and Insurance Against Illness’, *Journal of Development Economics*, 81(2): 337-56
- [14] De Weerd, Joachim and Marcel Fafchamps (2011), ‘Social Identity and The Formation of Health Insurance Networks’, *Journal of Development Studies*, 47(8): 1152-1177.
- [15] 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
- [16] Erdil, Aytak and Haluk Ergin (2007), ‘What’s the Matter with Tie-breaking? Improving Efficiency in School Choice’, Department of Economics, Oxford University
- [17] Fafchamps, Marcel and Susan Lund (2003), ‘Risk Sharing Networks in Rural Philippines’, *Journal of Development Economics*, 71: 261-87
- [18] Fafchamps, Marcel and Flore Gubert (2007), ‘The Formation of Risk Sharing Networks’, *Journal of Development Economics*, 83(2): 326-50
- [19] 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
- [20] Fox, Jeremy (2011), “Estimating Matching Games with Transfers,” Department of Economics, University of Michigan (mimeograph)
- [21] Glaeser, Edward, Bruce Sacerdote, and Jose Scheinkman (1996), “Crime and Social Interactions”, *Quarterly Journal of Economics*, 111: 507-48, 1996
- [22] 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)
- [23] Goyal, Sanjeev (2007), *Connections: An Introduction to the Economics of Networks*, Princeton University Press, Princeton and Oxford

- [24] Hitsch, Gunter J., Ali Hortacsu, Dan Ariely (2011), ‘Matching and Sorting in Online Dating’, *American Economic Review*, forthcoming
- [25] Hong, Han and Elie Tamer (2003), “A simple estimator for nonlinear error in variable models,” *Journal of Econometrics*, 117(1): 1-19
- [26] Jackson, Matthew O. (2009), *Social and Economic Networks*, Princeton University Press, Princeton
- [27] Jackson, Matthew O. and Asher Wolinsky (1996), “A Strategic Model of Social and Economic Networks,” *Journal of Economic Theory*, 71(1): 44-74
- [28] Kinnan, Cynthia (2012), ‘Distinguishing barriers to insurance in Thai villages’, Department of Economics, Northwestern University, July (mimeograph)
- [29] Kocherlakota, Narayana R. (1996), ‘Implications of Efficient Risk Sharing Without Commitment’, *Rev. Econ. Stud.*, 63(4): 595-609
- [30] 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
- [31] Ligon, Ethan, Jonathan P. Thomas, and Tim Worrall (2001), ‘Informal Insurance Arrangements in Village Economies’, *Review of Economic Studies*, 69(1): 209-44
- [32] 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
- [33] 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
- [34] Platteau, Jean-Philippe (2000), *Institutions, Social Norms, and Economic Development*, Harwood Academic Publishers, Amsterdam
- [35] Rosenzweig, Mark R. (1988), “Risk, Implicit Contracts and the Family in Rural Areas of Low-Income Countries,” *Economic Journal*, 98: 1148-1170, December

- [36] Roth, Alvin and Marilda Sotomayor (1990), *Two-Sided Matching*, Cambridge University Press, Cambridge
- [37] Schennach, Susanne (2004), “Estimation of Nonlinear Models with Measurement Error,” *Econometrica*, 72 (1): 33-75
- [38] Snijders, Tom A.B., Johan Koskinen, and Michael Schweinberger (2010), “Maximum Likelihood Estimation for Social Network Dynamics”, *Annals of Applied Statistics*, 4 (2): 567-588
- [39] Steglich, Christian E.G., Tom A.B. Snijders, and Michael Pearson (2010), ‘Dynamic Networks and Behavior: Separating Selection from Influence’, *Sociological Methodology*, 40 (1): 329-393
- [40] 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
- [41] Vandenbossche, Joost and Thomas Demuynck (2013), ‘Network Formation with Heterogeneous Agents and Absolute Friction’, *Computational Economics*, 42 (1): 23-45.
- [42] 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](http://www.nber.org/WNE/lect_6_controlfuncs.pdf)



# Appendix A

**Table A.1 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. Cut-off values computed on nonzero observations only. Values expressed in *tzs*.

**Table A.2 Marginal effects for Table 3 (gifts)**

	(1)	(2)	(3)	(4)	(5)
	$\tau_{ij}^{max}$	$\tau_{ij}^{min}$	$\Pr(\tau = 1)$	$\Pr(G = 1 \tau = 1)$	$\Pr(R = 1 \tau = 1)$
$d_{ij}^*$	0.3431	0.0766	0.6052	0.5388	
$d_{ji}^*$	0.4180	0.1485	0.6939		0.6625
$d_{ij}d_{ji}^*$	-0.0280	-0.0053	-0.0063		
$wealth_i$	0.0032	0.0002	0.0032	-0.0123	
$wealth_j$	0.0036	0.0000	0.0041		-0.0125
$same\ religion^*$	0.0243	0.0030	0.0256	0.0086	0.0033
$related^*$	0.1953	0.0123	0.3985	0.1634	0.2092
$distance$	-0.0931	-0.0144	-0.0663	-0.2036	-0.1496
$hh\ members_i$	0.0046	0.0006	0.0043		
$hh\ members_j$	0.0118	0.0014	0.0103		
$declared\ friends_i$				0.0090	
$female\ dependents_j$					-0.0420
$male\ dependents_j$					-0.0535

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

**Table A.3 Marginal effects for Table 4 (loans)**

	(1)	(2)	(3)	(4)	(5)
	$\tau_{ij}^{max}$	$\tau_{ij}^{min}$	$\Pr(\tau = 1)$	$\Pr(G = 1 \tau = 1)$	$\Pr(R = 1 \tau = 1)$
$d_{ij}^*$	0.1301	0.0040	0.7681	0.0425	
$d_{ji}^*$	0.1367	0.0259	0.7395		0.0707
$d_{ij}d_{ji}^*$	-0.0223	-0.0007	-0.0528		
$wealth_i$	0.0005	0.0000	0.0024	-0.0018	
$wealth_j$	0.0004	0.0000	0.0012		-0.0002
<i>same religion*</i>	0.0048	-0.0003	0.0355	-0.0025	-0.0006
<i>related*</i>	0.0040	-0.0002	0.1142	-0.0032	0.0022
<i>distance</i>	-0.0323	-0.0013	-0.0701	-0.0036	0.0003
$hh\ members_i$	0.0013	0.0001	0.0005		
$hh\ members_j$	0.0008	0.0000	0.0076		
<i>declared friends<sub>i</sub></i>				0.0049	
<i>female dependents<sub>j</sub></i>					-0.0006
<i>male dependents<sub>j</sub></i>					-0.0031

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

# Appendix B

To proxy for desire to link  $d_{ij}$  and  $d_{ji}$ , which is not directly elicited in our data, in Section 4 we use answers to the first-round survey question: “*Can you give a list of people from inside or outside of Nyakatoke, who you can personally rely on for help and/or that can rely on you for help in cash, kind or labor?*” The question was first piloted in the Philippines (Fafchamps and Lund 2003) and subsequently adopted in the Tanzania survey. This phrasing was used because respondents understand it and are willing to answer it.<sup>24</sup> However, given that the question in principle asks about links rather than desire to link, our choice requires some discussion that we provide in what follows.

## Desire to link vs. actual links

We first note that if responses perfectly captured actual links, then we would observe  $d_{ij} = d_{ji}$  for all  $i$  and  $j$ , which 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 (2014) 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 conclude that the data are best interpreted as desire to link.<sup>25</sup> In what follows 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 mutually beneficial, but  $i$  and  $j$  sometimes forget to report existing links. In this case,  $d_{ij} = 1$  only when  $i$  knows  $j$  wants to link with him. Similarly,  $d_{ji} = 1$  when  $j$  knows

---

<sup>24</sup>Other questions were tried both in the Philippines and in Tanzania, 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.

<sup>25</sup>The intuition behind their 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.

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 not grounded in mutual interest:  $i$  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  $j$ . As predictors of attractiveness  $z_j$  we use  $wealth_j$  and  $popularity_j$ .  $popularity_j$  is defined as the number of times  $j$  is listed by households other than  $i$  in response to the first-round question, and proxies for unobserved characteristics of household  $j$  such as sociability and charitable disposition that makes other households wish to link with  $j$ . Regression results are reported in Table B.1 below. We find  $b > 0$  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.

**Table B.1 Testing 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.

It is also important to realize that, if link formation is not necessarily mutually beneficial, 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 easily create such a link. So, in this sense our evidence is internally consistent.

## Mis-reported links

We now explore the possibility that first-round answers reflect mis-reported links rather than desire to link, as Comola and Fafchamps (2014) have concluded. The difference is important because it would bias our test results against mutually beneficial 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  $\tau_{ij}$  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 desire to link and used to estimate (1), results would induce us to conclude against mutually beneficial transfers.

We attempt to reassure the reader that this is not the case by replacing  $d_{ij}$  ( $d_{ji}$ ) with a variable proxying for  $i$ 's benefit from linking to  $j$  ( $i$ ). This variable is *popularity* as defined above. Results for gifts presented in Table B.2 below reconfirm the main results of Table 3: in the transfer equation of column (3)  $popularity_j$  and  $popularity_j$  are positive, significant and of similar order of magnitude, while their interaction is non-significant.

**Table B.2 Results with proxies for desire to link**

	(1)	(2)	(3)	(4)	(5)
	$\tau_{ij}^{max}$	$\tau_{ij}^{min}$	$\Pr(\tau = 1)$	$\Pr(G = 1 \tau = 1)$	$\Pr(R = 1 \tau = 1)$
<i>popularity<sub>i</sub></i>	0.077*** (0.010)	0.098*** (0.013)	0.051*** (0.016)		0.087*** (0.018)
<i>popularity<sub>j</sub></i>	0.097*** (0.018)	0.064*** (0.018)	0.070** (0.028)	0.087* (0.049)	
<i>popularity<sub>i</sub> * popularity<sub>j</sub></i>	0.000 (0.001)	0.002* (0.001)	0.002 (0.001)		
<i>wealth<sub>i</sub></i>	0.046*** (0.009)	0.024** (0.012)	0.055*** (0.012)	-0.030 (0.022)	
<i>wealth<sub>j</sub></i>	0.049** (0.021)	0.011 (0.017)	0.067 (0.045)		-0.037* (0.020)
<i>same religion</i>	0.514*** (0.112)	0.488*** (0.165)	0.469* (0.258)	0.206 (0.259)	0.186 (0.257)
<i>related</i>	2.722*** (0.300)	2.286*** (0.299)	2.741*** (0.650)	0.837* (0.483)	1.019*** (0.344)
<i>distance</i>	-2.204*** (0.320)	-3.317*** (0.708)	-1.922*** (0.489)	-0.868** (0.413)	-0.891** (0.364)
<i>hh members<sub>i</sub></i>	-0.003 (0.028)	-0.063* (0.038)	0.004 (0.031)		
<i>hh members<sub>j</sub></i>	0.069 (0.063)	0.048 (0.074)	0.095 (0.095)		
<i>declared friends<sub>i</sub></i>				0.065*** (0.020)	
<i>female dependents<sub>j</sub></i>					-0.098 (0.100)
<i>male dependents<sub>j</sub></i>					-0.191 (0.120)
<i>constant</i>	-3.282*** (0.210)	-3.853*** (0.266)	-2.805*** (0.326)	-0.787** (0.374)	-0.368 (0.297)

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

We have also run robustness checks using *popularity* and other observational variables as instruments for  $d_{ij}$  and  $d_{ji}$ , and the conclusion of our test are still remarkably consistent in rejecting mutual self-interest in link formation.

The usual caveat applies since these variables 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.

## Desire to give vs. desire to receive

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 one side’s 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 one side’s desire to receive help that determines  $\tau_{ij}$ , 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 Platteau and Hayami (1996) for sub-Saharan Africa.<sup>26</sup>

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,

---

<sup>26</sup>If  $j$  perfectly internalizes  $i$ ’s altruism towards him/her, then both  $\alpha$  and  $\beta$  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 mutually beneficial.

i.e.,  $d_{ij} = d_{ij}^g$ . If this were the case, then when we regress  $\tau_{ij}$  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 not grounded in mutual self-interest but 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 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  $\tau_{ij}$  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 not grounded in mutual self-interest but 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 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 significant in Tables 3 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  $\alpha$  and  $\beta$  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 Weerd and Dercon (2006) encountered when they sought to separately ask respondents who respondents would turn to and who would turn to them.



# Appendix C

Here we explain how our estimator can be implemented when researchers suspect that transfers are over-estimated instead of under-estimated, *i.e.*, when respondents may report transfers 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 rather unlikely. It should be noted that in our data few household pairs have declared a transfer from both sides (2.8% of dyads for gifts and 0.3% of dyads for loans). This means that, under the assumption of over-reporting, the number of observations for which  $\tau = 1$  is small, especially for loans. It is nevertheless instructive to illustrate the procedure.

Formally, we now assume that unless both  $i$  and  $j$  declare a transfer, it did not take place. As long as recall errors are not perfectly negatively correlated, it is also possible that a transfer did not take place even if both  $i$  and  $j$  declare it. As before, let us assume that recall errors are independent between  $i$  and  $j$ . We can write:

$$\Pr(G = 1, R = 0) = \Pr(\tau = 0) * \Pr(G = 1|\tau = 0) * \Pr(R = 0|\tau = 0) \quad (14)$$

$$\Pr(G = 0, R = 1) = \Pr(\tau = 0) * \Pr(G = 0|\tau = 0) * \Pr(R = 1|\tau = 0) \quad (15)$$

$$\Pr(G = 0, R = 0) = \Pr(\tau = 0) * \Pr(G = 0|\tau = 0) * \Pr(R = 0|\tau = 0) \quad (16)$$

$$\Pr(G = 1, R = 1) = 1 - \Pr(G = 1, R = 0) - \Pr(G = 0, R = 1) - \Pr(G = 0, R = 0) \quad (17)$$

Equations (14) to (17) 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  $\lambda(\cdot)$  as follows:

$$\Pr(\tau = 0) = \lambda(\beta'_\tau X_\tau) \quad (18)$$

$$\Pr(G = 1|\tau = 0) = \lambda_G(\beta'_G X_G) \quad (19)$$

$$\Pr(R = 1|\tau = 0) = \lambda_R(\beta'_R X_R) \quad (20)$$

The main equation of interest now is  $\Pr(\tau = 0)$ . Define  $h_{ij} = 1$  if and only if  $\tau_{ij} = 0$ , *i.e.*,  $h_{ij}$  is an indicator variable that takes value 1 if  $i$  does *not* give something to  $j$ .

We estimate a model of the form:

$$Pr(h_{ij} = 1) = \lambda(\theta'_\tau X_\tau^{ij}) \quad (21)$$

Equations (19) and (20) can be similarly transformed. The resulting likelihood function is equivalent to equations (6) to (8), but expressed in terms of  $h_{ij}$  instead of  $\tau_{ij}$ .

Estimation results are not presented to save space - unsurprisingly, the estimated coefficients for  $Pr(h_{ij} = 1)$  are similar to the coefficients of the standard logit regression where the dependent variable is  $\tau_{ij}^{min}$  and all discordances are imputed to over-reporting. Table C.1 reports the frequency of giving and lending estimated under the assumption of over-reporting. As expected, for gifts and loans the average fitted  $\Pr(\tau_{ij} = 1)$  is close to the share of transfers declared by both  $i$  and  $j$ . Accordingly, the estimated probabilities of reporting a transfer which did not take place range from 1.5% (loans received) to 5.4% (gifts given). These probabilities are substantially lower than those reported in Table 6 under the alternative assumption of under-reporting. Because the probability of over-reporting is estimated to be small, estimated coefficients for  $Pr(h_{ij} = 1)$  and for  $\tau_{ij}^{min}$  are close to each other, leading to similar inference in this case.

**Table C.1 Estimates of over-reporting**

	gifts	loans
in data: declared by $i$	0.071	0.025
in data: declared by $j$	0.059	0.017
in data: declared by $i$ or $j$ ( $\tau_{ij}^{max}$ )	0.101	0.039
in data: declared by $i$ and $j$ ( $\tau_{ij}^{min}$ )	0.028	0.003
average fitted $\Pr(\tau_{ij} = 1) = \Pr(h_{ij} = 0)$	0.023	0.002
average fitted $\Pr(G = 1 \tau = 0)$	0.054	0.023
average fitted $\Pr(R = 1 \tau = 0)$	0.043	0.015