Insurance networks outside the village and social network analysis: Evidence from rural Malawi

Author: Clifford Gerrit Westland

Student #: 5688384

Supervisor: Christopher Ksoll

A Master's Research Paper Submitted in Partial Fulfillment of the Requirements for the Degree of Master's of Arts in Economics

Department of Economics, University of Ottawa

Ottawa, Ontario

August 7th, 2015

Insurance networks outside the village and social network analysis: Evidence from rural Malawi

Clifford Gerrit Westland

Abstract. The development literature has consistently overlooked out of village links in the evaluation of informal insurance networks. This paper provides evidence that such an oversight may give a flawed impression of network formation decisions, given the high proportion of non-local transfer links in most studies to date – almost 60% in the Malawian data used for this study. Using a series of OLS and probit regressions, the results of this paper point to an endogenous propensity for households to link with non-local individuals for their informal insurance networks in a manner that distinguishes them from within-village insurance links.

1. Introduction

In this paper, I will try to demonstrate that there is a selection problem that characterizes inter-village informal insurance relationships that has been left unexamined in the literature on social network theory. In different regions of the developing world, the absence of formal financial markets requires individuals to make their own insurance arrangements for times of financial strain. Such informal insurance mechanisms entail networks of people that these individuals can rely upon as financial backstops, exchanging monetary transfers as needed. There is a growing body of research in development economics that looks to explain the dynamics of individual decisions to link with certain types of people for this purpose. With a better understanding of these incentives, we can gain some insight into the functioning of informal insurance arrangements and how to address shortcomings therein.

The research conducted in this area to date has insisted upon the need for extensive dyadic information – information on both members of a transfer network link – in order to evaluate network formation decisions. This necessitates sampling of both parties in the relationship, however, and thus necessarily restricts the data to the sample villages alone. The trouble with this restriction is that it creates an artificial geographic boundary to insurance networks that does not exist in reality. Social network analysts have seen fit to dismiss distant network links from examination, assuming that the same decision variables found to govern intra-village network formation can be extended to inter-village network formation.

This study suggests otherwise. Using unilateral survey data from rural Malawi, the results from a series of network decision models show evidence of a selection problem that arises from insurance connections between respondents and links who reside outside of the respondent household's locality. That is, there appears to be a qualitative difference between decisions to link within one's village or traditional authority and decisions to link with individuals elsewhere. My results reveal that differences in wealth, education, kinship and gender are intimately related to both decisions to form non-local insurance links and the depth of individual insurance networks, as measured by the flow of transfers between households and their network links. In the last section of the text, I explore the bias that is created by ignoring non-local links in the evaluation of social networks. If evidence of selection in the Malawian sample villages is common to informal insurance networks in the rest of the developing world, studies that regard network formation decisions at the local level alone suffer from a structural flaw that biases their findings.

2. Social Networks

The study of social networks is a sub-classification of social capital theory. The development literature on the topic is still small at this stage, but the breadth of research is growing quickly.[1] In essence, social capital theory seeks to assign measurable value to social organization and its ensuing impact on individual welfare (Santos and Barrett, 2007). Studying social networks is fundamental to research on social capital, because the benefits of interpersonal relationships are necessarily dependent on the structure and density of those connections.

The notion that the same assumptions that govern economic analysis in the western context can be transplanted into developing communities is misguided. The (relative) anonymity of the consumer-firm dynamic is irreconcilable with the interactions of low-income individuals in developing African villages for instance. However, Marcel Fafchamps writes that "market activity in Africa is not without form; it is only without economic formalization. It may escape our present understanding, but it does not defy explanation" (2004, p. 4). The economic agents and the means of production in these contexts must be specified differently. In the absence of formal social and financial structures, such is the case in most poor developing states, the social network is the fundamental mechanism for transmitting social value in a community. Networks are the economic unit of analysis in much of developing Africa (Chuang and Schechter, 2015; Fafchamps, 2004; Udry, 1990). Each is dynamic, variable in intensity and constantly in a state of flux (Durlauf and Fafchamps, 2005, p. 1654).

2.1. Social Networks and Network Formation

The social network is characterized by a community wherein agents are linked to other agents, but not to the entirety of the population under study. More specifically, it is a set of links between individuals oriented toward a common goal, whether that goal is information sharing, financial insurance, or any other social need (Durlauf and Fafchamps, 2005). Social networks are conceptually distinct, however, from exogenously determined relationships such as ethnicity,

[1] For a detailed treatment of the literature on social capital and social networks, I direct my reader to the literature reviews by Brieger et al. (2003), Durlauf and Fafchamps (2005), and Jackson (2007) who cover the central themes that continue to dominate discussion on the topics today.

gender, or religion. Santos and Barrett (2007) argue that it is far too common that this simplification is made. Studies that use these broad social groups as proxies for networks inevitably make false deductions, because the researchers make the spurious assumption that such community participation is a matter of choice. Research shows that links chosen for individual economic advantage are not synonymous with predestined social groups (Santos and Barrett, 2007). In his seminal work on endogenous social effects, Charles Manski (1993) arrives at the same conclusion. Peer-to-peer relationships and the influence of an individual's community on their own behaviour cannot be reliably inferred, he argues, without prior specification of how groups are formed – information that exogenously determined groups cannot offer.

Although the literature on social networks is diverse, it is rare that researchers neglect to weigh in on what determines the structure of individual networks. Indeed, this is the key question that underlies network analysis, because with an understanding of network formation decisions, we can begin to evaluate how well these informal organizations function and suggest means for their improvement (Aggarwal, 2007). The explanations vary from simple to incredibly complex. Casciaro and Lobo (2008) and McPhereson (2001) are good representations of the more minimalistic explanations. In their study, Casciaro and Lobo (2008) argue that social connections in the workplace are primarily determined by likeability, while the more typical determining factors like competence and intelligence are secondary. McPhereson's analysis is not greatly different. He argues that "homophilia" is the basis of network formation decisions. That is, people have a tendency to gravitate socially toward people with similar attitudes and experiences. These simple explanations suggest that informal networks are structurally suboptimal, since neither likeability nor homophily necessarily engender good performance. At the other extreme, some researchers include a dense web of determinants.[2] While their dynamism may add an element of realism not present in the basic models, they tend to suffer from loss in intuition due to overcomplication.

Most studies in development economics fall somewhere on the spectrum between these two poles. In their study of informal insurance links in rural Tanzania, De Weerdt and Dercon (2002) find that networks are determined both by social and economic variables. Their research shows that kinship, geographical proximity, common friends, clan membership, religious affiliation, and wealth have positive effects on network formation. Similarly, in their study of persistent poverty traps in subsistence villages in Ethiopia, Santos and Barrett (2006a) find that individuals' tendency to establish supporting financial links is strongly and positively influenced by membership in the same clan and gender when the links are male. More important to this latter study is the finding that recent loss in a link's herd size has a positive impact on the probability that an agent will lend to them, provided that their wealth is above a certain threshold. With this finding, the researchers conclude that informal transfer arrangements may best be understood as a means of preventing a link from slipping into persistent poverty (Santos and Barrett 2006a, 2011).

Santos and Barrett (2006a) is a good representative study of the wider research on poverty traps, which is important to keep in mind for any study on social networks. It would be too ambitious to try to reflect all of the nuances to social network theory arising from this branch of research. However, for the purposes of this study, it suffices to note that in poor communities

[2] For an example of such work, see Xu et al. (2013).

whose members substitute network links for formal insurance, there is a wealth threshold that precedes involvement in any wealth-sharing network. This is the idea behind poverty traps – transfer decisions are assumed to be based on "balanced reciprocity" such that agents make links with those whom they trust can reciprocate in the event that the agent themselves falls into financial distress. The result is that people who are desperately poor are socially invisible, because they are deemed unworthy connections (Lønborg and Rasmussen, 2014; Santos and Barrett, 2006a, 2011). In evaluating the results of the present study, it is important to keep in mind that networks are likely to systematically exclude the abjectly poor. And due to their social isolation, it is unlikely that they participated in the community meetings from which survey respondents were gathered in the first place.[3] This ought not to change the inference of the results on determinants of networks, because the absence of links is implicitly accounted for through inclusion of wealth variables. However, the impact of wealth is probably more pronounced than is observed in the analysis.

2.2. Selection and Non-Local Links

This study contributes primarily to the discussion on a specific undercurrent of the literature on network formation decisions – the unobserved bias arising from selection of economic agents. In effect, a problem that arises in studies of network formation is that there may be a non-random, but unobserved determining factor of network links. Certain types of agents tend to link with certain types of people, irrespective of the characteristics observed in survey data. Studying a similar phenomenon in a different context, Ackerberg and Botticini (2002) explain that the econometric results from their study on contract choice in Renaissance Tuscany suggest that endogenous matching undergirds network formation and that "naive" models that do not account for such effects give misleading results. They argue that contracting landlords and tenants link according to willingness to accept risk and the extent of contract oversight that can be expected from landlords, among other factors. As are the determinants of cross-border selection, these variables are largely unobserved and can only be proxied by what the researcher determines to be most appropriate to the specific case (Aggarwal, 2007, p. 476). This unsurety and potential misspecification of one's resulting model is a hallmark of the reflection problem in Manski (1993, 2000). Without appropriate information on the defining features of interacting individuals, it is impossible to know whether there is a systematic tendency for people to behave in a manner that is defined by these features.

Cassidy and Fafchamps (2015) also study individuals' tendency to match, looking at whether village savings and loan associations (VSLAs) allow potential borrowers to link with savers. They find evidence of negative assortative matching according to employment type, with farmers matching with small business owners – an efficient outcome, since the two jobs entail opposing needs for spending and saving at any given time.[4]

---

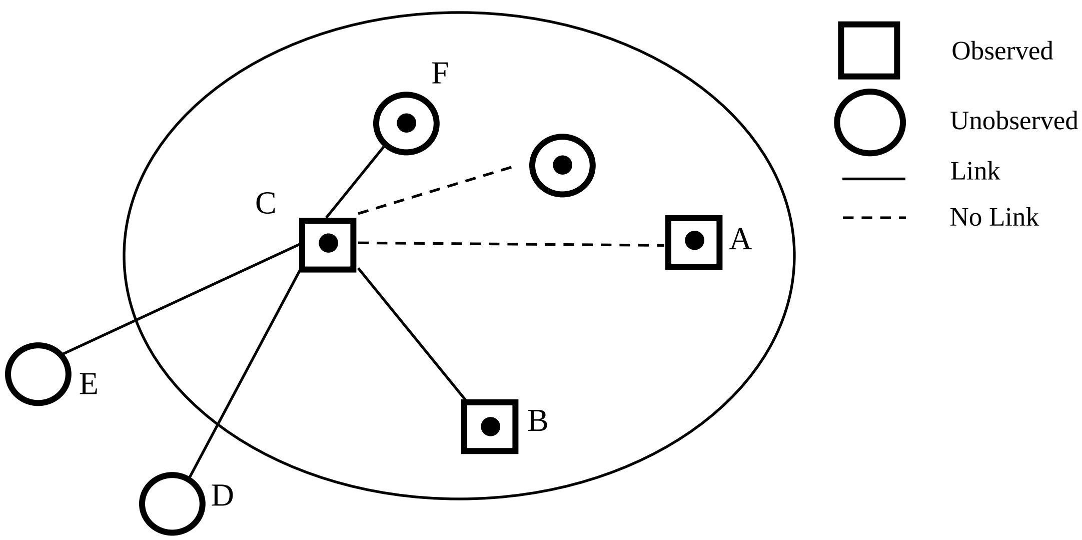
[3] Census data would be required to properly evaluate the extent of poverty traps and social exclusion in the Malawi dataset.

[4] In fact, the authors employ the same dataset used for this study, but focus only on VSLA members, whereas it was not necessary for this paper's research question to limit the data in the same way.

In this paper, I attempt to determine whether there is a selection problem that defines links made between respondents in survey villages in the traditional authority of Mwirang'ombe in Malawi and individuals outside of their villages. This type of selection is, to my knowledge, unique to the literature on social networks. Certainly, this is not the first instance where inter-village links are acknowledged as a potential fault line in the research (De Weerdt and Dercon, 2002; Santos and Barrett, 2007; Udry and Conley, 2004), but the development literature thus far handily dismisses these concerns by assuming that the same network formation decisions within villages hold across village borders. Santos and Barrett (2007) reason that missing data due to inter-village networks ought to be of no great concern unless there is reason to believe that there is a qualitative difference between the links established locally and those established elsewhere. I argue that there is such a qualitative difference.

The problem that arises from assuming a consistency of network formation decisions across geographic space is that those who fall outside of the sample region are ignored in the analysis. Ignorance of the determinants of non-local network connections could thus introduce a bias akin to the issue of the socially ignored class of individuals below the insurance poverty threshold shown in Santos and Barrett (2006a). This is no small matter. If there is a specific set of determining factors that characterize out of village links and these links are a substantial element of villagers' financial security mechanisms, as the evidence suggests[5], then studies that regard only within-village links are of questionable significance to the broader discipline of social capital (Udry and Conley, 2004, p. 6).

Figure 1: Network Diagram



[5] Out of village and out of traditional authority links make up 57% and 45% of respondents' rosters respectively, as seen in Table 2.

Figure 1 illustrates how this study differs from the literature.[6] Standard social network analysis has focused on the question of why agent C might link with B, but not with A. Likewise, in the inverted case, it asks why B links with C, but not with A. This is where the ordinary framework ends, however. If we think of the thin oval as the local boundary (either village or traditional authority), C has two links outside of the boundary, while neither A nor B has any. Yet, for the purposes of regular analysis, B faces the exact same circumstances as C, since both have one established link in the sample set and a question of why neither of them links with A. The research merely assumes that there is no substantive difference between B and C despite C's cross-boundary links. In this study, I attempt to demonstrate that such a difference exists and cannot be overlooked in further research.

In this paper, I employ all self-reported network links in the regression analysis, which would include C's links with B, D, E and F. Certainly, because D, E, and F are unobserved in the data, one cannot claim to have complete information on their dyadic relationships with C and thereby the potential influences of their specific identifiers. But through innovative structuring of the survey questions, the research team that collected the Malawi data has made a great deal of information available to the analyst. It surely does not obviate the value of having dyadic information, but it shows that focussing only on dyadic data may miss a large part of the story regarding insurance.

2.3. Sampling Techniques

There has been a reluctance to deal with the issue up to this point, primarily due to established norms regarding how data on social networks ought to be obtained and evaluated. The extant research insists on the use of dyadic sampling in order to understand networks, because relying only on one side of the dyadic relationship may create an omitted variables bias (Santos and Barrett, 2008, p. 8). After all, network links are dyadically related as a matter of course. However, in practice it is rare to observe a dyadic relationship, because both households need to be part of the collected data. For example, where a research team samples 10% of a population for social network analysis, the probability of sampling two specific members of that population and thus the probability of obtaining all information on a dyadic link, assuming simple random sampling, is 0.01. As a result, quite a bit of research debates the best means for reflecting these dyadic relationships while avoiding bias to the greatest extent.

The standard method for obtaining dyadic links is to perform either a census or random sample, asking people to create a roster of individuals on whom they could rely for financial support, and to pair people who acknowledge one another in their respective rosters. Santos and Barrett (2007) refer to this as the matches within sample approach. Like Santos and Barrett,

[6] See Santos & Barrett (2007), page 11 where the authors provide a diagram representing the "matches within sample" approach to explain patronage relationships. This is a modified diagrammatic explanation that hopes to clarify the distinction between ordinary network analysis and what I attempt to uncover in this research project. It is a simplified explanation, but in its simplicity, I hope that it is instructive.

McPherson (2001) recognizes that the census ought to be avoided to a large extent; it is a costly practice that often devolves into a large and extremely expensive availability sample.

The other prevailing sampling technique described by Santos and Barrett (2007) is referred to as the random match process of data collection, wherein a select group of randomly sampled people are surveyed, each individual being asked network formation questions only about others in that sample. Since it is rare that there is a pre-existing financial link between these randomly sampled people, Santos and Barrett (2007) promote the use of hypothetical questions about their willingness to establish a link. Thereby, all sampled relationships are dyadic. There are two implicit assumptions made in the random match sampling procedure that are of concern to this study on network theory. Firstly, if we can ascribe any importance, in the context of financial networks, to the personal likeability determinant of network formation as suggested by Casciaro and Lobo (2008), one must assume that respondents have the same information and social rapport with their random match that they do with their real links in order to establish meaningful hypothetical links. Otherwise, inference with respect to network formation decisions is questionable. Secondly, and most importantly for this study, one must also assume that the factors found to determine network formation decisions are transferable to links in different villages.

A serious concern, Breiger et al. (2003) explains, is that networks don't have definite boundaries and as a result, sampling decisions can have a great impact on how accurate one's results are. In the related statistical literature, this is called the boundary problem, and there is some attempt to formalize statistically the network formation processes. I will turn to that question in greater depth in the following section on methodology.

3. Inter-Village Transfers

In the study of social network analysis, due to the availability of data, researchers thus far have had to settle for the simplification that sample village boundaries define a reasonable limit on the scope of social networks and, by extension, social context itself (De Weerdt and Dercon, 2002; Fafchamps and Gubert, 2007; Santos and Barrett, 2007; Udry and Conley, 2004). In gathering network data, whether through a random matching approach or through compiling network rosters, resource availability has set a cap on researchers' ability to examine the dyadic relationships that extend beyond village boundaries. Meanwhile, these same researchers acknowledge that the process for establishing links among economic agents is likely non-random. Thus, in the development literature that examines links that are established on the basis of informal insurance – the focus of most papers in this field (Santos and Barrett, 2007, p. 5) – exclusion of links external to the sample village is a non-trivial matter. Udry and Conley argue, however, that the "analysis of connections entirely within the village may be profoundly misleading if the most important interactions generally cross village boundaries" (2004, p. 6).

Santos and Barrett (2007, p. 5) themselves recognise this oversight by reporting the extent of information loss in various analyses of network formation decisions. Among the works they cite are those of De Weerdt and Dercon (2002) and Udry and Conley (2004). In their study on the endogeneity of social network formation, De Weerdt and Dercon (2002) find that almost a third of self-reported lending partners are outside of their sample village in Tanzania. In this particular instance, the researchers conducted a census, meaning that the extent of information loss was limited at a third, where in the case of a random sample, the problem of missing data is exacerbated. This is the difficulty encountered by Udry and Conley (2004), where just under 60% of stated financial links are within-village.

According to Santos and Barrett (2007), Fafchamps and Gubert find an even greater extent of information loss in their paper on the determinants of risk-sharing networks. In that paper, however, they ignore inter-village transfers by narrowing the scope of their paper to within-village risk pooling, arguing that the consensus in the literature thus far has tended to make the same assumption (Fafchamps and Gubert, 2007, p. 332). That said, Fafchamps and Gubert (2007, p. 346) point out that risk pooling is frequently conducted with links outside of the village whose impact on consumption smoothing, and thereby informal insurance, is much more important than intra-village lending. Santos and Barrett (2007, p. 7) explain that expansion of one's insurance networks beyond village boundaries is just a rational reaction to uncertainty. Since economic shocks are generally more uniformly distributed within villages than beyond, lending partners in neighbouring villages serve to balance individual risk portfolios. If the intensity of insurance networks are then associated with other contributing factors – for instance, if only wealthy people establish links in other villages, as Udry and Conley (2004, p. 11) argue is a strong incentive to form links within villages – then the extent of bias from unobserved links is compounded. It is these types of patterns that have thus far escaped observation in the literature due to data availability.

Indeed, there is likely a structural component to the choice of distant or domestic network connections. In the absence of detail on relations outside of the village, Jaimovich (2013) uses out of village connections to proxy for market interactions and within-village transfers to represent reciprocated exchanges – an indicator of lesser developed economies. A greater proportion of individual networks that are outside of the domestic economy are taken to indicate market integration and a step toward transition from autarkic dependence to developed, modern economies. While this interpretation may not hold for informal credit networks in rural Malawi, there is certainly some consideration that ought to be lent to the significance of peripheral credit links.

In each of the studies just mentioned, the researchers admit to the limitation on inference that is introduced by this lack of data beyond their sample area, yet none of the research provides robustness checks that account for this particular failure. Presumably, this decision stems from the assumption that the existence of network links beyond the sample village introduces a bias that is fundamentally unknowable due to the non-random nature of social network formation.

This, of course, begs the question: to what extent does the existing literature miss the underlying trends in social network analysis by ignoring networks that exist between members of different villages? Presently, little research exists to clarify this point. In a study of what he refers to as the "boundary specification problem", Kossinets (2006, p. 265) finds that missing information on account of researchers' decisions to set boundaries on the network areas, however they justify these boundary decisions, can have momentous impact on their results. Though Kossinets does not deal with the same problem type that is addressed herein, it is of no little consequence to this study that the boundary specification problem is found to bias statistical analysis so severely.

This is where the Malawi data stand out. In collecting network rosters, the distributed surveys gathered information on the intensity of links, both real and hypothetical, between respondents and their stated lending partners. For each connection, respondents were asked to specify the amounts that he or she transferred to or received from them. The availability of this information allows us to explore the factors explaining network formation decisions with respect to informal insurance with some confidence, irrespective of the availability of details on the opposing member of a dyad.

In this study, I am not looking to uncover network formation decisions in the traditional sense. Ordinarily, development economists are trying to determine whether there is a particular series of features in individuals, both in the agent and the link, that makes them more likely financial partners. In this paper, however, I am trying to determine whether out of village links exhibit different characteristics than domestic links and whether the people who have out of village links are different from the ones who do not. Despite the lack of similar studies, there are several existing analyses that are instructive toward this end. De Weerdt and Dercon (2002) and Goldstein and Udry (1999) are probably the best points of comparison.

Rather than only looking at the statistical likelihood of a binary link variable between two respondents, De Weerdt and Dercon (2002) employ a set of endogenous variables indicating the strength of an individual's network. One example of this is a variable that adds the number of times members of the same household identify another household as a network link and the number of times that opposite household reciprocally acknowledges the respondent household. They also construct a measure of a household's network density they call "geodesic distance" that counts the degrees of separation between households.[7] Goldstein and Udry (1999) do not directly examine the informational network rosters collected in their samples; however, they discuss them as an avenue for further study. In so doing, they propose to relate the information network data to profitability of farm production in southern Ghana. Their regression models testing the impact of a series of social indicator variables on profitability give an idea of how they would incorporate the network information into the same regressions. The outcomes of interest to this study derive from these examples.

The extent of the data collected for this research project allows one to take a singular look at social networks, and more specifically out of village links, through the network roster and respondent household datasets across the three survey rounds. However, the expanded capacity to analyse cross-village interactions is by no means complete. The data continues to lack specifics on these connections, since certain details gathered from one node in a connection would suffer from measurement error. As Santos and Barrett (2007) explain, highly detailed information on lending partners' personal affairs tends to be inaccurate and could thus lead to

[7] See Table 1 in De Weerdt and Dercon (2002) for details of all their endogenous variables.

spurious conclusions if that data were used alongside the other details gathered in the survey. However, they support the argument that there is no reason to doubt information that is readily available to the respondent.

Even in the absence of highly detailed information on others' circumstances, it may be possible for researchers to gather dyadic data that is sufficient to indicate the factor that they are trying to explain. Toward this end, the Malawi survey asked non-technical, but readily observable questions about the individuals in respondents' rosters. For instance, while a link's income is not known with great precision, it may be enough to know the wealth of that link relative to the respondent. Thus, the research team created a categorical variable indicating whether the link's income was estimated to be higher than, lower than, or equal to the respondent's income at the time of the survey. Thereby, one is able to test whether credit networks are pursued with individuals of relatively higher or lower income for all members of a roster.

As explained in the next section, another feature that distinguishes the data employed for this study is detailed information on both real and hypothetical transfers to members of an individual's network roster. The use of hypothetical data is contentious in social network literature. On the one hand, using econometric comparison of real links and generated hypothetical links, Santos and Barrett (2006b) show that the two methods lead to no substantial difference in inference. This result is supported to a large extent by other research, but in more detailed analyses, hypothetical data is treated with some reservations. Harrison and Rutström (2008) contend that focus ought not to be placed on whether hypothetical bias exists; whether it exists is indisputable. The bias inherent to analyses of hypothetical data is a serious concern, but if the bias is systematic and observable in the data, a bias function can be employed to calibrate the results in such a way that outcomes can be more accurately predicted (Harrison and Rutström, 2008, p. 764).

Avoiding bias through ex post calibration is no simple task, however, because the extent of hypothetical bias is unknown in the social network literature at present and is likely contextually driven to a large extent.[8] In fact, there is reason to believe that hypothetical bias takes on a different character in making decisions that involve social networks, because people make what seem to be suboptimal decisions due to restrictions of their social circles (Casciaro and Lobo, 2008). As a result, adequate calibration of hypothetical data becomes a complex undertaking.

Although it is of no service to this study, calibration of survey questions themselves is a research area that could be valuable to social network analysts in the future. Hypothetical bias can be exacerbated when surveys ask leading questions that don't commit someone in the same way that real decision making would. Through reformulation of those survey questions in a manner that simulates the environment of decision-making, one can have respondents answer questions in a manner that better represents reality (Harrison and Rutström, 2008, p. 762). When

[8] Loomis (2014) provides a detailed treatment of how to work with hypothetical data in general. It could be highly instructive for later analysis, but his conclusions are overlooked in this paper because he does not focus on matters pertaining to social capital.

hypothetical questions imply exposure to risk, for example, respondents can be guided toward making choices that better reflect their attitude and they will be less inclined to exaggerate.

The adaptations to traditional survey designs that the Malawi dataset reflects is still an early contribution. Optimal configuration of surveys that allow for the greatest strength of inference with regard to social networks is at the core of most problems to date. In the concluding remarks of Durlauf and Fafchamps (2005), they argue that one of the most fertile avenues for further investigation is the "exploration of the extent to which ... existing survey questions are adequate in terms of dealing with the specification and identification problems" (Durlauf and Fafchamps, 2015, p. 1689) inherent to social network research.

4. The Malawi VSLA Dataset

Officially the Karonga Assessment of Vulnerability Datasets, the using data for this study was obtained from three rounds of surveys conducted in the traditional authority of Mwirang'ombe in the Karonga district of Malawi.[9] In this section, I will provide a brief description of the data, its relevance to this study, and some analytics to allow for better comprehension of the context behind this study and the study of social networks as a whole.

The data collection rounds were conducted on a yearly basis between 2009 and 2011 inclusive. Information was gathered on a random sample of individuals, evenly split (to the extent possible) among the 46 villages in Mwirang'ombe. Between 800 and 850 households were interviewed across the three rounds with the goal of maintaining the same households in each. In the case of single cross-section surveys, there is always the danger that researchers could arrive on a year that is unrepresentative of the natural state of the community – a drought year, for instance. One avoids this uncertainty by using multiple cross-sections, because in the event of a sudden and unexpected deviation of the data from its trend in one round, the researcher still obtains reliable information based on data collected in the other rounds.

The survey rounds were organized to coincide with the rolling out of a village savings and loan association (VSLA) intervention that the research team arranged in the region. Most simply, a VSLA is a small community-organized savings and loan group, provided with no external funds, that is introduced to subsistence communities. Their purpose is to formalize insurance mechanisms for the mutual benefit of all participant savers and borrowers alike.[10] Risks are assumed entirely within each VSLA. In the case of Malawi, the goal was to develop small, self-sustaining financial markets in the area, since no capital is provided from outside of the community.

The central purpose for the collection of the data was to evaluate the progress of the VSLA project, using a cluster randomized control trial (Ksoll et al., 2013). Although the present study

[9] The data is publicly accessible on the Rockwool Research Foundation website, along with a thorough description of its collection and stated goals (Lilleør et al., 2012). For a more detailed treatment, I direct my reader to the citation provided.

[10] For more on VSLA projects, see Allen and Staehle (2007).

does not directly evaluate VSLA activities – rather, it uses other questionnaires distributed in both the control and treatment villages that record non-pertinent information to the VSLA project itself – it is important to note that there is a non-random element that is introduced through the sample collection process, because the survey over-sampled respondents who were interested in partaking in the VSLAs (Ksoll et al., 2013, p. 11). It is unknown whether there is a bias that is introduced as a result, beyond the persistent poverty bias acknowledged in the *Social Networks* section of this paper.

There are two questionnaires from this research project that serve as the basis for analysis in this study – the household questionnaire and the network roster questionnaire. The first details information on each sampled household, from amount of land ownership and the household's main source of lighting to the age of the household head and his or her years of education. The network roster questionnaire asked respondent household heads to list all of the people who could be relied on for financial relief in times of hardship or that could rely on the respondent household in a similar predicament. In order to gauge the strength of network links, further questions were then asked regarding both real and hypothetical transfers between parties.

In the context of the Malawi dataset, it should be clarified that when I refer to real transfers, I mean the Malawian kwacha-denominated value of transfers that have been exchanged between individuals. This is not restrictive to the members of the roster who have either received or given the respondent money. If a link has yet to have any true exchange with the respondent, the transfer amount is given as zero; it is not ignored in the analysis. In order to gather information on hypothetical transfers, on the other hand, the surveys asked respondents to give the amounts that they would be willing to provide or ask from their network links if necessary.

As discussed in the introduction, the use of a combination of the network roster and household roster datasets, without subsequent merging of households into dyadic relationships[11], is rarely seen in research on social networks. However, it would serve no purpose to neglect non-dyadic relationships, because these very relationships serve as the object of this research project. A study of out of village links necessarily requires one to work with one-sided and thus limited instruments. Beyond the relative limitations in detail caused by using the network transfer rosters as the sole basis for analysis, the usual techniques known from social network analysis cannot be used, because there are no cases of unlinked individuals.[12] That is, the transfer network dataset observes unilateral links between respondent households and individuals outside the household and does not reflect relationships between people unknown to one another. Thus, it cannot make any statement with regard to distinctions between linked and unlinked individuals.

However, in many studies that aim to uncover the determinants of social network formation, the researchers use binary logistic regressions with dyadic data to examine the probability that respondents mutually acknowledge a link, using a series of respondents' characteristics as the independent variables (Santos and Barrett, 2006a, 2006b, 2007, 2008, 2010, 2011; Udry and

[11] See, among others, Ackerberg and Botticini (2002), Aggarwal (2007), Breiger et al. (2003), David (1988), and Santos and Barrett (2007)

[12] Certainly, previous studies such as De Weerdt and Dercon (2002) may include unilateral rather than bilateral links, but they are never the only type of link recognized in the analysis.

Conley, 2004). In these studies, the researchers want to know whether there are determinate features that distinguish links and non-links. In this research project, I am trying to distinguish between links and other links. Using the network roster dataset, I can test whether there are distinguishing features that offset domestic links and out of village links.

In order to give the reader an idea of the breadth of data included in the analysis, it is important to detail some of the dataset's key features. Keeping the number of survey households constant, the number of links provided increased in each successive round. Thus, among the 46 households who provided network transfer rosters of their trusted financial links, there is information on 6879 relationships in the first round, 7402 in the second, and 8469 in the third. This translates to 4.8, 5.8, and 6.2 people on the average network roster in rounds 1 to 3 respectively.

A common issue that complicates analysis in any survey-based study is inconsistency of information from one round to the next. The problem is twofold: firstly, it is common that survey input data is simply incorrect in one of the rounds; secondly, in the span of a year, people often forget about links that they provided in previous rounds. The extent of inconsistency between rosters across survey rounds in the Malawi dataset was mitigated, however, because after the inaugural round, rosters were pre-generated for the research team and respondents based on previous rounds' entries, with an option to make any necessary changes. As a result, a substantial majority of network connections (specifically 5954) were maintained throughout all three rounds.

Although individual rosters are largely consistent throughout the three rounds, there remains a fair bit of shuffling of links over the years. In 207 cases, lending partners were dropped from rosters in the third round. In 465 cases, lending partners were dropped in the second round and restored in the third round. In 1200 cases, new lending partners were added to the roster in the second round and maintained in the third round. In 250, 39, and 847 cases in each successive round, an individual was added to the roster who was not identified in either the following or preceding rounds. Some of the inconsistencies detailed here can be attributed to attrition in different rounds, but since attrition rates were generally low, the bulk of the remainder can be regarded as new links, dropped links, or input errors that were not caught upon review.

5. Econometric Framework

The econometric models used in this paper must necessarily diverge from those in the recent social network literature. As has been explained above, the data used for this paper is restricted to the network roster alone and does not include dyadic relationships. The result of this discrepancy is that the logistic regression that serves as the basic model for Santos and Barrett (2007) and Udry and Conley (2004), among others, is inappropriate to determining network formation decisions in the context of this study's research question. In those logistic models, the researchers test the probability of establishing a link between two individuals. In order to generate the binary link variable used as the endogenous term therein, one must necessarily have dyadic data, meaning that both individuals and information on their characteristics need to be in the sample. Not only is the data from the network roster non-dyadic, but each connection in the dataset represents what De Weerdt and Dercon (2002) call a unilateral link, meaning that exploring the probability of establishing a link between two individuals would be futile.

For consistency with the literature exemplified by Santos and Barrett (2007) and Udry and Conley (2004), one set of regressions uses the binary variable indicating whether a link is non-local as its dependent variable. Following the proposed methodology of Goldstein and Udry (1999), another one of the endogenous variables, meant to reflect the strength of a connection between households, is the Malawian kwacha value of inter-household transfers. The goal here is to determine whether there is a marked difference between the transfers exchanged with non-local households than with households in the respondents' villages or traditional authority. Without dyadic data, the endogenous variables employed by De Weerdt and Dercon (2002) escape investigation. However, I borrow their concept of employing network density as a dependent variable, with a respondent household's total number of cross-boundary links meant to indicate the density of that household's non-local insurance network.

In accordance with previous research on social networks, the empirical models chosen for this study follow either a multivariate probit or OLS regression structure. Details of each formulation of these models follow shortly. The regression models can be separated into two classes: link level and household level. As the titles suggest, link level regressions were run on a per-link basis while household level regressions collapsed data at the household level to assess whether there were significant characteristics of links and the households themselves that escaped observation at the link level. In this latter set of regressions, link-level variables were transformed into count and ratio variables in order to accurately represent the contents of household rosters. Note that only living members of network transfer rosters were considered in all of the regressions. Links that died between rounds were removed from analysis only in the year that they were acknowledged by respondents as having died.

5.1. <u>Link Level Regressions</u>

Using the merged household and network roster dataset, the following model was constructed

$$\Delta_i = \alpha + \mathbf{X}_{\mathrm{i}}\beta_1 + \mathbf{X}_{\mathbf{j}}\beta_2 + \epsilon_{\mathrm{i}} \tag{1}$$

Where Δ_i represents the net transfers between the household and the roster link. It is a constructed variable, calculated as assistance received less assistance given. A positive term on the Δ_i variable indicates that the link represents a net inflow of assistance, whereas a negative value indicates that the household provides more transfers to the link than it receives in return. Δ_i is meant to signify the terms of the respondent household's financial relationship with the link. The use of net transfers, as opposed to another metric representing the value of financial links, is consistent with McPeak (2004) and Bloch, Genicot, and Ray (2008). Variables such as total value of transfers or the square of net transfers would misrepresent the significance of links. The two variables would count a borrower of 1000 kwachas and a lender of 1000 kwachas as being indistinct, which removes the researcher's capacity to recognize a tendency to lend and borrow in the regression results. Net transfers is favoured because it most accurately captures the exchange between two individuals. See Table 1 for details on the link level regression dependent variables.

The $\mathbf{X_i}$ and $\mathbf{X_j}$ matrices are collections of explanatory variables associated with the respondent and the link respectively. The variables included in each set $\mathbf{X}$ are characteristics of both the respondent household and the link that are likely to impact whether a link is established. Each variable was selected in accordance with variables found to be significant contributors to network formation decisions in the literature. There are four factors that are seen to be most important to the establishment of financial links: family relation, education, wealth, and gender. Several variables were employed to represent each in the regressions. After completing an account of the regression models, I will provide a brief synopsis of the key factors evaluated in the regressions. Table 2 provides the summary statistics on all of the variables used in the regressions and considered relevant to research question, while Appendix B defines each of the variables using the survey questions that respondents were asked.

The reader should note here that the same regression structure was employed to test hypothetical net transfers between respondents and their links. It would be of little value to show a second equation for this, but the regressions on hypothetical links warrant separate mention. It is only the endogenous variable that changes in this case; the exogenous variables are still those seen in Table 2.

The second model used to exploit the link level data is a logistic regression, using as dependent variables the binary relations of whether the link is i) outside of the respondent's village; or ii) outside of the traditional authority altogether. Define the locality variable as

$$y_i = \begin{cases} 1 & \text{if the link is non-local} \\ 0 & \text{if the link is local} \end{cases}$$

The y_i endogenous variable is defined broadly to allow for both interpretations i) and ii). Using y_i, we can then fit the logit model

$$\log\left(\frac{y_i}{1-y_i}\right) = \alpha + \mathbf{X}_\mathrm{i}\beta_1 + \mathbf{X}_\mathbf{j}\beta_2 + \epsilon_\mathrm{i} \tag{2}$$

This second regression model indicates what independent variables increase the probability of establishing a link outside of the household's village or traditional authority. For the purpose of this study, I have kept to a rather basic formulation of a logistic regression such that the same set of $\mathbf{X_i}$ and $\mathbf{X_j}$ can serve as the regressors. Certainly, the interpretation of the resulting coefficient estimates on these variables become less functional than they are in the OLS model, but the goal is to uncover an overlooked bias in the social network research. Thus, nothing more specific than the signs of the independent variables' coefficient estimates and their statistical significance are required. The logit model designed above is an indispensable element of the analysis and is the mechanism that most clearly addresses the research question at hand.

Before turning to an explanation of the set of household level regressions, it is important to provide justification for the use of the key independent variables discussed both in the results section of this study and in the wider literature on network formation decisions. Recall that there is a certain perspective that the reader should take when reading through the following. Although the respondent household's and link's attributes are discussed as they pertain to social network formation in general, the target of this research project is to understand those features as they impact decisions to link with individuals outside the local area – either their village or traditional authority – as opposed to within it. With that in mind, these are the variables that are highlighted in this paper's analysis.

Family Relation

Survey respondents were asked how their links were related to them and their responses were subsequently coded as being family or non-family. Kinship is one of the most prevalent suggested determinants of network formation in the literature and represents the "dominant form of social insurance" in developing Africa according to Fafchamps (2004, p. 9). Udry and Conley (2004) explain that family members are often preferred sources of financial support, because the extent of social familiarity in the dyad ensures reciprocity.[13] But kinship involves more than self-interested assurance against harm. In their study on the implications of social capital to development research, Woolcock and Narayan assert that "a person's family, friends, and associates constitute an important asset, one that can be called on in a crisis, enjoyed for its own sake, and leveraged for material gain" (2000, p. 226). Their objective analysis certainly agrees with that of Udry and Conley (2004); yet, they admit a tendency to support one's kin for no other reason than intrinsic affection. The Malawian data tends to agree with this analysis. Table 2 shows that nearly 88% of network roster members are family members.[14]

However, the family composition of households' transfer network rosters would be of little consequence to the present research if family links are concentrated within the same village as the respondent household. While members of a family are less inclined to spread across villages than are unattached individuals, Mincer (1978) demonstrates that when the net benefit to the family unit from migration is positive, that hesitation tends to be overcome. Moreover, evidence shows that people's tendency to gravitate toward family members as insurance links is not mitigated by distance or time (Rosenzweig, 1988). Thus, the drive to link with distant family members is an important consideration for network formation.

Education

The main indicator for education used in this study is the respondent household's level of education in years, because it is straightforward and measured in unit increments. However, I have employed two measures of education as robustness checks. These measures are: i) share of adults in household who can read; and ii) proportion of household members without education, excluding head. In this research project, we know only the education level of the respondent household, but not the link. Education is expected to have a positive association with decisions to

[13] The authors actually regard family bonds as insurance against reneging on debts; however, the implication is one of mutual assurance against poor behaviour.

[14] In fact, the abundance of family links in network rosters may have been influenced by the survey structure. Respondent households were asked to list their links beginning with family members, moving onto non-family members thereafter. This may have created an implicit ranking of links that influenced the revealed networks, but we will assume for the purposes of this study that this is of minimal concern.

establish a link. Beyond the effect that higher education would tend to have on income, De Weerdt and Dercon (2002) reason that there are external benefits to forming links with an individual who has more education. More educated people tend to be better information links, for instance (Udry and Conley, 2004). If a more educated link is a close contact, that link can assist with tasks that require a formal education and thereby directly benefit the agent – the example given by De Weerdt and Dercon (2002) is assisting a family member to read pamphlets or fill in forms.

Table 2 shows that the sampled villagers generally demonstrate a low level of education, with a mean household head education level of under 7 years. While there are some heads of households who attain up to 19 years of education, the standard deviation of the variable suggests that such outliers are highly uncommon. That said, the other indicators of education tell a more positive story. There is a high mean proportion of household members who can read. Meanwhile, the mean value of the proportion of household members without education is very low with a concurrently low standard deviation. In fact, the maximum proportion of household members lacking education is 0.5. Thus, while the number of years of education obtained in the sample villages is low, most individuals are able to obtain a functional level of education.

Wealth

A household's property acreage is used as the primary indicator for wealth in the regressions. As can be seen in Figure 2, the distribution of land over the sample set is what one would expect of wealth in a community largely oriented around agriculture and subsistence farming. While there are a number of outliers, the distribution is smooth and unimodal, with a concentration of households with small plots of land. The metrics on lighting and the size of the household's home are good alternatives.

As explained above, wealth is a difficult concept to reflect in the data when the subjects under study live in subsistence communities such as the villages in Mwirang'ombe. Year by year consumption and income can be unreliable determinants, because in circumstances where agriculture and small business make up the majority of employment, these variables can rapidly deviate from average. If surveys are conducted in off-years, they can gain a poor understanding of business-as-usual (McPeak, 2004). Thus, possessions and associated characteristics at the baseline survey period are used to proxy for wealth, because these things tend to remain fairly constant across time and within households.

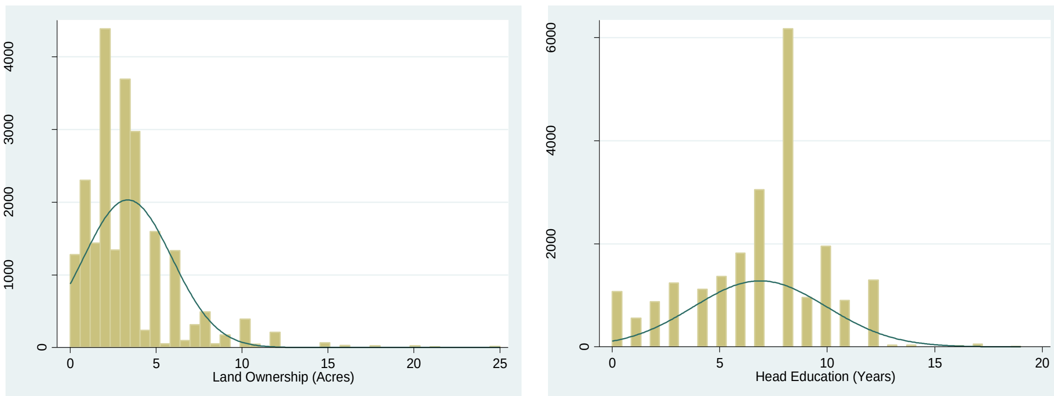
Wealth is of interest to the question of whether there is a distinction to out of village links for two reasons. The first is that the absolute value of transfers given and received is expected to increase with wealth (Udry, 1990, p. 255).[15] Secondly, one would expect people who are most wealthy to have business or family links outside of their villages. Wealth and family links could well be associated, because wealthy families likely have greater mobility between villages.[16]

[15] Admittedly, Udry (1990) refers to credit transactions when he recognizes the impact of wealth on borrowing and lending. It is assumed here that for the same reasons that more wealthy people are able to lend and receive more credit, they are also able to lend and receive more in transfers.

[16] Mincer (1978) provides good coverage of family migration decisions.

The indicators of wealth in Table 2 confirm that the sample villages are not affluent. Only a third of households sampled employ modern means of lighting their homes – meaning they use some form of electricity or battery. The mean number of rooms in a respondent's household is around three, while the mean number of people in a household is close to 6, indicating that space is limited.

Figure 1: Land Ownership and Education of HH Head in Mwirang'ombe



Gender

Gender is also of interest to the analysis on types of network formation. Udry and Conley (2004) include gender dynamics in their analysis where this study does not, but their discussion regarding gender is still important to recognise. They explain that in rural Ghana, people tend to link according to gender with a low probability of cross-gender links being established (Udry and Conley, 2004, p. 10). While this study cannot confirm their finding due to the lack of a gender dynamics indicator, the data herein allows one to look at whether households have a greater tendency to link with a particular gender or whether they have a proclivity to establish out of village links with members of another gender. This question would be trivial if there was little representation of one sex or the other in the data, but that is not the case. Table 2 shows that network links are distributed almost evenly between genders.

5.2. <u>Household Level Regressions</u>

As explained in the introduction to this section, the household level dataset is a configuration of the link level dataset, such that one observes the household's network formation decisions as opposed to individual agent-link analysis. The intuition behind the decision to include these regressions is that such aggregate statistics can reveal an element to out of village selection that would otherwise be overlooked. For example, the density of a household's out of village network is impossible to account for with link level observations.

The first model employed for household level analysis is

$$\widetilde{\Delta}_{\iota} = \psi + \widetilde{\mathbf{X}}_{\imath}\theta_1 + \widetilde{\mathbf{X}}_{\jmath}\theta_2 + \widetilde{\epsilon}_{\imath} \qquad (3)$$

Hereafter, for the purposes of this study, I will use tilde to represent aggregate household variables. $\widetilde{\Delta_\iota}$ refers to the net transfers exchanged by the household. A positive value of $\widetilde{\Delta_\iota}$ indicates that the responding household is a net receiver of assistance as a whole, while a negative value signifies net outflow from the household. The household net transfer variable represents either real or hypothetical flow of assistance into or out of each sampled household.

The interpretation is not very different from the non-aggregate Δ_i, but it must be admitted that the aggregate variable is less informative in absolute terms, because households' assistance relationships can be distorted where decisions are dynamic and outflow is canceled out by inflow. The reason that it continues to be included is that it is assumed that households generally tend either toward inflow or outflow depending on individual need. In terms of real transfers, households that must turn to their financial network in emergency circumstances – which is the stated purpose of these transfers – ought to be unlikely to give out any significant transfers themselves. Doing so would effectively just make them intermediate agents in financial transactions. Even in the case of hypothetical assistance, where respondents are more inclined to embellish their decisions, households in poor financial condition are still unlikely to offer a net outflow of assistance. The reasoning for the use of household level net transfers rather than another form of representing the value of transfers follows directly from the description given in the link level regressions. I refer my reader to that discussion for further details. See Table 3 for details on these and the rest of the household level regression dependent variables.

As in the link-level regressions, the **X** matrices included in the OLS regressions represent household and link attributes. Like the endogenous variable in this model, the exogenous variables track those in the individual analysis closely and in all cases of household characteristic variables, they are exactly the same. Thus, reinterpretation of their characteristics would be of little practical value. Note that when variables indicated particular elements of a relationship between a household and a link, however, the variables were either transformed into aggregate level statistics or were dropped if they could not be included in a meaningful way. There were three such variables that made it into the final regressions – the ratio of family to non-family members in the household's transfer roster and the fraction of links with either higher or lower economic status relative to the respondent household. For the full set of regressors, see Table 4.

The second household level regression is similar in form to the first. However, the set of variables that are chosen to serve as the exogenous variables give a different interpretation to the model and therefore require separate recognition. The model is given as

$$\widetilde{n_\iota} = \psi + \widetilde{\mathbf{X}_\iota}\theta_1 + \widetilde{\mathbf{X}_\jmath}\theta_2 + \widetilde{\epsilon_\iota} \tag{4}$$

where $\tilde{n}_i$ specifies either the number of non-local links – links that are either out of village or out of traditional authority – that household *i* maintains in each survey round. Table 3 shows that the mean number of links outside of the respondent's village is over 4, while the mean number of links outside of the respondent's traditional authority is over 3. Since the average household maintains several links that reside outside of their village or traditional authority, the regression results can be expected to be more reliable than they would be if only select households maintained links outside of the villages. When little data exists, it becomes harder to generalize results to a wider population. With regard to the independent variables in the model, the $\widetilde{\boldsymbol{X}}$ matrices are interpreted in the same manner as in the first household model above for all intents and purposes.

This set of models is crucial to the analysis on this paper's research question, because the endogenous variable is indicative of the network density, delineated into out of village density or out of traditional authority density depending on the identity of $\tilde{n}_i$. The OLS regressions demonstrate the impact of each independent variable on the number of certain types of financial relationships that the typical household maintains.

It may seem inconsistent with the rest of the paper that the number of hypothetical links is not assessed in this set of regressions. But it would be meaningless to study the number of hypothetical links as a dependent variable due to the extent of overlap between real and hypothetical observations. Certainly some network roster members exist only as hypothetical links, but households who identified links as real transfer links also responded to the hypothetical questions regarding the same link in most cases. Thus, it would be misleading to talk about numbers of hypothetical links as a distinct group.

Given that the Malawi VSLA dataset consists of panel data and that several observations are sourced from the same household in each survey round, one would expect the error terms returned in each regression to be correlated and heteroscedastic within households. In order to account for this problem, I have employed cluster-robust standard errors, using households as the cluster variable. The reported results in the next section all follow this pattern. As a robustness check, I have included regressions using the respondents' villages as the cluster variable. Each of these regressions is given at the end of this paper in Appendix A.[17]

6. Results

6.1. Link Level Regressions

The results of the link level regressions are shown in tables 5 and 6. What is immediately apparent is that the variables that were expected to be significant to the respective models are indeed significant, despite some notable exceptions. In the set of net transfer regressions (Table 5), there is a positive and significant effect of non-local links in all but one of the model configurations. Given the interpretation of the net transfer endogenous variable at the link level, the results of regressions 1 through 3 indicate that non-local links tend to provide between 200 and 300 kwachas more in transfers than local individuals do.[18] Variation in the result depends mostly on what indicator of wealth is chosen in the regression. Since the majority of links are outside of respondents' villages, it appears that non-local links are a greater source of incoming

[17] An alternative strategy that was pursued during the model creation phase of the research was using a random effects regression, anchoring the data by household and survey round. Both time constraints and a relative loss of intuition prevented its inclusion in this paper, but it will be given due recognition in later work on this topic.

[18] We can compare this to the standard deviation in net transfers which is 2579.88.

transfers than local links. This finding is robust to links either outside of the respondent household's village and its traditional authority.

Comparing the results for non-local links' association with real and hypothetical transfers, one notices a falling off in the significance and magnitude of the coefficient estimates in each respective regression. This can be seen in comparing the values of "Does (…) live in another village?", "Is (…) a family member?", or "Age of household head" among others in columns 1 to 3 with the same variables in columns 4 to 6. Looking at the descriptive statistics for real and hypothetical transfers respectively, it is apparent that the magnitude of hypothetical transfers is generally lower than that of real transfers, so the regression result is not too surprising. However, as Harrison and Rutström (2008) remark, it is well recognized in the literature that hypothetical bias tends to inflate individuals' perceptions of monetary values, making this result comparatively odd. Since net transfers are the variable of interest, perhaps when households are contemplating their hypothetical transfer decisions, they tend to inflate their generosity relatively more than they do their willingness to accept. This would tend to drive the value of net transfers down. That assessment is consistent across coefficient estimates on the various independent variables, irrespective of the particular variation of the net transfer regression.

The behaviour of the binary family variable in both sets of regressions seen in tables 5 and 6, is consistent with earlier predictions. If a network roster link is a family member, there is a statistically significant tendency for the household to transfer money to them. Looking at real transfers in regressions 1 to 3, we see that people lend between 150 and 200 kwachas more to family members than to others. The fact that a link is family is positively and significantly associated with the probability of that link being outside of the household's village or traditional authority, as is observable in the logit models shown in Table 6. These results are consistent with general theory on network formation decisions and the predictions given in the methodology section of this paper. The result demonstrates that people are more inclined to link with members of their family and to be more generous in their transfers to these individuals than they are with friends, neighbours, or colleagues.

The education parameter estimates are statistically insignificant in all constructions of the link level models. This result is robust to all variants of the education variable as provided in Table 2. Theory predicts that there is a higher drive to connect with educated people due to the range of benefits they confer on people's lives, as evidenced by De Weerdt and Dercon (2002). Moreover, the direction of a transfer relationship would likely more often be from higher educated to lesser educated people, if only due to the advantages in working skills that education confers. The distribution of education shown in Figure 2 suggests that the Malawi sample communities lend themselves well to this type of exchange structure. Household heads' education is concentrated at the fairly low level of around 7 or 8 years, with a number of outliers nearing 20 years. Despite this, the link level regressions can make no clear statement on the value of education to link formation decisions.

A relative difference in wealth between households and their links is generally significant and follows a predictable relationship in the regressions. People who are recognized as better off tend to be sources of transfers, while people who are worse off tend to be recipients. Respondents who are much better off than their real links tend to transfer around 1200 kwachas more to these individuals when needed than they would to links who have the same economic status. In almost perfect opposition, respondents who are much poorer than their real links will receive around 1000 kwachas more than they would from links whose wealth is comparable. This result is unsurprising. A notable trend in the logit models, however, is that households are more likely to connect with individuals living in other villages or traditional authorities if there is a perceived difference in wealth between those people, irrespective of what that relative difference entails. The fact that all coefficient estimates on the relative wealth variables are positive suggests that there is a tendency to either help or receive help from links that reside outside of the respondent household's locality, while people thought to be on equal financial terms are less likely to interact across village borders.

Household estimates of their insurance links' finances may not be dependable, so I also look at the wealth proxy variables from the respondents' households. The regressions were checked for consistency in all cases with no significant differences in results. It is interesting that land ownership only becomes significant for the hypothetical transfer regressions. If relative income can be believed, there is little difference in the impact of wealth on transfer decisions. However, the wealth proxies indicated that people may only factor wealth into their considerations when faced with hypothetical interactions.

Along with the discussed relation of education to transfers in the previous paragraph, this would lend some weight to the arguments of Casciaro and Lobo (2008) and McPherson (2001) who say that in practice, people have a tendency to rely on their emotions rather than their rationality in forming social networks. Taking this line, the results could suggest that Malawian villagers link to family members, older people and men more readily than they will consider who the most appropriate lending partners are. The argument falters when it comes to establishing links outside of the village, however. The extent of a household's property ownership improves the likelihood that their link is outside their village or traditional authority, as demonstrated in all four variants of the logit model with at least 10% significance.

The regressions suggest that gender is not a dependable inspiration for network formation beyond the settlement of transfers. It is clear that gender is connected with the magnitude of net transfer decisions, both real and hypothetical. In Table 5, one can see male links are consistently associated with the highest coefficient estimates on the complete set of dummy variables across the regressions. This would indicate both that male links are much more likely to be donors than recipients of transfers and also they tend to transfer more than women do. The robustness checks reveal that having a female head of household is also strongly correlated to the amount of transfers that that household receives. However, when it comes to the logit regressions, gender has no clear connection with decisions to link.

6.2. Household Level Regressions

For the household level regressions, refer to Table 7 for the net transfer regression model results and Table 8 for the network density regression results. Unlike in the link level regressions, neither variable indicating a household's connection with non-local individuals is significantly related to the net transfer variable. What is curious is that there is a strong negative and significant association between non-local links and hypothetical transfers between links. Thus, the number of links outside of the respondent household's area corresponds to a hypothetical outflow of funds. Households with one more link outside of their village than another are hypothetically willing to transfer almost 400 kwachas more, whereas their real activity shows that they tend to receive between 100 and 200 kwachas more in transfers.[19] On a superficial level, it appears that households' hypothetical willingness to lend is at odds with their real activity when more of their links reside outside of their locality. Like with the wealth proxy variables in the link level regressions, it could be that households don't systematically factor non-local links into the values of their real exchanges, but consider them only with prospective connections. Without more information on the households and those relationships, however, it is difficult to speculate as to the reasoning behind this anomaly.

There is an unequivocally positive and significant relationship between higher ratios of family members on household rosters and the number of links that the household has outside of its village or traditional authority. However, the ratio of family relations to non-family relations on a household's transfer roster appears to have no significant interaction with household real net transfer decisions. This is surprising, given the extent of significance on the relationship between the roster link being a family member and link level transfer decisions. However, perhaps there is an underlying relationship between the types of individuals who list higher proportions of family members on their rosters and the amount of transfers they provide and receive. For instance, in accordance with the link level analysis, it may be that ordinarily households more readily lend to family links, but that older people and parents generally have a higher proportion of family links than the general population. Since the elderly and aged parents can be expected to be net recipients of transfers, their inclusion in the regression would confuse inference on the family variable's association with net transfer decisions in the household regressions.

Both the education and wealth indicator variables are consistent with previous understandings of their relationship to network formation decisions, as discovered in the link-level regressions.[20] They are of little explanatory value for net real transfers, but hypothetically, households with more education and wealth appear to be more willing to extend funds to their transfer roster connections than are other households. Furthermore, the results in Table 8, show that more wealthy and educated households are positively associated with the number of links that the respondent household has outside of its village or traditional authority.

The link gender variable was omitted from the household level set of regressions, because no constructed proportion figure properly illustrated the gender presence within an individual household roster. The female respondent household head variable is thus the only indication of gender in this set of regressions. Households with a female head have a strong and significant

[19] We can compare these figures to the standard deviation in net household transfers and hypothetical net household transfers, which are 9934.87 and 4886.43 respectively.

[20] There was some suspicion that education and wealth may have been jointly significant to the regression, but a joint significance test yields no marked increase in significance for the two variables.

tendency to be net receivers of transfers to the effect of between 2000 and 3000 kwachas. This result is robust to real and hypothetical transfers. However, there is little indication that these households are associated with the density of non-local links in a household's transfer network roster.

7. An Assessment of Bias

The results reported in the previous section are suggestive of the determinants of network formation and how financial relationships change based on the characteristics of both providers and recipients of transfers. While there is some indication that links residing outside of one's village or traditional authority are chosen for different reasons, I have not given clear evidence that the exclusion of non-local links from regression analyses leads to biased coefficient estimates on network formation decision variables. In this section, I will attend to that gap. If members of a village have different types of connections within their village than they do outside of their locality, any attempt to predict the factors influencing network formation that do not account for both local and non-local links will be flawed.

In order to investigate whether non-local links are systematically distinct from local links, I have restructured the link level regressions in equation 1 as follows

$$\Delta_i = \alpha + \beta_1 nloc + \mathbf{X}_\mathrm{i}\beta_2 + \mathbf{X_j}\beta_3 + (nloc \times \mathbf{X}_\mathrm{i})\beta_4 + (nloc \times \mathbf{X}_\mathrm{j})\beta_5 + \epsilon_\mathrm{i} \qquad (5)$$

where *nloc* refers to the binary variable indicating whether a link is non-local – either out of the respondent's village or traditional authority. The use of interaction terms for each of the independent variables used in the regression allows one to see if the determinants of net transfer flows are the same for different types of links. The results of equation 5 are provided in Table 9. Note that only regressions 1 and 2 from Table 5 are recreated therein.

The results indicate that non-local links are indeed distinct from those within an individual's village or traditional authority. The most telling indicator to that effect is the set of coefficient estimates associated with the household head's years of education in regression 1. The base coefficient estimate on the household head's years of education alone tells us that higher levels of education increase the household's net transfer to local links by 34 kwachas per additional year. Meanwhile, if a link resides outside of the respondent's village, the household head's being educated another year results in a net reduction in transfers by 38 kwachas. The fact that links are non-local results in a complete reversal of the relationship. Using a t-test, one finds that the sum of the base and interaction terms is statistically different from zero. Thus, education is shown to have a statistically significant and opposite association to links within and outside an individual's village.

The household head's age also appears to separate links according to locality. In both regressions 1 and 2, the results show that all else equal, a household head that is one year older than another can expect to receive a net inflow amounting to between 4 and 7 kwachas from local links. If a link is non-local, the same household head can expect to receive around 10 kwachas in addition to the amount received from local links.

The impact of gender, wealth, and family variables is less clear. While the gender and family base variables are significant – most beyond the 1% significance level – with the same signs and interpretations found in earlier discussion of Table 5, none have a significant interaction term, meaning that they do not clearly distinguish local links from non-local according to these criteria. The reverse is true of the property acreage indicator for wealth. Though the base variable is not significant within the 10% range, regression 2 shows that non-local links provide larger transfers to individuals with more land.

Despite the above results, the most comprehensive test for bias is not to look at individual coefficient estimates, but to see if the variables jointly have a different association for local and non-local links. The results of f-tests for whether the determinants of these two types of links have a jointly significant difference are reported at the bottom of the regressions in Table 9. In the case of the first regression, we fail to reject the test while it is rejected in the second. Thus, we find some indication that the two types of links have different determinants.

The evidence from the t-tests and f-tests provided in this section indicate that the practice of ignoring links outside of the researcher's sample area can lead to biased results and subsequently biased interpretation of the evidence of network formation. Given that this is particularly true for such important variables as education and property ownership, the bias is likely to be non-trivial.

8. Conclusion

This study finds evidence of a systematic distinction between informal insurance link formation at a local level and at an inter-village or inter-traditional authority level. Among other factors, it appears that households that choose to link with people beyond their locality are demarcated by kinship, education, wealth and gender in a manner that others are not. This finding suggests that studies that have aimed to generalize network formation decisions by looking only at dyads within their sample villages are inclined to miss a substantial part of the nature of social networks. An evaluation of the distinction between decisions to link with local and non-local individuals reveals that excluding the latter category of links from analysis would bias the results. Given the apparent extent of financial interaction across village boundaries, the implication of this result is non-negligible.

A point that is worth reiterating is that by necessity, the data that was used in this study breaks from the norm set in the social network literature. Whereas others have studied the propensity to establish a link using information on both nodes in dyadic link, the data for this research project had to be restricted to unilateral survey responses, since the dyadic link responses are consistently unavailable in the case of cross-boundary links. That said, this is an avenue of study that I intend to pursue further. The sample set for the Malawi VSLA data included all 46 villages in Mwirang'ombe. With this data, it is possible to pair all dyads within the traditional authority and examine whether there is any discernable difference between intra- and inter-village links using models designed to test dyadic relationships as in Santos and Barrett (2007) and Fafchamps and Gubert (2007).

More research is required both to verify the results of this paper in other contexts and to further refine survey designs meant to assess social network structure. I believe that the research mentioned in the previous paragraph will be a significant contribution to the former suggestion. As for the second, there has been tremendous progress toward survey development for use in social network analysis, as evidenced by the Malawi dataset. However, there remains little consensus or commonality regarding survey structure in the literature.[21] The results of this study indicate that random matching surveys ought to be revised in some manner to reflect non-local links. The systematic exclusion of non-local individuals in random matching surveys make the practice irreconcilable with true network formation decisions if there is indeed a selection problem that characterizes cross-boundary linkages. With more research and resources devoted to understanding underlying propensities to link, accounting for all links within and between villages or traditional authorities, it becomes necessary to harmonize practices to some extent.

[21] See Table 1 in Santos and Barrett (2008) for a comparison of sampling techniques used in social network analysis.

References

Ackerberg, Daniel A, and Maristella Botticini. 2002. "Endogenous Matching and the Empirical Determinants of Contract Form." *Journal of Political Economy* 110 (3): 564–91. doi:10.1086/339712.

Aggarwal, Rimjhim M. 2007. "Role of Risk Sharing and Transaction Costs in Contract Choice: Theory and Evidence From Groundwater Contracts." *Journal of Economic Behavior and Organization* 63 (3): 475–96. doi:10.1016/j.jebo.2005.06.010.

Allen, Hugh, and Mark Staehle. 2007. "Programme Guide: Field Operations Manual." doi:10.1109/CICEM.2013.6820130.

Bloch, F, G Genicot, and D Ray. 2008. "Social Networks and Informal Insurance." *Journal of Economic Theory* 143 (1): 36–58. http://neumann.hec.ca/neudc2004/fp/genicot_garance_aout_31.pdf.

Breiger, Ronald, Kathleen Carley, and Philippa Pattison. 2003. *Dynamic Social Network Modeling and Analysis: Workshop Summary Papers*. http://www.nap.edu/catalog.php?record_id=10735.

Casciaro, Tiziana, and Miguel Sousa Lobo. 2008. "When Competence Is Irrelevant: The Role of Interpersonal Affect in Task-Related Ties." *Administrative Science Quarterly* 53 (4): 655–84. doi:10.2189/asqu.53.4.655.

Cassidy, Rachel, and Marcel Fafchamps. 2015. "Can Community-Based Microfinance Groups Match Savers With Borrowers? Evidence from Rural Malawi." WPS/2015-13. CSAE Working Paper. Oxford, UK.

Chuang, Yating, and Laura Schechter. 2015. "Social Networks in Developing Countries." *Annual Review of Resource Economics* 7: 1–23.

De Weerdt, Joachim, and Stefan Dercon. 2002. "Risk-Sharing and Endogenous Network Formation." *World Institute for Development Economics Discussion Papers*. Leuven. doi:10.1093/0199276838.003.0011.

Durlauf, Steven N., and Marcel Fafchamps. 2005. "Social Capital." In *Handbook of Economic Growth*, edited by Philippe Aghion and Steven N. Durlauf, 1:1640–99. Elsevier Masson SAS. doi:10.4324/9780203868850.

Fafchamps, Marcel. 2004. *Market Institutions in Sub-Saharan Africa : Theory and Evidence*. Cambridge, Mass: MIT Press.

Fafchamps, Marcel, and Flore Gubert. 2007. "The Formation of Risk Sharing Networks." *Journal of Development Economics* 83 (2): 326–50. doi:10.1016/j.jdeveco.2006.05.005.

Goldstein, Markus, and Christopher Udry. 1999. "Agricultural Innovation and Resource Management in Ghana." https://www.academia.edu/2411701/Agricultural_Innovation_and_Resource_Management_in_Ghana_Final_Report_to_IFPRI_under_MP17.

Harrison, Glenn W., and E. Elisabet Rutström. 2008. "Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods." In *Handbook of Experimental Economics Results*, edited by Charles R. Plott and Vernon L. Smith, 1st ed., 1:752–67. doi:10.1016/S1574-0722(07)00081-9.

Jackson, Mo. 2007. "The Study of Social Networks in Economics." In *Formation and Decay of Economic Networks*, edited by James E. Rauch, 1–25. New York: Russel Sage Foundation. http://www.stanford.edu/~jacksonm/netsocialecon.pdf.

Jaimovich, Dany. 2013. "Missing Links, Missing Markets: Internal Exchanges, Reciprocity and External Connections in the Economic Networks of Gambian Villages." 44080. MPRA Working Paper. Munich.

Kossinets, Gueorgi. 2006. "Effects of Missing Data in Social Networks." *Social Networks* 28 (3): 247–68. doi:10.1016/j.socnet.2005.07.002.

Krackhardt, David. 1988. "Predicting with Networks: Nonparametric Multiple Regression Analysis of Dyadic Data." *Social Networks*.

Ksoll, Christopher, Helene Bie Lilleør, Jonas Helth Lønborg, and Ole Dahl Rasmussen. 2013. "Impact of Village Savings and Loans Associations: Evidence from a Cluster Randomized Trial." 56. Study Papers.

Lilleør, Helene, Chris Ksoll, Jonas Helth Lonborg, and Ole Dahl Rasmussen. 2012. "User's Guide to the Karonga Assessment of Vulnerability Datasets." Rockwool Foundation.

Lønborg, Jonas Helth, and Ole Dahl Rasmussen. 2014. "Can Microfinance Reach the Poorest: Evidence from a Community-Managed Microfinance Intervention." *World Development* 64: 460–72. doi:10.1016/j.worlddev.2014.06.021.

Loomis, John B. 2014. "2013 WAEA Keynote Address: Strategies for Overcoming Hypothetical Bias in Stated Preference Surveys." *Journal of Agricultural and Resource Economics* 39 (1): 34–46.

Manski, Charles F. 2000. "Economic Analysis of Social Interactions." *Journal of Economic Perspectives* 14 (3): 115–36. doi:10.1126/science.151.3712.867-a.

Manski, Charles F. 1993. "Identification of Social Endogenous Effects: The Reflection Problem." *The Review of Economic Studies* 60 (3): 531–42. doi:10.2307/2298123.

McPeak, John. 2004. “Contrasting Income Shocks With Asset Shocks: Livestock Sales in Northern Kenya.” *Oxford Economic Papers* 56 (2): 263–84. doi:10.1093/oep/gpf040.

McPherson, Miller. 2001. “Sampling Strategies for the Arts: A Hypernetwork Approach.” In *Poetics*, 28:291–306. doi:10.1016/S0304-422X(01)80005-X.

Mincer, Jacob. 1978. “Family Migration Decisions.” *The Journal of Political Economy* 86 (5): 749–73.

Rosenzweig, Mark R. 1988. “Risk, Implicit Contracts and the Family in Rural Areas of Low-Income Countries.” *The Economic Journal* 98 (393): 1148–70. http://resolver.scholarsportal.info/resolve/00130133/v98i0393/1148_ricatfiraolc.xml.

Santos, Paulo, and Christopher B Barrett. 2006a. “Informal Insurance in the Presence of Poverty Traps : Evidence from Southern Ethiopia.” Mimeo.

Santos, Paulo, and Christopher B Barrett. 2006b. “Why and How to Sample Social Networks.” 211. Cornell Food and Nutrition Policy Program.

Santos, Paulo, and Christopher B Barrett. 2007. “Understanding the Formation of Social Networks.” Mimeo.

Santos, Paulo, and Christopher B Barrett. 2008. “What Do We Learn About Social Networks When We Only Sample Individuals? Not Much.” Mimeo. doi:10.2139/ssrn.1141838.

Santos, Paulo, and Christopher B Barrett. 2010. “Identity, Interest and Information Search in a Dynamic Rural Economy.” *World Development* 38 (12): 1788–96. doi:10.1016/j.worlddev.2010.04.003.

Santos, Paulo, and Christopher B Barrett. 2011. “Persistent Poverty and Informal Credit.” *Journal of Development Economics* 96 (2). Elsevier B.V.: 337–47. doi:10.1016/j.jdeveco.2010.08.017.

Udry, Christopher. 1990. “Credit Markets in Northern Nigeria: Credit as Insurance in a Rural Economy.” *The World Bank Economic Review* 4 (3): 251–69. http://resolver.scholarsportal.info/resolve/02586770/v04i0003/251_cminncaiiare.xml.

Udry, Christopher, and Timothy Conley. 2004. “Social Networks in Ghana.” 888. Economic Growth Center Working Papers. doi:10.4324/9780203799659.

Woolcock, Michael, and Deepa Narayan. 2000. “Social Capital : Implications for Development Theory, Research, and Policy.” *The World Bank Research Observer* 15 (2): 225–49.

Variable	Mean	S.D.	Minimum Value	Maximum Value
OLS Models				
Net transfers with match (MK)	153.90	2579.88	-90000	98500
Hypothetical net transfers with match (MK)	-374.00	838.59	-24000	4500
Probit Models				
Does (…) live in another village?	0.57	0.49	0	1
Does (...) live in another T/A?[1]	0.45	0.50	0	1

Notes:
[1] *T/A refers to the traditional authority of Mwirang'ombe that provided the source data for this study*

Table 1: *Descriptive Statistics of Dependent Variables (Link Level)*

Variable	Mean	S.D.	Minimum Value	Maximum Value
Household (HH) Variables				
Does the HH use modern lighting?	0.34	0.47	0	1
Number of HH members at time of interview	5.96	2.44	1	18
Age of HH head	42.98	14.26	2	87
Share of adults in HH who can read	0.74	0.34	0	1
Proportion of HH members without education excluding head	0.03	0.07	0	0.5
HH head years of education	6.87	3.10	0	19
HH is female-headed	0.13	0.33	0	1
Size of house (number of rooms)	3.16	1.24	0	7
Land ownership (acres)	3.35	2.58	0	25
Match Variables				
Is (...) male?	0.44	0.50	0	1
Does (…) live in another village?	0.57	0.49	0	1
Does (...) live in another T/A?[1]	0.45	0.50	0	1
Is (…) a family member?	0.88	0.33	0	1
How does (...)'s income compare to respondent's?[2]	2.65	1.13	1	5

Notes:
[1] *T/A refers to the traditional authority of Mwirang'ombe that provided the source data for this study*
[2] *This is a categorical variable with five categories specifying the compared status and the extent thereof. Equal economic status is used as the variable's base for regressions.*

Table 2: *Descriptive Statistics of Independent Variables (Link Level)*

Variable	Mean	S.D.	Minimum Value	Maximum Value
Number of roster matches ouside village	4.30	2.41	0	14
Number of roster matches ouside T/A[1]	3.30	2.27	0	14
Net HH transfers with match (MK)	1492.20	9934.87	-164000	119500
Hypothetical net HH transfers with match (MK)	-3610.40	4886.43	-33000	4000

Notes:
[1] *T/A refers to the traditional authority of Mwirang'ombe that provided the source data for this study*

Table 3: *Descriptive Statistics of Dependent Variables (Household Level)*

Variable	Mean	S.D.	Minimum Value	Maximum Value
Household (HH) Variables				
Number of HH members at time of interview	5.91	2.43	1	18
Age of HH head	42.44	14.17	2	87
Share of adults in HH who can read	0.74	0.34	0	1
Number of HH members without education excluding head	0.03	0.07	0	0.5
HH head years of education	6.81	3.15	0	19
HH is female-headed	0.14	0.35	0	1
Size of house (number of rooms)	3.11	1.24	0	7
Land ownership (acres)	3.28	2.56	0	25
Match Variables				
Ratio of family to non-family in roster	0.81	0.19	0	1
Number of roster matches ouside village	2.61	2.03	0	12
Number of roster matches ouside T/A[1]	1.95	1.82	0	11
Fraction of matches with higher economic status	0.38	0.38	0	1
Fraction of matches with lower economic status	0.1	0.24	0	1

Notes:
[1] *T/A refers to the traditional authority of Mwirang'ombe that provided the source data for this study*

Table 4: *Descriptive Statistics of Independent Variables (Household Level)*

	(1)	(2)	(3)	(4)	(5)	(6)
		Real Net Transfers			*Hypothetical Net Transfers*	
Household (HH) Variables						
HH head years of education	8.386	8.459	17.77	-5.397	-5.398	-4.059
	(11.64)	(11.62)	(11.19)	(4.534)	(4.529)	(4.522)
Age of HH head	11.14***	11.29***	12.06***	1.814*	1.819*	0.535
	(3.008)	(2.987)	(2.862)	(1.026)	(1.023)	(0.967)
HH is female-headed	386.3***	381.1***	276.4***	81.19**	80.36**	76.79*
	(92.75)	(92.86)	(85.18)	(39.46)	(39.42)	(39.36)
Land ownership (acres)	19.66	19.60		-21.08***	-21.13***	
	(14.59)	(14.48)		(5.454)	(5.444)	
Does the HH use modern lighting?	17.31	22.72	81.02	25.66	26.28	40.50
	(54.95)	(54.64)	(54.40)	(27.01)	(26.96)	(26.72)
Match Variables						
Does (...) live in another village?	300.3***		220.0***	35.69*		11.38
	(55.32)		(50.90)	(19.71)		(19.62)
Does (...) live in another T/A?[1]		298.8***			41.92**	
		(59.50)			(19.68)	
I am much poorer than (...)			1,098***			249.0***
			(109.4)			(29.89)
I am slightly poorer than (...)			414.0***			149.8***
			(44.22)			(22.04)
I am slightly richer than (...)			-462.5***			-153.4***
			(59.90)			(26.45)
I am much richer than (...)			-1,245***			-344.5***
			(187.8)			(45.02)
Is (...) male?	406.4***	403.2***	263.1***	100.1***	99.74***	61.25***
	(55.83)	(55.82)	(54.44)	(16.08)	(16.05)	(15.90)
Is (...) a family member?	-191.9***	-167.6***	-145.5***	-73.66***	-73.68***	-61.67**
	(45.42)	(44.88)	(44.01)	(25.15)	(23.93)	(25.30)
Constant	-637.1***	-624.0***	-773.6***	-459.8***	-457.8***	-516.5***
	(186.5)	(185.1)	(180.9)	(61.27)	(61.13)	(61.60)
Observations	16,857	16,857	16,697	16,857	16,857	16,697
R-squared	0.013	0.013	0.052	0.009	0.009	0.037

Notes:
i) Robust standard errors in parentheses
ii) ***, **, * *denote statistical significance at the 1, 5, 10 percent levels, respectively.*
[1] *T/A refers to the traditional authority of Mwirang'ombe that provided the source data for this study*

Table 5: *Net Transfer Regressions (Link Level)*

	(1)	(2)	(3)	(4)
	Does (…) live in another		*Does (…) live in another T/A[1]?*	
Household (HH) Variables				
HH head years of education	0.00120	0.00232	0.000424	0.00167
	(0.00253)	(0.00255)	(0.00278)	(0.00278)
Age of HH head	0.00246***	0.00280***	0.00196***	0.00230***
	(0.000606)	(0.000594)	(0.000657)	(0.000633)
Land ownership (acres)	0.00567*		0.00540	
	(0.00322)		(0.00348)	
Does the HH use modern lighting?	0.0219**	0.0236**	0.00487	0.00796
	(0.0105)	(0.0107)	(0.0122)	(0.0121)
Match Variables				
I am much poorer than (…)		0.132***		0.165***
		(0.0148)		(0.0154)
I am slightly poorer than (…)		0.0821***		0.0929***
		(0.0127)		(0.0126)
I am slightly richer than (…)		0.0327**		0.0259*
		(0.0134)		(0.0139)
I am much richer than (…)		0.0617***		0.0462**
		(0.0210)		(0.0231)
Is (...) male?	-0.00645	-0.0170	0.00396	-0.00980
	(0.0118)	(0.0118)	(0.0112)	(0.0112)
Is (…) a family member?	0.533***	0.529***	0.456***	0.453***
	(0.0154)	(0.0156)	(0.0129)	(0.0129)
Observations	16,857	16,697	16,857	16,697
Pseudo R-squared	0.1226	0.1300	0.1013	0.1122

Notes:
i) Robust standard errors in parentheses
ii) ***, **, * *denote statistical significance at the 1, 5, 10 percent levels, respectively.*
[1] *T/A refers to the traditional authority of Mwirang'ombe that provided the source data for this study*

Table 6: *Marginal Logit Regressions (Link Level)*

	(1)	(2)	(3)	(4)	(5)	(6)
	Net household transfers with match (MK)			*Hypothetical net household transfers with match (MK)*		
Household (HH) Variables						
Land ownership (acres)	163.2	161.3		-155.7***	-168.0***	
	(112.0)	(110.4)		(45.99)	(47.71)	
Age of HH head	85.45***	85.27***	88.46***	27.04***	26.28***	16.90**
	(21.28)	(21.15)	(20.64)	(7.767)	(8.063)	(7.334)
HH is female-headed	2,926***	2,898***	2,186***	783.0***	879.0***	800.0***
	(674.1)	(673.6)	(616.7)	(291.9)	(302.4)	(287.7)
HH head years of education	73.07	72.12	161.0*	-60.40*	-72.17*	-43.17
	(86.55)	(85.88)	(84.65)	(35.54)	(36.91)	(35.73)
Match Variables						
Number of roster matches outside village	103.3		167.4	-397.3***		-396.2***
	(110.5)		(105.8)	(55.76)		(54.87)
Number of roster matches outside T/A[1]		135.6			-250.8***	
		(127.8)			(54.27)	
Ratio of family to non-family in roster	1,369	1,341	1,004	932.9*	504.8	880.4
	(1,055)	(1,033)	(1,036)	(545.8)	(546.7)	(553.2)
Fraction of matches with higher economic status			4,319***			1,138***
			(577.1)			(273.2)
Fraction of matches with lower economic status			-4,797***			-1,845***
			(1,583)			(523.8)
Constant	-5,148***	-5,104***	-6,376***	-3,009***	-3,405***	-3,423***
	(1,710)	(1,714)	(1,671)	(640.3)	(665.3)	(620.5)
Observations	2,327	2,327	2,325	2,327	2,327	2,325
R-squared	0.030	0.031	0.082	0.061	0.038	0.077

Notes:
i) Robust standard errors in parentheses
ii) ***, **, * *denote statistical significance at the 1, 5, 10 percent levels, respectively.*
[1] *T/A refers to the traditional authority of Mwirang'ombe that provided the source data for this study*

Table 7: *Net Transfer Regressions (Household Level)*

	(1) *Number of of roster matches outside village*	(2) *Number of of roster matches outside T/A* [1]
Land ownership (acres)	0.0763***	0.0718**
	(0.0287)	(0.0303)
HH head years of education	0.0656***	0.0569**
	(0.0244)	(0.0238)
Age of HH head	0.00530	0.00537
	(0.00558)	(0.00540)
HH is female-headed	-0.212	0.0468
	(0.228)	(0.212)
Ratio of family to non-family in roster	2.330***	1.984***
	(0.345)	(0.320)
Constant	1.523***	0.833**
	(0.413)	(0.401)
Observations	2,327	2,327
R-squared	0.053	0.041

Notes:
i) Robust standard errors in parentheses
ii) ***, **, * *denote statistical significance at the 1, 5, 10 percent levels, respectively.*
[1] *T/A refers to the traditional authority of Mwirang'ombe that provided the source data for this*

Table 8: *Network Density Regressions (Household Level)*

	(1)	(2)
	Net Transfers	
	Non-local = Out of Village	Non-local = Out of T/A[1]
Base Variables		
Is (...) non-local?	192.3	175.3
	(122.2)	(141.3)
HH head years of education	-34.89***	-16.40
	(10.45)	(10.04)
Age of HH head	4.341*	6.719***
	(2.563)	(2.452)
HH is female-headed	270.8**	385.9***
	(117.3)	(111.9)
Land ownership (acres)	-1.345	-9.960
	(19.07)	(15.16)
Does the HH use modern lighting?	33.30	37.27
	(68.40)	(60.91)
Is (...) male?	352.1***	388.5***
	(74.78)	(68.58)
Is (...) a family member?	-188.0***	-173.0***
	(49.33)	(45.66)
Interaction Variables		
Is (...) non-local? x HH head years of education	72.58***	50.62**
	(19.97)	(22.83)
Is (...) non-local? x Age of HH head	11.48**	9.840*
	(5.024)	(5.806)
Is (...) non-local? x HH is female-headed	193.6	-13.60
	(156.0)	(157.9)
Is (...) non-local? x Land ownership (acres)	31.36	62.43**
	(31.22)	(29.96)
Is (...) non-local? x Does the HH use modern lighting?	-30.86	-41.97
	(100.3)	(107.2)
Is (...) non-local? x Is (...) male?	89.67	21.91
	(104.3)	(114.1)
Is (...) non-local? x Is (...) a family member?	63.90	130.9
	(129.9)	(142.9)
Constant	-30.56	-29.63
	(33.51)	(31.71)
Observations	16,857	16,857
R-squared	0.015	0.015
Joint Significance of the Local Dummy and Interactions[2]		
F(7, 826)	2.30	1.53
Prob > F	0.0249	0.1533

Notes:
i) Robust standard errors in parentheses
ii) ****, **, * denote statistical significance at the 1, 5, 10 percent levels, respectively.*
iii) All regressions in Table 5 were tested with interactions and yielded no markedly different results.
[1] *T/A refers to the traditional authority of Mwirang'ombe that provided the source data for this study*
[2] *Joint test of whether local and non-local links follow the same econometric model*

Table 9: *Evidence of Bias*

Appendix A: Village Cluster Regressions (Robustness Check)

	(1)	(2)	(3)	(4)	(5)	(6)
		Real Net Transfers			*Hypothetical Net Transfers*	
Household (HH) Variables						
HH head years of education	8.386	8.459	17.77	-5.397	-5.398	-4.059
	(12.07)	(12.08)	(11.96)	(4.291)	(4.283)	(4.257)
Age of HH head	11.14***	11.29***	12.06***	1.814*	1.819*	0.535
	(3.371)	(3.326)	(3.301)	(1.077)	(1.069)	(1.015)
HH is female-headed	386.3***	381.1***	276.4***	81.19**	80.36**	76.79**
	(83.25)	(82.91)	(74.84)	(37.68)	(37.60)	(36.79)
Land ownership (acres)	19.66	19.60		-21.08***	-21.13***	
	(13.68)	(13.61)		(4.967)	(4.969)	
Does the HH use modern lighting?	17.31	22.72	81.02	25.66	26.28	40.50
	(57.02)	(57.22)	(58.66)	(29.96)	(29.91)	(29.47)
Match Variables						
Does (…) live in another village?	300.3***		220.0***	35.69*		11.38
	(61.92)		(55.63)	(19.38)		(19.27)
Does (…) live in another T/A?[1]		298.8***			41.92**	
		(65.88)			(20.46)	
I am much poorer than (…)			1,098***			249.0***
			(110.6)			(33.97)
I am slightly poorer than (…)			414.0***			149.8***
			(46.59)			(24.71)
I am slightly richer than (…)			-462.5***			-153.4***
			(49.64)			(25.04)
I am much richer than (…)			-1,245***			-344.5***
			(170.7)			(52.05)
Is (…) male?	406.4***	403.2***	263.1***	100.1***	99.74***	61.25***
	(59.61)	(59.69)	(56.23)	(18.85)	(18.77)	(17.52)
Is (…) a family member?	-191.9***	-167.6***	-145.5***	-73.66***	-73.68***	-61.67**
	(50.59)	(50.90)	(52.12)	(24.36)	(21.24)	(25.01)
Constant	-637.1***	-624.0***	-773.6***	-459.8***	-457.8***	-516.5***
	(217.4)	(213.7)	(208.8)	(60.40)	(59.82)	(59.08)
Observations	16,857	16,857	16,697	16,857	16,857	16,697
R-squared	0.013	0.013	0.052	0.009	0.009	0.037

Notes:
i) Robust standard errors in parentheses
ii) ***, **, * *denote statistical significance at the 1, 5, 10 percent levels, respectively.*
[1] *T/A refers to the traditional authority of Mwirang'ombe that provided the source data for this study*

Table 10: *Net Transfer Regressions - Village Cluster (Link Level)*

	(1)	(2)	(3)	(4)
	Does (...) live in another		*Does (...) live in another T/A [1] ?*	
<u>Household (HH) Variables</u>				
HH head years of education	0.00120	0.00232	0.000424	0.00167
	(0.00269)	(0.00269)	(0.00295)	(0.00292)
Age of HH head	0.00246***	0.00280***	0.00196***	0.00230***
	(0.000617)	(0.000597)	(0.000666)	(0.000658)
Land ownership (acres)	0.00567*		0.00540	
	(0.00327)		(0.00360)	
Does the HH use modern lighting?	0.0219**	0.0236**	0.00487	0.00796
	(0.00973)	(0.00978)	(0.0133)	(0.0131)
<u>Match Variables</u>				
I am much poorer than (...)		0.132***		0.165***
		(0.0168)		(0.0141)
I am slightly poorer than (...)		0.0821***		0.0929***
		(0.0139)		(0.0121)
I am slightly richer than (...)		0.0327**		0.0259**
		(0.0133)		(0.0123)
I am much richer than (...)		0.0617***		0.0462**
		(0.0212)		(0.0215)
Is (...) male?	-0.00645	-0.0170	0.00396	-0.00980
	(0.0128)	(0.0129)	(0.0135)	(0.0136)
Is (…) a family member?	0.533***	0.529***	0.456***	0.453***
	(0.0145)	(0.0152)	(0.0135)	(0.0139)
Observations	16,857	16,697	16,857	16,697
Pseudo R-squared	0.1230	0.1304	0.1014	0.1123

Notes:
i) Robust standard errors in parentheses
ii) ***, **, * *denote statistical significance at the 1, 5, 10 percent levels, respectively.*
[1] *T/A refers to the traditional authority of Mwirang'ombe that provided the source data for this study*

Table 11: *Marginal Logit Regressions – Village Cluster (Link Level)*

	(1)	(2)	(3)	(4)	(5)	(6)
	Net household transfers with match (MK)			*Hypothetical net household transfers with match (MK)*		
Household (HH) Variables						
Land ownership (acres)	163.2	161.3		-155.7***	-168.0***	
	(114.4)	(112.0)		(42.21)	(45.02)	
Age of HH head	85.45***	85.27***	88.46***	27.04***	26.28***	16.90**
	(24.23)	(23.96)	(23.71)	(8.468)	(8.632)	(8.268)
HH is female-headed	2,926***	2,898***	2,186***	783.0***	879.0***	800.0***
	(615.7)	(616.5)	(558.1)	(259.8)	(272.3)	(252.1)
HH head years of education	73.07	72.12	161.0*	-60.40*	-72.17**	-43.17
	(91.50)	(91.18)	(92.79)	(30.53)	(30.94)	(31.05)
Match Variables						
Number of roster matches outside village	103.3		167.4	-397.3***		-396.2***
	(125.4)		(120.6)	(54.85)		(56.53)
Number of roster matches outside T/A[1]		135.6			-250.8***	
		(135.7)			(46.00)	
Ratio of family to non-family in roster	1,369	1,341	1,004	932.9*	504.8	880.4*
	(994.7)	(987.3)	(980.6)	(478.5)	(491.3)	(500.9)
Fraction of matches with higher economic status			4,319***			1,138***
			(617.5)			(325.2)
Fraction of matches with lower economic status			-4,797***			-1,845***
			(1,492)			(468.1)
Constant	-5,148**	-5,104**	-6,376***	-3,009***	-3,405***	-3,423***
	(2,025)	(2,005)	(1,944)	(625.6)	(632.7)	(587.9)
Observations	2,327	2,327	2,325	2,327	2,327	2,325
R-squared	0.030	0.031	0.082	0.061	0.038	0.077

Notes:
i) Robust standard errors in parentheses
ii) ****, **, * denote statistical significance at the 1, 5, 10 percent levels, respectively.*
[1] *T/A refers to the traditional authority of Mwirang'ombe that provided the source data for this study*

Table 12: *Net Transfer Regressions – Village Cluster (Household Level)*

	(1) *Number of of roster matches outside village*	(2) *Number of of roster matches outside T/A* [1]
Land ownership (acres)	0.0763***	0.0718***
	(0.0251)	(0.0255)
HH head years of education	0.0656***	0.0569**
	(0.0236)	(0.0243)
Age of HH head	0.00530	0.00537
	(0.00558)	(0.00539)
HH is female-headed	-0.212	0.0468
	(0.233)	(0.211)
Ratio of family to non-family in roster	2.330***	1.984***
	(0.399)	(0.338)
Constant	1.523***	0.833*
	(0.410)	(0.452)
Observations	2,327	2,327
R-squared	0.053	0.041

Notes:
i) Robust standard errors in parentheses
ii) ***, **, * *denote statistical significance at the 1, 5, 10 percent levels, respectively.*
[1] *T/A refers to the traditional authority of Mwirang'ombe that provided the source data for this study*

Table 13: *Network Density Regressions – Village Cluster (Household Level)*

Appendix B: Survey Forms

What follows is the set of questions asked of respondents in the household and transfer network roster surveys that were of direct relevance to this study. The variables considered relevant are given in tables 1-4.

Preliminary Questions - Network Roster and Transfers Questionnaire

i. Who are your parents?
ii. Who are the parents of the head of the household?
iii. Do you have any children who are living in a different household?
iv. Are there any other people that you can turn to for monetary transfers or in-kind assistance
v. Are there any people who can turn to you for monetary transfers or in-kind assistance?
vi. Do you or the head have any siblings? Could you turn to them?
vii. Do you or the head have any uncles or aunts? Could you turn to them?
viii. What about cousins, friends, neighbours
ix. If you needed 3000 Kwacha, are there any other people you could turn to?
x. Is (…) still alive?

Survey Questions Used in Regressions - Network Roster and Transfers Questionnaire

Variable (as given in text)	Survey Question
Does (…) live in another village?	**Where** does or did (…) live?
Does (...) live in another traditional authority?	**Where** does or did (…) live?
Is (...) male?	**What** is/was the gender of (…)?
Is (…) a family member?	**How** is (…) related to you?
How does (...)'s income compare to respondent's?	**In** general, how do you compare the household of (...) with your own in terms of economic status?
Net transfers with match (MK)	**What** was the total value of the money or in-kind assistance that (...) gave you?
	What was the total value of the money or in-kind assistance that you gave (...)?
Hypothetical net transfers with match (MK)	**If** you needed 3000 Kwacha because of an illness or a very bad year. How much would you ask from each of these persons if they all had above average income/harvests? **Now** imagine that (...) needed 3000 Kwacha. Think about the people he might ask. How much do you think he/she would ask you for if he needed 3000 Kwacha and you and all people he can turn to had above average income/harvest?

Survey Questions Used in Regressions - Household Questionnaire

Variable (as given in text)	**Survey Question**
Does the household use modern lighting?	**What** is your main source of lighting?
Number of household members at time of interview	**Same** as variable
Age of household head	**Same** as variable
Share of adults in household who can read	**Same** as variable
Proportion of household members without education excluding head	**Same** as variable
Household head years of education	**Same** as variable
Household is female-headed	**Same** as variable
Size of house (number of rooms)	**Same** as variable
Land ownership (acres)	**Same** as variable