DEPARTMENT OF ECONOMICS
UNIVERSITY OF COPENHAGEN

# PhD Thesis

Benedikte Alkjærsig Bjerge

# Essays on social network formation in sub-Saharan Africa

Academic Advisors: Finn Tarp and John Rand
Submitted: April, 30 2014

# Contents

# Acknowledgements

This thesis is the result of my doctoral studies at the Department of Economics, University of Copenhagen. I would like to thank numerous people who in different ways have contributed to my work during the past three years.

My sincere gratitude goes to Finn Tarp, my principal advisor, for encouraging me to enroll in the PhD program, guidance and contineuous support throughout the entire process. Finn has been excellent at introducing me to the international research community. In particular, for making my visit to University of Oxford possible.

I am indebted to John Rand, my secondary advisor, for this competent guidance, friendship and support during all stages of my studies. John Rand's analytical competency and broad knowledge have been an immense source of inspiration. Thanks also to Ulrik Richardt Beck, who have been my co-author and good friend during the last three years, and for extremely enjoyable and productive collaboration from which I have learnt a lot. Additionally, I have gained valuable knowledge for this thesis through regular discussions with Nina Torm, Sam Jones, Henrik Hansen and others at the Department of Economics, University of Copenhagen, not to mention the members of the Develoment Economic Research Group (DERG), Christel Brink Hansen and the IT staff.

In the academic year 2012 I had the wonderful opportunity to visit University of Oxford. It was a highly rewarding experience, both academically and personally. I would like to thank Marcel Fafchamps for his time and discussions of my work. Sincere thanks also to the doctoral students and my fellow visiting researchers for interesing discussions and many fun experiences. In relation to my visit to University of Oxford I gratefully acknowledge the financial support I received from Oticon Fonden, Augustinus Fonden and Knud Højgaards Fond.

My parents and brother; thank you for invaluable optimism, and for backing me in everything I put my mind to. Finally, I would like to thank my boyfreind - Tobias Carøe Siemenge - for loving support and endless encouragement throughout my studies. Thank you!

Benedikte Alkjærsig Bjerge

Copenhagen, April 2014

# Summary (English)

The focus of this thesis is social network formation in a development economic context. The main objective is to achieve a better understanding of how networks are formed and why they might have certain characteristics. These related objectives are addressed empirically from different angles using different methodological approaches. Three of the chapters investigate social network formation based on strategic network formation theory, while a fourth chapter takes a random graph approach using exponential random graph models.

Based on a unique social network dataset consisting of households in rural Gambia, the first two chapters investigate mechanism underlying inter-household land transactions. The first chapter documents the importance of pre-existing social networks in terms of social and geographical proximity, and show that only geographical proximity has an additional impact on allocative efficiency. The second chapter, documents that land is allocated in a pro-poor way consistent with the presence of the norm-based access rule to vital resources. However, land allocation come with an obligation to reciprocate in the labor market, and poor households are only allocated land in relatively less population dense and more ethnic homogenous villages. The third chapter is based on the same dataset from rural Gambia, but examines how the network of land, labor and agricultural input transactions is formed. Findings suggest that structural mechanisms in the form of reciprocity and transitivity explain network formation to a greater extent than standard household attributes. The chapter further demonstrates that inability to account for structural mechanisms leads to upward biased parameter estimates of household attributes and dyad-specific characteristics. The fourth chapter is based on a different dataset concerned with micro, small and medium sized enterprises in Mozambique. The chapter establishes presence of positive assortative matching on co-membership in a business association, and find limited evidence supporting diffusion of business practices between co-members. This finding is consistent with the large heterogeneity across firms and slow convergence of productivity observed both across and within sectors in sub-Saharan Africa.

# Resumé (Danish)

Denne afhandling omhandler social netværksformation i en udviklingsøkonomisk kontekst. Hovedformålet er at opnå en bedre forståelse for, hvordan netværk dannes, og for de underliggende mekanismer som driver dem. Disse relaterede målsætninger er behandlet empirisk og belyst fra forskellige vinkler ved brug af to relaterede metoder indenfor litteraturen om netværksformation. Tre af kapitlerne fokuserer på de underliggende mekanismer bag tilblivelsen af sociale netværk, og hvorledes de kan bidrage til at forstå og forklare fordelingen af knappe ressourcer. Det fjerde kapitel fokuserer på, hvordan sociale netværksstrukturer etableres, uafhængig af økonomisk adfærd.

Ved brug af et socialt netværks datasæt omfattende husholdninger bosiddende i landdistrikterne i Gambia, undersøger de første to kapitler mekanismerne bag transaktioner af landbrugsjord mellem husholdninger. Det første kapitel dokumenterer vigtigheden af allerede eksisterende sociale netværk indenfor social og geografisk nærhed og viser, at det udelukkende er geografisk nærhed, som har en mereffekt på den efficiente allokering af landbrugsjord. Det andet kapitel dokumenterer, at land allokeres med et fattigdomsorienteret sigte i relativt mindre tætbefolkede og i mere etnisk homogene landsbyer. Dette er i overensstemmelse med tilstedeværelsen af en normbaseret regel om, at alle skal have adgang til vitale ressourcer såsom jord. Landtildeling kommer imidlertid med en forpligtelse om en modydelse i form af arbejdskraft. Det tredje kapitel er baseret på det samme datasæt men undersøger, hvordan netværk dannes. Resultaterne viser at endogene strukturelle mekanismer i højere grad forklarer tilblivelsen af netværk sammenlignet med traditionelle husholdningskarakteristika. Kapitlet demonstrerer ydermere, at udeladelsen af strukturelle mekanismer i økonomiske analyser skaber systematiske fejl for betydningen af husholdnings- samt links-specifikke karakteristika. Det fjerde papir er baseret på et andet men tilsvarende datasæt omfattende små og mellemstore virksomheder i Mozambique. Kapitlet fastslår tilstedeværelsen af positiv selektion i virksomheders medlemskab i samme erhvervsorganisation. I overensstemmelse med den fraværende konvergens i produktivitet både på tværs samt indenfor virksomheder i Afrika syd for Sahara finder kapitlet ikke grundlag for at konkludere, at vidensdeling om forretningsmetoder og innovationer finder sted blandt medlemmer af samme erhvervsorganisation.

# Chapter 1

# Introduction to social network formation

As social capital theorist have stated for some time, and economists have recognized more recently, the social context of economic interaction is one of the possible drivers of behaviors and outcomes. The interest in social networks partly derives from the acknowledgment that a deeper understanding of social structures underlying human behavior can enrich economic models. Within the field of economics, the focus on social networks has sparked two different, though related, trends. One trend focuses on how social networks can help understand and explain allocation of scarce resources. The other focuses on the understanding of social structures independent of economic behavior. Today, there is a large and rapidly growing literature on social networks, both within economics and in other fields. This thesis contributes to the expanding body of empirical work on social networks, arguing that the traditional approach is inadequate for understanding a number of phenomena embedded in social networks.

A social network is a social structure made up of social actors, such as individuals, firms, or organizations. In the terminology of social networks, these actors are referred to as nodes, while the set of links that connect actors are referred to as edges or dyadic ties. A dyad is thus the link between a pair of nodes. For example, consider Padgett and Ansell's (1993) famous study of a network of marriages between key families in Florence in the 1940s. Here the nodes correspond to the families, and the links between the families represents marriages between members of the two families. Another example of a network, taken from this thesis, is the land market, where nodes represent households. A link between two households is established if they both agree to transact land and the value of the transaction corresponds to the amount of land transacted.

For empirical work related to social networks, it is important to distinguish between the study of network effects and the study of network formation. The study of network effects is concerned with the node-specific outcome, which may depend on the actions or/and characteristics of the nodes to which node $i$ is directly or indirectly connected. The study of network effects is generally interested in externalities. Examples of network effects include diffusion and search externalities. In contrast, the study of network formation investigates existing ties between nodes. These studies are closely related to the literature on matching processes, of which marriage and labor markets are two examples. This dissertation falls within the latter area of network formation focusing on the study of dyadic ties.

The area of network formation has been studied in multiple disciplines, including sociology, anthropology, mathematics, economics, and, more recently, the fields of statistical physics and computer science, and it is therefore a complex and multi-faceted area. The literature concerned with network formation is largely organized around two different methodologies. One has its roots in the random graph literature, and the other approach is based on economic fundamentals presuming that agents choose their relationship based on payoffs that emerge as a function of the network. This later approach is also referred to as strategic network formation. These approaches are not mutually exclusive but complement each other through their different strengths. This thesis examines network formation using both approaches.

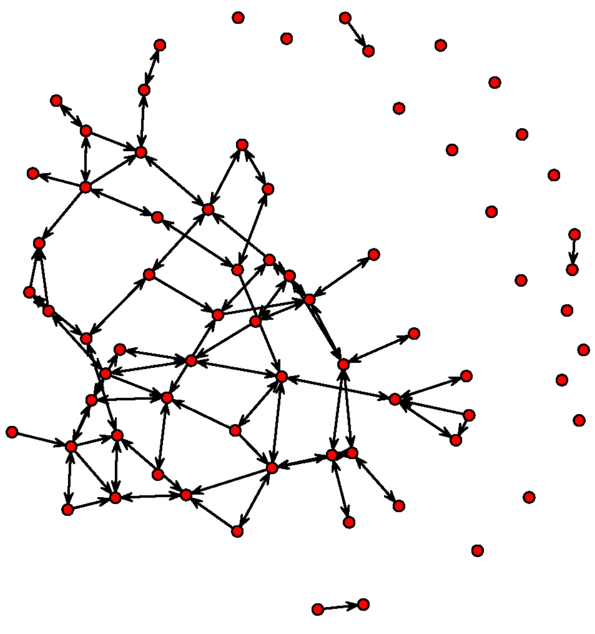
In order to start from a common level of insight into the basic concepts and features of social networks, Section 1 provides a brief introduction into these. Depending on the reader's familiarity with the concepts in social networks, this subsection may be skipped. In order to contribute to our understanding of network formation, a basic introduction in Section 2 is indispensable. Given the breadth of the literature, my aim here is not to give a comprehensive overview of the literature, but rather to provide a brief primer on network formation, organized around the literature concerned with random graph models and strategic network formation. Section 3 positions the core chapters of this thesis in the network formation literature and discusses the findings. Finally, Section 3.1 reflects on the principal contributions of this thesis and addresses the limitations and areas for future research.

# 1 Common features of social networks

Before pointing out some of the common features of social networks, a minimum level of understanding of the basic notation is in order. Let $N = \{1, 2, ..., n\}$ denote a set of nodes which represent the social actors who might be tied into a network of social relationships. A network $g$ can be represented by an $n \times n$ matrix taking the value 1 or 0. If $g_{ij} = 1$, then nodes $i$ and $j$ are said to be linked or that a tie exists between them. In some applications, these links are *undirected*, meaning that the area above the diagonal in the network matrix is identical to the area below the diagonal. Formally: $g_{ij} = g_{ji}$. An example of an undirected network is when two actors are observed to have spoken on the phone. In other applications, the network can be *directed,* which implies that $g_{ij} \neq g_{ji}$. Building on the above example, we can think about a directed network representing whether it was actor $i$ or actor $j$ made the call. If $i$ called $j$, then $g_{ji} = 1$ and $g_{ij} = 0$. Nodes not connected to other nodes are called isolates. Depending on the application, the link may also carry a value representing the strength of the link or the size of a transaction.

Figure 1 illustrates a directed network for a rural village in The Gambia. The nodes represent households, and a link indicates a transaction between two households as well as the direction of the transaction. The network includes 71 households and 124 directed links. The *distance* between two households is the maximum path length connecting $i$ and $j$. For example, if $i$ is connected to $j$ and $j$ is connected to $k$, then the path length between $i$ and $k$ is 2. Hence, the distance does not refer to the geographical distance between nodes. Rather, the geographical distance between households is a link-specific characteristic and not a feature of the network itself.

Figure 1: Transaction network in rural Gambia



Note: Data come from a baseline surveyconducted by the World Bank for the purpose of evaluating a nationalCommunity Driven Development Project (CDDP).

The first basic feature is that social networks exhibit small diameters and average path lengths relative to the number of nodes in the network. The average path length refers to the average number of links needed to get from node $i$ to node $j$, while the diameter of a network is the maximum distance between any two nodes in a network. One of the most influential studies of social networks first established this small world feature Milgram (1967). Stanley Milgram studied the full chain of booklets being sent from one person in one geographical location, through intermediate acquaintances, before reaching the target person in another geographical location. Of the chains that were successful, the average number of links that a booklet took was only 5, despite the fact that the booklets would generally not have taken the shortest route from the initial sender to the target person. Disregarding isolates, the average path length in Figure 1 is also 5, and the maximum distance between two households is 14.

What tends to be a more distinguishing feature of social networks is their clustering tendency. Clustering measures what fraction of node $i$'s connections are connected to each other, leading to networks of triangles. Hence, clustering is a measure of the frequency with which transitivity (if node $i$ is linked to $j$, and $j$ is linked to $k$, then $i$ is linked to $k$) is present. The average across all nodes in the network is then the average clustering coefficient, which gives an idea of the extent to which transitivity occurs in the network. Social networks generally tend to have significantly higher clustering than what would emerge if links were generated by an independent random graph process (more on this in Section 2). The network shown above has 36 triangles, which are many more than would be expected to form by chance in a network of 71 nodes and 124 links (edges).

Another established property of social networks is their degree distribution, where the degree represents the number of links that each node has, $d_i(g)$. In a directed graph, we distinguish between the in-degree, which is the number of links leading to the node, and the out-degree, which is the number of links emanating from the node. In Figure 1 the largest out- and in-degree of any household is 5, and only four and five households, respectively, have this many links. In comparison, 16 households are defined as isolates. The degree distribution gives an idea of the variation in the number of links across different nodes. It provides information about more- and less-connected nodes compared to the degree distribution generated by a random process of link formation where all links are equally likely. However, it is important to note that the degree distribution varies considerably across different social networks. One extreme distribution corresponds to a regular network where all nodes have the same degree. Another extreme distribution corresponds to a complete graph in which each node is connected to all other nodes.

Another extensively documented aspect of social network structures is that nodes tend to be linked more frequently to other nodes with similar characteristics than to nodes with dissimilar characteristics. This is generally referred to as homophily (for a review of the many dimensions of homophily, see McPherson, Smith-Lovin and Cook, 2001). To minimize cost of forming links, individuals often choose individuals they already know through for example, school, residential area, extended family, ethnic group, etc. Empirical evidence general support that homophily based on social and geographical proximity is strong. Presence of homophily also provide an explanation for the observed tendency towards clustering. However, the tendency to observe homophily depends on the network analyzed and the level of analysis. For instance, consider transaction of land between households in an agrarian society. If transactions are altruistically motivated than we expect transactions of land to flow from land abundant towards land-poor households resulting in asymmetric relations (i.e., heterophily). However, a land abundant household may choose to transfer to a poor households that belong to the same ethnic group, and thus within the asymmetric network structure we would expect to observe homophily along ethnic groups.

## 2 Network formation

Given the impact of network structures, it is important to understand *how* networks are formed and *why* they might have certain characteristics. This chapter distinguishes between two different approaches taken in the network formation literature. These two approaches lead to complementary insights regarding networks, each of which is adapted to answer different sorts of questions. They also have different strengths and weaknesses, as pointed out below (see Jackson (2008) for a more extensive discussion).

The first approach, random network models, originates in the random graph literature. The main focus is on how specific assumptions about the random emergence of links lead to various properties of network structures, such as degree distribution and clustering. Network formation is modeled by specifying either a stochastic process where links appear at random according to a distribution, or an algorithmic process through which the links in a network are formed. What these models do is match observed characteristics back to specific processes in order to estimate which patterns and correlations appear in social network data. This approach allows one to answer questions related to *how* a network was formed by showing how observed networks at a given point in time might have resulted from some stochastic or mechanical process.[1] The basic random graph models face two main limitations. First, basic random graph models are unable to capture important features of many observed networks. In particular, the models are unable to capture the combination of relatively small diameters and high levels of clustering and degree distribution.[2] This limitation is partly due to the underlying assumption of dyad independence. Second, while these models can be used in empirical analysis of social networks, they are highly imperfect in terms of the characteristics they allow for. For instance, basic random graph models do not incorporate features to study homophily or how node characteristics influence network formation. This has led to the development of another class of models for network analysis. These models were specifically developed for empirical analysis of social networks, and I refer to these types of models here as statistical models. Two important classes of statistical models are community detection models and exponential random graph (ERG) models.

Community detection models derive from the idea that society has natural underlying "communities" that can be identified by examining social network data. These communities do not necessarily coincide with standard boundaries but impact observed behaviors. For instance, one might wish to investigate the presence of biases in hiring workers in the labor market caused by latent structures not directly related to the natural boundaries between worker skills and experience. The community structure is then detected by identifying blocks of nodes that are comparable or equivalent such that their relationships with other nodes are highly interchangeable. If nodes have similar connections, then they should belong to the same community. The derived outcome is a hierarchy of communities; however, the approach suffers greatly from subjectivity due to the lack of an exact definition of communities and a guide to how the network was formed.[3]

[1] Beyond the scope of these models is *why* one process operates in one setting while another operates in a different setting.

[2] These problems have been solved in more advanced random graph models. However, no single class of random graph models allows for all the important observed social network characteristics.

[3] This is only one branch of community detection models. For a review of different community

The underlying idea of ERG models is to express the probability that a given network arises as a function of a set of different network statistics, also referred to as network configurations. Network configurations can include the number of links in the network, the number of triads, the number of stars, and so forth. The purpose is to test for various correlation patterns in order to provide answers to questions such as whether networks with a certain pattern, such as triad closure (tendency to befriend friends of friends), are more likely to appear than networks without this pattern. For this reason, this type of model is well suited for identifying network-formation patterns but less suitable for identifying causal relationships. As stressed above, observed patterns of network clustering are often significantly higher than what would occur at random. To solve this, the ERG model incorporates richer network statistics, as compared to the basic random graph model, to govern the network-formation probability.[4] By allowing for the richer configurations, it is possible to incorporate a range of dependencies into a given network. For instance, if we were to investigate the presence of clustering, then a configuration detecting the number of triads can be included. Another feature of the ERG model is that the likelihood of links is allowed to be affected by a range of observed attributes such as socioeconomic and demographic variables. A discussion of the practical difficulties in estimating ERG model is outlined in Chapter 4.

The second approach, strategic network formation, stems from the economics literature and is linked to game theory. Strategic network formation moves the study of social networks beyond the purely descriptive state, by modeling network formation based on two key aspects. First, nodes (i.e., actors) derive utility from the network. Second, links are formed by actors, and the resulting networks can be predicted through notions of equilibrium. In other words, strategic network formation requires that agents create relations that are beneficial and drop those that are not. In contrast to the broader random graph models, the economic models are well suited to help answer questions as to why certain network features appear.

Strategic network formation models can be distinguished according to their link-formation processes: *one-sided* and *two-sided*. One-sided link processes are those in which each node can freely decide with which other node to link. This process is also often referred to as *unilateral* as it only takes one party to create a link. Examples of

detection models, see Newman (2004) and Chapter 8 in Jackson (2008).

[4]The problem in basic random graph models arises due to the assumption of independence between links.

this include hyperlinks to web pages: Your agreement is not required for someone to link his website to yours. Unilateral link-formation processes can further be divided into versions of one-way flows and two-way flows. In models with two-way flows, only one actor incurs the cost of forming the link, but both actors benefit from the link, and vice versa for one-way flows. Two-sided or *bilateral* link processes are those in which the agreement of both nodes is required in order for a link to be formed. Examples include exchange of information, agricultural input, and marriage agreements. While the standard Nash equilibrium concept can be used to detect the equilibrium in unilateral link-formation processes, it is not restrictive enough in the sense that it allows for unrealistic equilibria when the link-formation process is bilateral. For this reason other equilibrium concepts specific to the network have been proposed. The most studied is *pairwise stability*, which was first developed by Jackson and Wolinsky (1996) and has since been extended in numerous ways.

The basic principle underlying pairwise stability is that agents derive utility based on the network structure in place. Links are formed through the decision of self-interested maximizing agents who form and sever links in order to maximize their benefits, net of the cost of links. A network is pairwise stable if (i) no player has an incentive to sever a link, and (ii) no two players both want to create a link. The former implies that players have the discretion to unilaterally delete a relationship. In equilibrium, the latter implies that if some link would benefit both players involved, then the overall network is not stable, as it would be in the players' interests to create the link. Limitations of the equilibrium concept of pairwise stability are that it only considers one link at a time and, at most, two players at a time.

Another limitation of the game-theoretical approach to strategic network formation is that the explicit characteristics of equilibrium networks are often so simplistic that the predicted network structures are too simple to represent the observed network structures. Hence, compared to random graph models, game-theoretical models have a hard time predicting things like what degree distribution the network has. While random graph models are able to account for various features of the observed data, the processes are largely ad hoc and structured in order to match the features. Furthermore, game-theoretical structure provides a framework for evaluating networks and understanding why certain networks are likely to emerge. This is in fact the main limitation of the general random graph approach: The network-formation process detected does not provide us with methods to evaluate the network that emerged. The weaknesses and strengths of each approach greatly illustrate their complementarity. The approach selected to study network formation should therefore be guided by the question of interest.

# 3 Chapter overview

The previous sections outlined the common features of social networks and discussed the two lines of literature within social network formation. While all chapters of this thesis include elements of network formation, three of the chapters add to the literature on strategic network formation (Chapters 2, 3, and 5), while the third chapter take a random graph approach (Chapter 4). Hence, three of the chapters focus on questions related to why networks are formed, whereas one chapter focuses on how networks are formed. This division between chapters reflects an important learning process starting with the understanding of network structures (Chapter 4) and moving towards questions concerned with why networks are formed. A second similarity of the chapters is that they deal with networks in markets; however, different markets are analyzed using two different datasets. The first three chapters are closely related and focus on unregulated economic markets using complete social network data on households from rural Gambia (Chapters 2, 3, and 4). Common to these chapters is that they all consider directed networks. The last chapter considers an undirected network and employs survey data on micro, small, and medium-sized enterprises (MSMEs) in Mozambique to analyze knowledge diffusion (Chapter 5).

Chapter 2, "*Efficiency of land markets: Network level evidence of the importance of social ties*" (joint work with Ulrik Richardt Beck, submitted to Economic Development and Cultural Change), models imperfections in the land market at the network level to capture the impact of pre-existing social networks on allocative efficiency. In the absence of market imperfections, economic theory predicts that land is re-allocated so all farmers cultivate the same amount of land. Skoufias (1995) proposed a testable model, which has been widely used in the literature. The main drawback of this model is that agents are assumed to be independent. However, for mutual transactions to take place, agents must pair up. We extend a standard rural household model with a theory of network formation in order to model a land market where farmers decide on those specific others with whom they will transact. Since a transaction between two farmers requires the consent of both, we build on the pairwise stability equilibrium concept. To test the hypothesis of allocative efficiency, we develop predictors for the allocation of land based on dyad-specific optimization as well as optimization by a benevolent social planner. In contrast to the current literature, we demonstrate the critical role of pre-existing networks at the link level and how this leads to market segmentation. We find that land is exchanged in efficiency-enhancing ways, but not labor. Interpersonal relations in terms of family ties and geographical proximity are found to increase households' access to land. However, only transactions between geographically close households are efficiency enhancing. Simulations of the network structure reveal that allocative efficiency can be substantially improved by reducing enforcement and monitoring costs in ways similar to the proposed efficiency-enhancing effect that operates through geographical proximity.

Chapter 3, "*Are inter-household transactions pro-poor?*" (joint work with Ulrik Richardt Beck, to be submitted to Journal of Development Economics), documents a different mechanism governing network formation in rural Gambia. This chapter investigates whether land transactions are related to differences in income of households, which would provide evidence for the existence of the norm-based access rule guided by social security considerations in traditional village communities. We further investigate whether land transactions come with an obligation to directly or indirectly reciprocate using either the labor or agricultural input markets, before we test the strength of the norm-based land-access rule and reciprocity patterns against increasing population pressure and ethnic diversity. The chapter has three main findings. First, we find that inter-household transactions are pro-poor and that poorer households receive more land. Second, evidence supports the direct reciprocity of labor, whereas no evidence suggests that households are more inclined to give to those that have been generous towards others (indirect reciprocity). Finally, transaction behavior is different in villages characterized by higher population density and ethnic diversity, where land does not flow towards relatively poorer households. In contrast, we did not find evidence that population pressure and high ethnic diversity influence patterns of direct reciprocity.

Chapter 4, *"The importance of structural mechanisms in network formation: The case of rural Gambia,"* investigates the explanatory value of network architecture. I allow for endogenous networking mechanisms to overcome the assumption of dyad independence using the recent development of exponential random graph (ERG) models. I focus on two structural mechanisms: reciprocity (tendency for friendship to be returned) and transitivity (tendency of friends or friends to become friends). The empirical analysis is based on network data collected for a large number of rural villages in The Gambia. The nodes represent households, and the edges indicate whether a transaction has been made in the form of labor, land, or agricultural input. I demonstrate that structural mechanisms are important for the formation of exchange networks. Inability to account for structural mechanisms leads to upward biased parameter estimates of household attributes and dyad-specific characteristics. For example, part of the effect attributed to kinship ties is in fact driven by households' valuation of symmetry in relations. This suggests that the network architecture is important for network formation, and that households take into account the structure resulting from additional partnerships.

Chapter 5, "*Network benefits from co-membership in Mozambican business associations*" (to be submitted to World Development), looks for assortative matching into business associations and examines the idea that business associations facilitate the exchange of information about new technologies and business practices. Recent years have brought about a more positive attitude towards the potential of business associations to help promote firm productivity. Whether enterprise development can be achieved through business associations largely depends on their composition. This chapter finds that membership in business associations more generally is not restricted to specific firms based on sector and geographical location. Next, if productivity-enhancing knowledge diffuses between co-members, then entrepreneurs who are members of the same business association are expected to have more similar business practices compared to non-members. The chapter corrects for self-selection using distance to association headquarters, and finds limited evidence in support of knowledge diffusion of business practices between co-members. This result is consistent with slow convergence of productivity, both across and within sectors in sub-Saharan Africa.

## 3.1 Reflections

This thesis considers mechanisms of network formation in various settings, on a range of outcomes, and from different angles using different empirical approaches. In conclusion of this introduction, I reflect on the main contributions and consider potential weaknesses and gaps for future research.

The principle contribution of this thesis is that it considers different network-formation mechanisms using different network-formation processes. Social networks do not form based on one rule; rather, multiple mechanisms underlie the formation of social networks. Thus, the challenge of rigorously evaluating network formation in order to improve our understanding is particularly important. The general trade-off faced in the analysis of network formation – and by this thesis – is that these mechanisms cannot be examined simultaneously. Notwithstanding this shortcoming, this thesis innovatively considers the related mechanisms and finds that these, in line with the theoretical literature, exist in a developing country context.

A second overall contribution is that this thesis combines the best available social network data and recent econometric techniques to shed light on both traditional economic issues and more nuanced issues within the topic of social network formation. This is where more specific contributions are made. Two methodological contributions are made to the strategic network formation literature. First, Chapter 2 contributes to the resource-based literature on land markets by allowing households to be interdependent, while pairs of households are assumed to be independent. The model developed allows transaction costs between households to vary at the link level. This implies that the transaction costs between households $i$ and $j$ are allowed to differ from the level of transaction costs faced by households $k$ and $m$. This improved methodology can be extended to other markets, and can thereby provide a consistent and robust treatment of transaction costs at the micro level (see Chapter 2 for extensions to the labor market). Second, Chapter 3 contributes to the inter-household gift literature and overcomes omitted variables bias. Specifically, the chapter carefully takes into account the characteristics of the participants on each side of the land market in accordance with the underlying utility function. Generally, the network data used in Chapters 2 and 3 lead to the application of more suitable econometric techniques than have been used previously, and the nature of the data allows us to correct for unobserved heterogeneity even though the empirical analysis is performed using a cross-sectional dataset.

Moreover, both Chapter 2 and 3 contribute to an under-explored area. Chapter 2 contributes by testing the narrative that pre-existing social networks are important for network formation. Chapter 3 contributes by providing detailed quantitative evidence in support of the norm-based access rule. The importance of the norm-based rule regarding access to vital resources has previously been pointed out by sociologists and anthropologists; however, no previous study has applied rigorous econometric methods to examine the rule using a large network dataset. Moreover, Chapter 3 contributes to the experimental literature that has recently highlighted the importance of indirect reciprocity in human behavior in the laboratory (i.e. Seinen and Schram, 2006; Kolm, 2006).

While Chapter 4 uses an exponential random graph approach, it also contributes to the literature on strategic network formation by illustrating how omission of structural mechanisms under the assumption of dyad dependence can lead to upward biased results. Chapter 4 further contributes to the general understanding of network formation in rural communities in a developing country by demonstrating the importance of both the number of links and structural mechanisms in determining link formation.

In addition to the connection to strategic network formation, Chapter 5 explores the under-researched topic of assortative matching by small and medium-sized firms into business associations. The chapter contributes new data, a rigorous identification strategy, and interpretation of how and why firms collocate in business associations. Contribution to the testing strategy of network similarity is not the prevailing focus of this chapter; nonetheless, the chapter provide new insights on a channel of diffusion not previously examined.

Some words of caution are also in order. The empirical analysis throughout this thesis is based on non-experimental, observational data. Despite efforts to address potential sources of bias, parameter estimates may still be biased. This concern is particularly relevant for the result derived from the cross-sectional dataset used to study collocation in Chapter 5. The main concern in Chapter 5 is the presence of unobserved heterogeneity as well as endogenous location of business association headquarters. In part this is because the characteristics of the firms prior to joining a business association are unknown.

A further limitation of these studies is that they consider network formation as an exclusively static game. In principle, network formation is a dynamic and constantly changing process (at least in the longer run) adapting to changes in network patterns and sociodemographic structures. However, the cross-sectional data considered in this thesis does not allow for investigation of dynamic network formation. As panel network data becomes available, it will be possible to study network dynamics and to control for unobserved heterogeneity. Notwithstanding these limitations, this thesis has striven to push forward our understanding of network formation in different social contexts.

# References

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

Jackson, M. O. and A. Wolinsky (1996): “A Strategic Model of Social and Economic Networks,” *Journal of Economic Theory*, 71(0108), 44–74.

Kolm, S.-C. (2006): *Introduction to the Economics of Giving, Altruism and Reciprocity*. In: Kolm, S.-C., Mercier Ythier, J. (Eds.), Handbook of Economics of Giving, Altruism and Reciprocity, vol. 1. Elsevier.

McPherson, M., L. Smith-Lovin and J. M. Cook (2001): “Birds of a Feather: Homophily in Social Networks,” *Annual Review of Sociology*, 27, 415–444.

Milgram, S. (1967): “The small-world problem,” *Psychology Today*, 1, 61–67.

Newman, M. (2004): “Detecting Community Structure in Networks,” *The European Physical Journal B*, 38, 321–330.

Padgett, J. and C. Ansell (1993): “Robust Action and the Rise of the Medici, 1400-1434,” *American Journal of Sociology*, 98, 1259–1319.

Seinen, I. and A. Schram (2006): “Social status and group norms: Indirect reciprocity in a repeated helping experiment,” *European Economic Review*, 50, 581–602.

Skoufias, E. (1995): “Household Resources, Transaction Costs, and Adjustment through Land Tenancy,” *Land Economics*, 71(1), 42–56.

# Chapter 2

# Efficiency of land markets: Network level evidence of the importance of social ties

Ulrik Richardt Beck and Benedikte Alkjærsig Bjerge*

**Abstract**

A fast growing literature highlights the importance of social networks in offsetting the negative effects of market imperfections. This paper models market imperfections at the network level to capture the impact of social networks. We propose a test of the impacts of link characteristics on allocative efficiency using a household survey of input transaction networks in Gambian villages. Geographical proximity has an additional positive effect on allocative efficiency in the land market, while transactions between kin do not enhance efficiency beyond transactions among non-kin. Kinship ties lead to market segmentation where households that are kin-related to landowners have easier access to land. Simulations of the land network reveal that while the impact of geographical proximity on allocative efficiency is limited due to the sparcity of the geographical proximity network, extending the positive effects that arise from neighborship to all links would increase allocative efficiency substantially. Re-estimation of the model using labor market transactions yields no effect on allocative efficiency. This finding is in line with the previous literature, where adjustments towards allocative efficiency are found to take place through the land market rather than the labor market.

**Keywords:** allocative efficiency, land markets, social networks
**JEL classifications:** C21, D85, O12

*Ulrik Richardt Beck (ulrik.richardt.beck@econ.ku.dk) and Benedikte Alkjærsig Bjerge (benedikte.bjerge@econ.ku.dk) are both employed at the Department of Economics, University of Copenhagen. The authors thank Marcel Fafchamps, Elisabeth Sadoulet, John Rand, Edward Samuel Jones, Finn Tarp, and participants at seminars at the University of Oxford, the Nordic Conference in Development Economics 2013, the Development Economics Workshop at University of Gothenburg, and the CSAE conference 2014 for helpful comments. We also thank Dany E. Jaimovich and the research team, including principal investigator Jean-Louis Arcand, investigators Yiu Por Chen and Yong He, and finally, field coordinators Cheikh Ibrahima Fall Diop, Eric Djimeu Wouabe, Malek Garbouj, and Slavica Zec for making the household survey available to us.

# 1 Introduction

A vast empirical literature aims to establish whether land market transactions lead to allocative efficiency (for a review see Otsuka, 2007; Holden, Keijiro and Place, 2009). Following the seminal paper by Skoufias (1995), which presented a testable model of the magnitude of allocative efficiency, the majority of studies have found significant inefficiencies in rural land markets caused by market imperfections often present in developing countries (i.e. Kevane, 1997; Teklu and Lemi, 2004; Ghebru and Holden, 2009; Deininger, Ali and Alemu, 2008, 2009; Jin and Jayne, 2013).

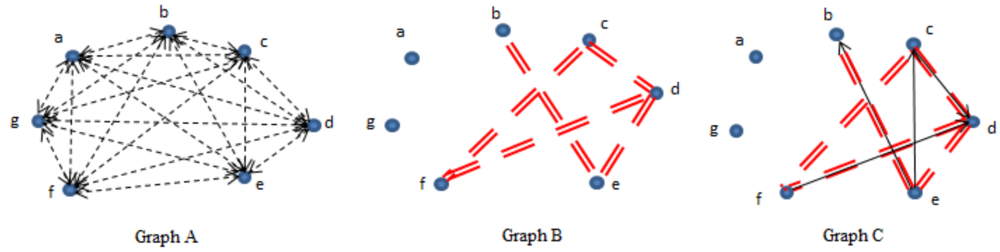
At the same time, there is a substantial literature arguing that preexisting social networks may play an important role in offsetting the negative effects of market imperfections (for a review see Cox and Fafchamps, 2007). Two examples are that diffusion of information in the job market often follows preexisting links (Topa, 2001) and that social networks between agents participating in international trade can help overcome weak enforcement of international contracts and asymmetric information problems (Rauch, 2001). Despite the importance of social networks in market exchange, less effort has been directed towards understanding the effect of social networks on land exchanges. While the main focus of the social network literature has been on overcoming endogeneity of network links by taking into account link-specific characteristics and the structure of the network, the empirical work on the effect of social ties on land markets typically concentrates on links with family members, something which is susceptible to this issue (Sadoulet, de Janvry and Fukui, 1997; Holden and Ghebru, 2006).

This paper examines the relation between social ties and local land transactions in rural Gambia, using a new dataset on inter-household transactions and two types of preexisting social connections. We show how a simple model of network formation offers testable predictions of allocative efficiency in the land market. In contrast to the current literature, we demonstrate the critical role of preexisting networks at the link level and how this leads to market segmentation. The results also show that different social networks can have different effects on economic outcomes: Family connections increase the probability of a transaction but do not increase efficiency further. Geographical proximity, however, does both. The model and estimation framework takes the problem of link endogeneity into the network structure seriously by accounting for preexisting network structures as well as link-specific characteristics. In this way, we are able to test the efficiency characteristics of the land market.

The efficiency of the land market is of crucial importance in many African countries, since it has direct implications for the prospects of reducing poverty and inequality as well as economic growth (Holden, Keijiro and Place, 2009). The issue is particularly important in The Gambia: Land tenancy and land distribution is an increasing concern as land scarcity increases as a consequence of population growth and an already high population density (World Bank, 2005). Contributing factors are the continuing problem of low agricultural productivity, food insecurity, and poverty (Gajigo and Saine, 2011). While steps have been taken to reform the tenancy system in and around urban areas, a complex indigenous system of land tenure dominates the rural areas of The Gambia (Freudenberger, 2000). Land usage rights in rural Gambia are vested in the hands of the descendants of the first settlers. This results in a highly unequal distribution of land ownership rights. The unequal land distribution and opposition to sale of land to non-family and non-residents mean that land transactions are common. Land ownership rights are often transferred on an annual basis, and land recipients are not necessarily allocated the same field every year. These factors make The Gambia a suitable candidate for testing how preexisting social networks in the form of kinship and geographical proximity affect the efficiency properties of the land transaction market.

In rural areas of developing countries, land and labor are regarded as the most important input factors for agricultural production. Empirical studies have found that transactions that are made in order to adjust land–labor ratios most often take place in the land market due to lower monitoring and enforcement costs in this market (Otsuka, 2007). If the magnitude of land market imperfections varies depending on the household one transacts with, transactions will take place where the costs associated with the market imperfections are lowest. Figure 1 illustrates this reasoning. The dotted lines between the nodes in Graph A represent all possible links between five hypothetical households. Each potential link, also known as a dyad, represents a potential exchange of land in either direction. The links in Graph B represent preexisting social ties between households, such as a network family relationship. Finally, Graph C adds realized land transactions to Graph B. If costs associated with land transactions are lower between kin-related households, transactions will tend to follow these paths. However, it is still possible that some non-related households have sufficiently large gains from a transaction that the transaction takes place even though there is no preexisting social tie (households c and e). If transactions follow lower-cost paths, this can lead to

Figure 1: Potential, preexisting and realized ties




market segmentation, where some households are excluded from participating (households a and g) or where transactions take place between households in segmented groups with no or few ties linking the different groups. There are two main questions which we test in the empirical analysis: first, whether social networks lead to market segmentation where some households have preferential access to land and labor, and second, whether social networks are efficiency enhancing in the sense that their presence leads to more efficiency-enhancing transactions, as is often believed in the social networks literature.

We test this by extending a standard agricultural household model with a theory of network formation in order to model a land market where farmers decide on those specific others with whom they will transact. The magnitude of the market imperfections associated with land and labor exchanges is allowed to vary at the link level in order to capture the effects of social networks on realized transactions, as well as the efficiency properties of these transactions. Farmers differ in terms of individual characteristics, and links between farmers differ in terms of preexisting social ties. To test the hypothesis of allocative efficiency, we develop predictions for the allocation of land based on dyad-specific optimization as well as optimization by a benevolent social planner. We test the predictions on data from a household survey conducted in 2009 in rural Gambia. The survey includes all households residing in 52 villages located in different regions of the country. The data not only offer information on household characteristics, but also contain information on all land and labor transactions at the network level, as well as information on the preexisting social networks of kinship ties and neighborship.

We find that land is exchanged in efficiency-enhancing ways. Interpersonal relations in terms of family ties and geographical proximity are found to increase households' access to land. However, only the geographical network gives rise to additional efficiency enhancing transactions. The family network does not help offset the inefficiencies created by the unequal initial land distribution. In line with previous findings in the literature, we find that land, but not labor, is exchanged in efficiency-enhancing ways, but that the most efficient outcome is not achieved.

The results have important consequences for the impact of social networks on land market transactions. The evidence suggests that there are substantial imperfections in the rural Gambian land market, and that the kinship network does not offset the lack of a more formal land market. However, simulations of the network structure reveal that allocative efficiency can be substantially improved by reducing enforcement and monitoring costs in ways similar to the proposed efficiency-enhancing effect that operates through neighborship. Hence, action should be taken to reduce transaction costs in the informal land market in The Gambia, or to improve the functioning of a more formal market system.

The structure of the present paper is as follows. In Section 2, we present in more detail the tenure system of rural Gambia and describe the data. In Section 3, we present a theoretical model to explain the importance of preexisting social networks for allocative efficiency. Section 4 describes the empirical testing strategy, while Section 5 provides empirical tests of the predictions developed in the theoretical model. Section 6 concludes.

## 2 Background and data description

The Gambian economy is dominated by agriculture, which contributes about 33 percent of GDP (IMF, 2007) and employs 68 percent of the labor force.[1] Rural villages in The Gambia are typically organized into compounds, which correspond to a group of people (usually from the same family) who work jointly on common fields, eat together, and organize daily activities under the management of a single decision maker. Depending on the size of the compound and presence of friction between adult males within the kin residence group, independent cooking and consumption units (*dababas*) can co-exist within the compound. Hence, each individual *dababas* corresponds to what is normally understood as a household unit.

Although the state formally owns all land, de facto usage rights are determined by a complex indigenous land tenure system. Two principal types of usage rights exist, referred to as *primary* and *secondary* rights. A household

[1]World Bank Indicators (2010). http://data.worldbank.org/country/gambia

with primary rights over a plot of land can decide which crops to grow and whether to lend or rent all or some of the land out to other farmers. Traditionally, the descendants of those who first settled and cultivated the land as well as the village chief, called the *Alkalo*, retain considerable amounts of the primary usage rights in the village (Pamela, 2010). This creates a highly unequal distribution of primary land rights between households. The descendants of the first settlers who possess surplus land have a moral obligation to lend out land to those in need, and the *Alkalo* can allocate land to households with no or few landholdings. In this sense, the role of the *Alkalo* resembles that of a social planner who decides land allocations for everyone involved. Transactions of secondary usage rights are thus often norm-driven and non-monetary in nature: The senders of land rarely receive monetary payment for the land that they lend to other farmers. Sometimes there will be a symbolic payment of kola nuts or cash (Freudenberger, 2000; Arcand and Jaimovich, 2010).[2]

Hence, rural households can access land in three ways: through inheritance from primary rights holders, through secondary usage rights transactions of land by the *Alkalo* and descendants of the first settlers, and finally, through normal market-based transactions, primarily rental transactions (Freudenberger, 2000).

Smallholder farmers in Gambia mostly work on their own farm (Pamela, 2010), as one would expect in a setting of highly imperfect labor markets. The seasonal nature of agricultural production makes family labor periodically insufficient. The shortage is predominately before and during the rainy season in relation to weeding and harvesting, particularly with respect to groundnuts, which are highly labor-intensive undertakings. Thus, land and labor transactions have a sequential nature that follows the agricultural seasons: First, land transactions occur, and later, during weeding and harvesting, additional labor may be hired (Swindell, 1987).

## Data

The data come from a baseline survey conducted between February and May 2009 for the purpose of evaluating of a national Community Driven Development Project (CDDP). The survey covers 60 randomly selected villages, representative of six out of eight Local Government Areas across different agro-

[2] We use the term “senders of land” to refer to households that give land to other households, regardless of whether the transaction is a norm-driven non-monetary transaction or a more standard rental agreement. Likewise, we use the term “receivers of land” to refer to households that receive land from any form of transaction.

ecological zones, and with populations between 200 and 1,000. The dataset contains three categories of information: (1) village-level information, (2) a standard household survey, and (3) information on six networks: land transactions, labor transactions of the household head, input, credit, marriage, and detailed kinship information. The dataset is unique in that it contains information on these six networks for transactions between all households residing in the village, as well as information on household endowments and structures.[3]

For the land and labor networks, the transacted amount is available, measured in hectares of land and working days, respectively. The land network contains both non-monetary transfers of secondary usage rights as well as any cash-rental transactions and sharecropping agreements, if present. One potential issue with self-reported link data is that the respondent may give information about their desire to link instead of their actual links (Comola and Fafchamps, Forthcoming 2014). In our case, this could pose a problem if respondents reported their desired transactions of land and labor instead of their actual transactions. However, as households also had to state the actual amount of land and labor transacted and not just the presence of a link, the risk of this is minimized.

Five villages were dropped due to substantial amounts of missing household-level information. Second, as we are concerned with land transactions in rural areas, three semi-urban villages were also dropped.[4] Moreover, we restricted the sample to households where the main activity of the household head is related to farming, as we are interested in allocative efficiency among farmers. Applying these selection criteria, the sample used for the empirical analysis consists of 1,625 households across 52 villages, corresponding to 57,060 within-village household dyads.

The definition of households adopted in the survey closely follows Matlon (1988 cited from Udry, 1996). Notably, we observe households and not compounds. This means that if several households exist within one compound, the network between these will be present in our data. Around 14 percent of the household heads in the sample are not the head of the compound in which

[3]The data were collected using a structured group approach with a median household coverage rate in the villages of 94 percent. For detailed information on the sampling methodology and data description, see Jaimovich (2011).

[4]In terms of network activity, some 2 percent of the households in the semi-urban villages participate in the land market, while 10 percent participate in the labor market. The main reason for the absence of land sharing could be the very small landholdings in these areas (0.243 hectares per household compared to 10.282 in the rural villages), as well as increased options for employment outside the village.

they live.

While the dataset does not contain information on households that reside outside the village, it includes information about the connections between households in the village and households outside the village. It is therefore necessary to assume that the village is the natural domain for potential exchanges. This assumption is supported by the fact that external actors are not very important in the land network.[5] The low level of land transactions with households outside the village is likely to be explained by the immobility of land, which leads to high transaction costs as inputs must be brought to the land and outputs must be transported to the places of consumption or sale.

Table 2 provides household-level information for all farmers, separately for those participating on the two different sides in the land market and farmers in autarky.[6] The data is consistent with the description of rural Gambian markets given in the last section: The largest households in the sample have more than 50 members (only 0.01 percent of the sample). These large households are partly explained by the polygamous nature of rural Gambian society (50 percent of household heads have more than one wife). The households in our sample are predominately led by poorly educated men: Only 9 percent have any formal education. The average monetary income per capita is 2,750 Gambian Dalasis a year, PPP-equivalent to 282 USD a year, of which 16 percent stems from agricultural activities.[7] Households participating on either side of the land market tend to be larger in terms of household size and working adults than households in autarky. Interestingly, receivers of land and households in autarky have similar amounts of land with ownership rights (9.0 and 8.3 ha. respectively), whereas senders of land have larger primary rights landholdings (17.5 ha.). The initial land–labor ratio is lowest for the receivers of land, indicating that those with the lowest land–labor ratios are indeed more likely to receive land.

Furthermore, the villages have a dense kinship network: 85 percent of house-

[5] Using the same data, Jaimovich (2013) examines the substitutability between internal and external links and the impact on reciprocity. Summary statistics suggest that only 5 percent of households send land to external partners, while 8 percent receive land from non-village members. The author does find evidence supporting substitutability between internal and external links.

[6] The category of autarky includes households that do not participate in the land market on either side.

[7] Using Penn World Tables PPP-adjusted exchange rate. Note that consumption and bartering of own production is not included in this figure. The Gambia Integrated Household Survey of 2010 found that mean consumption of own production and gifts in Gambia in 2010 amounted to 6,283 dalasis per household per year, or 776 dalasis per person, using the mean household size of 8.1 from that survey (Gambia Bureau of Statistics, 2011).

Table 1: Household-level descriptive statistics

| | All farmers | | By land participation status | | |
|---|---|---|---|---|---|
| | Mean | Std. Dev. | | Mean | |
| | (N=1625) | | Autarky (N=672) | Senders (N=459) | Receivers (N=687) |
| Household size | 14.100 | 15.002 | 12.242 | 15.638 | 15.425 |
| Age of head | 54.319 | 16.229 | 53.280 | 56.353 | 54.593 |
| HH has family links in the village | 0.958 | 0.202 | 0.947 | 0.987 | 0.953 |
| HH head has family links in the village | 0.850 | 0.357 | 0.829 | 0.917 | 0.841 |
| Wife of HH head has family links in the village | 0.508 | 0.500 | 0.484 | 0.603 | 0.497 |
| HH have marriage links in the village | 0.642 | 0.480 | 0.623 | 0.741 | 0.606 |
| Female headed household | 0.057 | 0.232 | 0.074 | 0.040 | 0.047 |
| Illiterate | 0.489 | 0.500 | 0.492 | 0.440 | 0.507 |
| Formal education | 0.087 | 0.282 | 0.093 | 0.092 | 0.077 |
| Compound head | 0.858 | 0.349 | 0.827 | 0.917 | 0.863 |
| Monogamous | 0.446 | 0.497 | 0.472 | 0.413 | 0.438 |
| Polygamous | 0.501 | 0.500 | 0.467 | 0.531 | 0.521 |
| Christian | 0.004 | 0.066 | 0.002 | 0.002 | 0.009 |
| Ethnicity: Mandinka | 0.541 | 0.498 | 0.580 | 0.518 | 0.534 |
| Ethnicity: Fula | 0.186 | 0.390 | 0.184 | 0.181 | 0.189 |
| Ethnicity: Wollof | 0.101 | 0.301 | 0.085 | 0.134 | 0.090 |
| Land owned with official rights (hec.) | 10.502 | 24.478 | 8.307 | 17.493 | 9.009 |
| Land-labour ratio (ha. land per active worker) | 2.762 | 7.386 | 2.568 | 4.517 | 2.005 |
| Number of working adults | 5.371 | 10.390 | 4.802 | 5.625 | 5.919 |
| Income per capita (1,000 GMD/PPP) | 0.282 | 0.285 | 0.293 | 0.254 | 0.279 |
| Agricultural share (share of income) | 0.161 | 0.271 | 0.141 | 0.189 | 0.171 |
| Emigrated household member | 0.518 | 0.500 | 0.489 | 0.605 | 0.496 |
| Receive remittances | 0.455 | 0.498 | 0.429 | 0.538 | 0.441 |

Note: The number of observations in the three land participation columns sum to more than 1,709 since some households are both senders and receivers of land. HH is applied as abbreviation for household.

Table 2: Land market participation rates by initial land–labor ratios

| | All | Landless | 0.1-0.6 ha/w | 0.6-1.6 ha/w | 1.6-3.0 ha/w | > 3.0 ha/w |
|---|---|---|---|---|---|---|
| % in land market | 45.2 | 36.1 | 42.4 | 41.1 | 53.8 | 56.5 |
| % Land sender | 21.0 | 2.0 | 14.9 | 19.0 | 30.1 | 40.5 |
| % Land receiver | 28.9 | 35.7 | 33.3 | 26.2 | 30.5 | 21.4 |
| Observations | 1,625 | 294 | 255 | 516 | 266 | 294 |

Note: ha/w corresponds to the number of hectares per working adult.

hold heads have relatives living in the village. Similarly, in 51 percentof households, the wife (or wives) has relatives living in the village, and 64 percent have marriage ties to other households in the village. Some 96 percent of households have at least one of these three kinds of links. Senders of land are more likely to have all three kinds of kinship ties to other households, stressing the fact that sender households, often being the first settlers, play an important role in the kinship networks in the village.

Table 2 shows that receivers of land have lower ex-ante landholdings. Table 2 reports in more detail the probability that a household is in the land market on either side against the initial landholdings in ha. per working adult. A striking pattern emerges: Overall, 45.2 percent of the households in the sample engage in land transactions. Households with higher initial land–labor ratios are simultaneously more likely to send land and less likely to receive land. Furthermore, 36 percent of landless households in the sample receive land from other households.[8]

# 3 Empirical framework

In this section, we introduce a simple theoretical framework inspired by Sadoulet, Murgai and Janvry (2001) in order to explain land transactions at the dyad level while incorporating the features outlined above. In the description of the model, the focus is kept on land transactions. Labor transactions can be described in the same way.

[8]At first glance, the 2 percent of landless households who send land seems like a paradox. However, these are households that also receive land and thus end up with a nonnegative amount of land. Households that have a negative amount of land, taking into consideration all households in the villages, are excluded from the sample and therefore do not appear in this table.

We extend the original model to explicitly take into account transactions and transaction costs at the dyad level. Consider a household optimization problem where the prices of both input factors and output are exogenously given.[9] A household $i$ maximizes the value of its production by choosing input levels of two essential inputs: land, $A_i$, and labor, $L_i$. The price of renting land, $r$, and the price of renting labor, $w$, are exogenously given. The amount of land used in production by household $i$ is determined by the endowment, $\bar{A}_i$, the amount of land the household receives from $j$, $A_{ij}$, and the amount sent to other households, $A_{ji}$. Thus, $A_{ij} \geq 0$ and $A_{ji} \geq 0$. The production function $q$ is identical for all households in a village,[10] increasing in both inputs, concave, and twice differentiable. Furthermore, it exhibits constant returns to scale.[11]

Imperfections in the land and labor market are modeled as both variable and fixed costs. Variable costs are captured by the parameters $\alpha_{ij}^A$ and $\alpha_{ij}^L$, respectively. These costs are modeled as extra costs for receivers of land, but could also be modeled as extra costs for senders without affecting the predictions. Variable transaction costs are likely to be affected by social ties. Thus, the size of these imperfections can vary between dyads. Fixed costs in the land and labor markets are modeled by the parameters $\psi_{ij}^A$ and $\psi_{ij}^L$, respectively, and are also allowed to vary between dyads.

For a household $i$, the maximization problem is (where $j$ is other households

[9]As described in Section 1, many land transactions are moneyless assignments and do not carry a price. The price of land in the model is the price on the actual rental market. However, non-monetary transactions have a shadow value that must be at least as high as the price in the rental market in order for them to take place.

[10]In order for the production function to be constant across all households, there must not be any differences in technology. Chavas, Petrie and Roth (2005) find that technical inefficiency among Gambian farmers is modest, indicating that differences in the factor mix are not driven by technological differences. However, the authors find substantial allocative inefficiency in farm input allocations, which motivates the focus of the present paper. As an additional attempt to investigate heterogeneity in household land and labor quality, we create a measure for the extent of variations in land and labor quality if allocative efficiency is assumed to hold and production technology is assumed to be a Cobb–Douglas function of land and labor. The measure developed in the Appendix predicts how different the relative skill-to-land-quality levels must be if the allocation after transactions has taken place is efficient. Results suggest that there must be substantial variation in household parameters for this to explain the actual distribution: The relative quality of labor and land available to households within a single village must differ by a factor greater than 40 for allocative efficiency to hold. In the small rural communities considered here, these differences are unrealistically large. Hence, heterogeneity in household parameters is unable to explain the large differences in within-village land–labor ratios. Finally, section 5 re-estimates the main specification including sender and receiver fixed effects, and the results are unchanged.

[11]It has been argued that in the absence of large-scale mechanization, the assumption of constant returns to scale is not unrealistic (Hayami and Otsuka, 1993; Deininger and Feder, 2001).

in the village)

$$\begin{aligned}
\max_{A_{ij},L_{ij}} \quad & p * q(\bar{A}_i + \sum_j [A_{ij} - A_{ji}], \bar{L}_i + \sum_j [L_{ij} - L_{ji}]) \\
- \quad & (\sum_j [A_{ij}(r + \alpha^A_{ij}) - A_{ji} r]) \\
- \quad & (\sum_j [L_{ij}(w + \alpha^L_{ij}) - L_{ji} w]) \\
- \quad & \psi^A_{ij} I[A_{ji} > 0] - \psi^L_{ij} I[L_{ji} > 0]
\end{aligned} \tag{1}$$

Links can be deleted unilaterally, but the link formation process is regarded as bilateral. Hence, for a transfer to take place (i.e., for a link to be established), the consent of both parties is required.[12] Equilibrium is achieved when all links are pairwise stable, i.e., when no agent has an incentive to delete an existing link and there is no pair of agents such that both members of the pair have an incentive to form a new link with each other (Jackson and Wolinsky, 1996). When a given exchange between two households is determined, all other exchanges in the network are taken as given. We denote this exchange the marginal exchange. Furthermore, we abstract from potential second-order benefits from indirect links.

Alternatively, one could model the link decision formation in a principal–agent framework with varying costs of linking between different agents and principals. This is the approach taken by Macours, Janvry and Sadoulet (2010), where expected costs of forming a link between a landlord and a tenant vary with socioeconomic attributes of the tenant, as this is thought to affect the probability that the tenant will squat on the rented land. In The Gambia, however, it is not obvious who potential landlords and potential tenants are without looking at the actual land network. Therefore, we prefer to treat senders and receivers symmetrically.

Recently, Comola and Fafchamps (Forthcoming 2014) have questioned the assumption of bilateral link formation, noting that in some networks, consent may not be required of both parties. Instead, links can be formed unilaterally. An example of such a network is exchange of information. In the present model, bilateral link formation describes the link formation process well: Households choose with whom to transact, but pay a price in the form of lower value of production if they do not conduct viable efficiency-enhancing transactions.

[12]Bilateral link formation means that $i$ can only receive land if $j$ agrees to send land, i.e., $A_{ji} = -A_{ij}$. See Comola and Fafchamps (Forthcoming 2014) for examples of bilateral and unilateral link formation.

Combining the first-order conditions for households $i$ and $j$, we get the following expression, which implicitly defines the size of the transaction from $j$ to $i$. Conditional on exchanging land (i.e., that the fixed transaction costs are not prohibitively high), the households will exchange land until the marginal value of land is equalized between the two households:

$$\begin{aligned} f(A_{ij}, \alpha^A_{ij}) &= p\frac{\partial q(\bar{A}_i + \sum_k [A_{ik} - A_{ki}] + A_{ij}, \bar{L}_i)}{\partial A_{ij}} - \alpha^A_{ij} \\ &- p\frac{\partial q(\bar{A}_j + \sum_k [A_{jk} - A_{kj}] - A_{ij}, \bar{L}_j)}{\partial A_{ij}} = 0 \end{aligned} \tag{2}$$

where $i \neq j$, $j \neq k$, $i \neq k$ and the production functions are evaluated taking all other exchanges in the network as given. Intuitively, if marginal values of land are different enough to overcome the cost of exchanging land, a land transaction will take place.

Now, consider the case where the marginal value of production of land is higher for household $i$ than for household $j$ before the marginal exchange, i.e., $\frac{\partial q(\bar{A}_i + \sum_k [A_{ik} - A_{ki}], \bar{L}_i)}{\partial A_{ij}} > \frac{\partial q(\bar{A}_j + \sum_k [A_{jk} - A_{kj}], \bar{L}_j)}{\partial A_{ij}}$. In this case, a transfer from $j$ to $i$ will increase efficiency but will not necessarily take place due to the presence of transaction costs. However, in the absence of transaction costs in the land market, land–labor ratios will equalize perfectly. In this case, the equalizing land exchange is

$$A^*_{ij} = \frac{\bar{L}_i A_{j/-ij} - \bar{L}_j A_{i/-ij}}{\bar{L}_j + \bar{L}_i} \tag{3}$$

where $A_{i/-ij} = \bar{A}_i + \sum_k A_{ik}$ is the land usage rights of $i$, taking all transactions into account except the transactions between $i$ and $j$. In practice, transaction costs will be present, resulting in less than perfectly equalized land–labor ratios.[13] Applying the implicit function theorem to equation 2, we find for $A_{ij} > 0$: $\frac{dA_{ij}}{d\alpha^A_{ij}} = -\frac{\partial f/\partial \alpha}{\partial f/\partial A_{ij}} = (p * (\frac{\partial^2 q_i}{\partial A^2_{ij}} + \frac{\partial^2 q_j}{\partial A^2_{ij}}))^{-1} < 0$, where $q_i$ and $q_j$ are the production functions of the two households evaluated taking all transactions into account. Thus, the optimally exchanged amount from $j$ to $i$ will be lower than the equalizing amount the larger $\alpha^A_{ij}$ is. A non-zero fixed transaction cost will not affect the amount of land transacted, but will make some transactions prohibitively expensive, resulting in fewer transactions. We approximate these

[13] The predictor is also relevant under an alternative assumption of decreasing returns to scale of the production function. If there are decreasing returns to scale, both land–labor ratios and farm sizes should equalize across households under allocative efficiency.

results by allowing transaction costs to affect the amount transacted from $j$ to $i$ in a linear fashion: $A_{ij} = max[0, -\alpha_{ij}\theta_0 + (1 - \alpha_{ij}\theta_1)A^*_{ij}]$.

## A social planner's problem

Alternatively, consider the allocation of land from the perspective of a social planner. The social planner wants to equalize land–labor ratios but realizes that transactions are costly. Thus, the social planner wants to minimize the amount of land that needs to be transferred and the number of transactions.[14] This approach also corresponds to a setting where transactions are driven by norms of equity and where the desired land–labor ratio is the village average level. In this case, households with higher than average land–labor ratios will send land until the land–labor ratio is equal to the village average $v$. The receiving households will be those with initial land–labor ratios below the village average. They will receive land until their land–labor ratio is equal to the village average. Thus, the land transaction between $i$ and $j$ is given by

$$A^{*v}_{ij} = max(0, min(A^{vi}_{ij}, A^{vj}_{ij})) \tag{4}$$

where $A^{vi}_{ij}$ and $A^{vj}_{ij}$ are implicitly defined by

$$p\frac{\partial q(\bar{A}_i + \sum_k [A_{ik} - A_{ki}] + A^{vi}_{ij}, \bar{L}_i)}{\partial A_{ij}} - \alpha^A_{ij} = v \tag{5}$$

$$p\frac{\partial q(\bar{A}_j + \sum_k [A_{jk} - A_{kj}] - A^{vj}_{ij}, \bar{L}_j)}{\partial A_{ij}} = v \tag{6}$$

where $i \neq j$, $j \neq k$, $i \neq k$, and the production functions are evaluated taking all other exchanges in the network as given. If there are no transaction costs and one household has a higher land–labor ratio than the village level average, the transaction that will take place is the smallest transaction that allows one of the households to reach the village level average: $A^{*v}_{ij} = min(\bar{A}_i + \sum_k [A_{ik} - A_{ki}] - v\bar{L}_i,\ \bar{A}_j + \sum_k [A_{jk} - A_{kj}] - A^{vj}_{ij} - v\bar{L}_j)$

Again, transaction costs will lower the transacted amount if the transacted amount is positive in the absence of transaction costs.[15] We approximate this

[14] A similar result emerges from a theoretical analysis of altruism effects in a network setting (Bourlès and Bramoullé, 2013). Here, if it is equally costly to give gifts to all persons, equilibrium transfers minimize the aggregate transfer needed to reach equilibrium outcomes.

[15] In the model, transaction costs only affect $A^{vi}_{ij}$ and not $A^{vj}_{ij}$. This is an artifact of the simplified representation where transaction costs are placed on the receivers of land and labor. In reality, one could easily think of transaction costs affecting the price for both receivers and senders of land.

result by allowing transaction costs to affect the transacted amount in a linear fashion: $A^v_{ij} = max[0, -\alpha_{ij}\theta^v_0 + (1-\alpha_{ij}\theta^v_1)A^{*v}_{ij}]$. Similarly, the presence of fixed transaction costs means that fewer transactions will take place.

# 4 Estimation strategy

Even after modeling parts of the dyad-specific transaction costs, much dyad-specific information is unobserved but affects the cost of conducting a transaction between two households. An example of such unobserved information that affects $A_{ij}$ is the level of trust between the households. If the level of trust is low, a potential sending household will be more reluctant to send land to a potential receiver. This is represented by an unobserved shock $\epsilon_{ij}$ to the amount that the two households want to transfer: $A_{ij} = max[0, \alpha_{ij}\theta_0 + (1-\alpha_{ij}\theta_1)A^*_{ij} + \epsilon_{ij}]$. If $\alpha_{ij}\theta_0 + (1-\alpha_{ij}\theta_1)A^*_{ij} + \epsilon_{ij} > 0$, a transfer will take place.

Assuming dyad-specific shocks to the transacted amount are normally distributed, the model above can be estimated by a tobit model. Denoting link-specific variables affecting transaction costs by $w_{ij}$ and household-specific attributes affecting the transaction costs by $(z_i,\ z_j)$, the equation to be estimated is

$$\begin{aligned} A_{ij} \quad = \quad & max(0, \gamma_o + \gamma_1 A^*_{ij} + \gamma_2 A^{*v}_{ij} + \gamma_3 w_{ij} + \gamma_4 w_{ij} A^*_{ij} + \gamma_5 w_{ij} A^{*v}_{ij} + \\ & \beta_1 z_i + \beta_2 z_j + \epsilon_{ij}) \end{aligned} \tag{7}$$

where $w_{ij}$ are dyad-specific characteristics and $z_i$ and $z_j$ are household-specific characteristics of household $i$ and $j$, respectively. Included in $z$ is a set of variables to capture differences in household labor quality and managerial abilities, namely the gender of the household head and whether the household head has any formal schooling. The amount of land $i$ receives from $j$, $A_{ij}$, is measured in hectares. Allocative efficiency can still be achieved even though *some* dyads experience transaction costs. However, efficiency is only achieved if households have sufficient links where there are no transaction costs, in such a way that these households are able to equalize land–labor ratios. Thus, if there is allocative efficiency and land–labor ratios are not equalized before the marginal transaction (meaning that $A^*_{ij} \neq 0,\ A^{*v}_{ij} \neq 0$), the marginal transaction must equalize land–labor ratios. Thus, a test of $\gamma_0 = 0\ \wedge\ \gamma_h = 1,\ h = \{1, 2\}$ in the restricted model where $\gamma_3 = \gamma_4 = \gamma_5 = \beta_1 = \beta_2 = 0\ \wedge (\gamma_1 = 0\ \vee\ \gamma_2 = 0)$ is a test of allocative efficiency in the sense that land–labor ratios equalize perfectly through transactions.[16] If we fail to accept the hypothesis of allocative efficiency, land may still flow in the predicted direction. $\gamma_1 > 0 \vee \gamma_2 > 0$ supports the prediction that land flows such that it enhances allocative efficiency. This could also have been observed by simply investigating the land–labor ratios of households directly. However, such a simple model provides a baseline against which the effect of specific link-characteristics can be examined.

Apart from the predicted household-level and village-level efficiency-achieving transactions ($A^*_{ij}$ and $A^{*v}_{ij}$), our main variables of interest are preexisting social ties. In order to examine whether transaction costs affect the transacted amount and whether social ties are efficiency enhancing, family ties and geographical proximity are included as link-specific indicators in $w_{ij}$. If $\gamma_3 \neq 0$, this is a sign of market segmentation: Link-specific attributes affect access to land, and some households have easier access than others. If social ties affect transaction costs negatively, we expect to find $\gamma_3 > 0$. To test whether market segmentation enhances allocative efficiency, we introduce interaction terms between link-specific characteristics and the efficiency-achieving predictors. If $\gamma_h > 0,\ h = \{4, 5\}$, the market segmentation due to social ties results in an efficiency gain. On the other hand, $\gamma_3 > 0 \wedge \gamma_h = 0,\ h = \{4, 5\}$ would be evidence indicating that the presence of social ties creates a system of insiders and outsiders where connected households have preferential access to land but these additional transactions do not increase efficiency. Moreover, if the efficiency-achieving predictors become insignificant when the interaction terms are included, then this is evidence in support of efficiency transactions only going through preexisting ties ($\gamma_h = 0,\ h = \{1, 2\} \wedge \gamma_h > 0,\ h = \{4, 5\}$).

Two measures of preexisting social ties are considered. The first is an aggregated kinship measure taking the value 1 if two households are family related through the household head, the wife, or marriage arrangements. In order to more closely examine what kind of family links matter for the possibility of receiving land, the kinship measure is disaggregated into kin of the household head, kin of the wife of the household head, and marriage links. Table A1.1 reports dyad-level descriptive statistics.

Second, geographical distance is measured by two variables: (i) a dummy variable taking the value 1 if the households $i$ and $j$ live in neighboring compounds, and 0 otherwise, and (ii) a dummy variable taking the value 1 if

[16]In the case of superfluous land exchanges compared to the minimum required to reach allocative efficiency, one could end up with parameter estimates on our efficiency-achieving predictors greater than one, even though this would be difficult to rationalize using the model presented earlier. However, the subsequent empirical results eliminate this concern.

households $i$ and $j$ have neighboring plots, and 0 otherwise.[17] The data on geographical distance were only collected for a subsample of the villages. This means that the sample of neighboring compounds (plots) is reduced to 25 (19) villages, corresponding to 842 (624) households. In the reduced sample, almost 10 percent of the households are regarded as compound and/or plot neighbors.

It is possible that reducing imperfections is not the only effect of preexisting social ties. Transactions between family-tied households may involve aspects of altruism, and strategic exchanges in order to establish a contract of reciprocity which is repaid when parents get old and need care (Laferrère and Wolff, 2006). These considerations are not present in the case of geographical proximity. In this sense, geographical proximity has a cleaner effect on transactions of land, working only through reduced market imperfections.

A potential issue is that if land and labor quality vary, or if other household-specific unobserved factors affect the production function, households will not want to perfectly equalize land–labor ratios, even in the absence of market imperfections. Similar, if there are no transaction costs in the labor market, households may choose to use the labor market rather than the land market to equalize land–labor ratios. In both cases, this would imply that the predictors overestimate the amount of land that will be transacted in the absence of transaction costs. For this reason, the test of whether allocative efficiency is reached should be interpreted with caution. However, a significant effect of the predictor variable still yields valuable information about the direction of land transactions and the relative characteristics of senders and recipients. Finding that the predictors help explain land transactions even in the presence of these potential issues can therefore be considered strong evidence that an effect is present.

To the extent that other variables are correlated with the actual amount of land exchanged and the household efficiency-achieving predictors, it is essential to control for these in order to obtain the true effect of social and geographical proximity. We therefore include various control variables ($z$) related to the characteristics of the households (including the number of household members, whether the household receives remittances, and the relative wealth level) and the household head (including age, gender, and educational level). Summary statistics are reported in Table 2.

We do not include $A^*_{ij}$ and $A^{*v}_{ij}$ in the same model due to collinearity issues. As seen from equation 2, the price of output will affect which transactions take

[17]The two geographical variables are positively correlated but only imperfectly (0.18).

place, as well as the transacted amount. This may vary between villages. This is one reason for including village fixed effects.[18] Even though we include village fixed effects, it is essential to correct standard errors for non-independence across observations, which arises principally because the residuals from dyadic observations involving the same individual $i$ and $j$ are correlated. However, we cannot rule out that all households in a village are dependent on each other. Therefore, we cluster standard errors at the village level. This approach is conservative in the sense that we do not assume anything about the dependency of dyadic observations inside the villages (Fafchamps and Söderbom, 2014).

Due to the sequential nature of land and labor transactions, we do not take labor transactions into account when constructing land market predictors. When we investigate the labor market, we take all land transactions as given. However, neither taking the labor market transactions as given when the land market predictors are calculated, nor taking land transactions as given when calculating the labor market predictors affects the results.

# 5 Results

## Baseline results

The results of the estimation of equation 7 are shown in Table 3. Columns (1) and (3) include only the household- and village-level efficiency-achieving predictor of land transactions, respectively. Both predictors are significant and positive, as expected. However, they are also significantly smaller than 1. Combined with the significantly negative constant terms, this implies that some adjustment towards allocative efficiency does take place, but that the adjustment is only partial.[19]

Part of the correlation between the predicted efficiency-achieving transactions and the amount of land transacted may be due to household-specific characteristics. For example, the ethnicity of a household may affect the potential land transaction if households with the same ethnicity face lower dyad-specific transaction costs. To net out the effect from observables, columns (2) and (4)

[18] Usually, the inclusion of fixed effects is not feasible in a tobit model due to the problem of incidental parameters. However, this is an issue only when there are few observations for each fixed effect (Wooldridge, 2010, p. 495-496). Due to the dyadic nature of the regressions, there are many observations for each fixed effect, varying from 198 to 3600. Thus, we can include village fixed effects. This setup is similar to Fafchamps and Gubert (2007).

[19] Employing a test of efficiency of the land market first proposed by Skoufias (1995), viz., taking the household as the level of analysis (as opposed to the dyad in the present paper), gives a similar result. The results are shown in the Appendix.

include additional regressors, including household characteristics (household size, whether the household receives remittances, and the relative wealth level of the household), as well as characteristics of the household head (ethnicity, age, and educational level). The coefficient estimates of the predictors change only marginally. Thus, the effect picked up by this predictor cannot be explained by any of the additional explanatory variables.

Turning to the additional regressors, the magnitude and significance do not depend on the predictor used. The results suggest that the characteristics of the sending household is the primary household-level driver of land transactions. In line with the nature of the Gambian tenure system, where land is typically inherited, we find that households with older heads send more land. This is also consistent with a previous study examining a non-directed land network in rural Ghana, which found that differential age among household heads increases the probability of exchanging land (Udry and Conley, 2004). Second, illiterate household heads send less land. A possible explanation is that households with more human capital can afford to share more land. The relative wealth levels of both sending and receiving households are important: Households that are relatively wealthier both send and receive more land. A possible explanation is that wealthy households have primary usage rights over more land, making them more likely to be senders of land. On the receiver side, it is possible that wealthy households are better able to post collateral, making a land transaction less risky for the sender. There is no significant effect of having the same ethnicity. This result echoes that of Arcand and Jaimovich (2012), who, using the same dataset, do not find evidence that ethnic fragmentation causes sub-optimal economic exchanges. The estimation results on our main variables of interest are still positive and statistically significant.[20]

## Market segmentation caused by social ties

Table 4 considers market segmentation and whether transaction costs decrease with social and geographical proximity. Column (1) includes a dummy variable for kinship ties at the dyad level.[21] The coefficient estimate is positive and significant, as expected. This implies that kin-related households have easier access to land, leading to market segmentation in rural Gambian villages.

[20]In the rest of this paper, the results using the village-level efficiency-achieving predictor $A_{ij}^{*v}$ are shown in the Appendix since the sign and significance throughout the empirical investigation are identical to those from $A_{ij}^{*}$.

[21]Estimation output for the additional explanatory variables is available from the authors upon request.

Table 3: Regression results: Baseline

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| $A^*_{ij}$ | 0.048*** | 0.042*** | | |
| | (0.009) | (0.008) | | |
| $A^{*v}_{ij}$ | | | 0.087*** | 0.076*** |
| | | | (0.024) | (0.020) |
| Same ethnicity | | 0.189 | | 0.213 |
| | | (0.184) | | (0.181) |
| *Receiver characteristics (i):* | | | | |
| Household size | | 0.003 | | 0.002 |
| | | (0.006) | | (0.006) |
| Age of head | | -0.007 | | -0.006 |
| | | (0.007) | | (0.007) |
| Illiterate | | 0.002 | | -0.008 |
| | | (0.203) | | (0.203) |
| Formal schooling | | -0.448 | | -0.434 |
| | | (0.335) | | (0.336) |
| Female head | | -0.553 | | -0.548 |
| | | (0.520) | | (0.518) |
| Receive remittances | | -0.188 | | -0.189 |
| | | (0.192) | | (0.192) |
| Wealth level | | 0.308** | | 0.311** |
| | | (0.137) | | (0.137) |
| *Sender characteristics (j):* | | | | |
| Household size | | 0.008 | | 0.008 |
| | | (0.008) | | (0.008) |
| Age of head | | 0.024** | | 0.025** |
| | | (0.010) | | (0.010) |
| Illiterate | | -0.897*** | | -0.910*** |
| | | (0.338) | | (0.336) |
| Formal schooling | | 0.049 | | 0.051 |
| | | (0.486) | | (0.497) |
| Female head | | -0.292 | | -0.244 |
| | | (0.665) | | (0.671) |
| Receive remittances | | 0.198 | | 0.212 |
| | | (0.305) | | (0.301) |
| Wealth level | | 0.461*** | | 0.465*** |
| | | (0.148) | | (0.150) |
| Observations | 57,060 | 57,060 | 57,060 | 57,060 |
| Households | 1,625 | 1,625 | 1,625 | 1,625 |

Note: Dependent variable: Amount of land $i$ receives from $j$. $A^*_{ij}$ denotes the predicted household-level efficiency-achieving transaction. $A^{*v}_{ij}$ denotes the predicted village-level efficiency-achieving transaction. All regressions include village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Column (2) splits the kinship variable into kin of the household head, kin of the wife, and marriage kin. The coefficient estimate on kin of the household head is large in size and statistically significant at the 1 percent level. In comparison, the coefficient estimate on relatives of the wife is smaller in magnitude and only statistically significant at the 10 percent level. The patrilocal nature of Gambian society likely explains why relatives of the wife (wives) are less important in determining land formation. The positive estimate of both confirms that market segmentation is caused by family ties related to the household head as well as the wife (wives) of the head. Being connected by marriage is not significant in the two models. We interpret this as evidence that not all family links are created equal: Some family connections are stronger and thus decrease transaction costs more.

As with social proximity, geographical proximity can help alleviate transaction costs, as households that reside close to each other have clear informational advantages. To ensure that the results are not driven by the smaller sample, the baseline regression is re-estimated on the smaller sample in column (3). Compared to the baseline results in Table 3, the sign and significance of the variables of interest are unchanged.

Columns (4) and (5) include link-specific regressors of whether two households are compound neighbors and share neighboring plots. Both measures of geographical proximity are positive and statistically significant. As expected, geographical proximity, especially plot proximity, positively affects the amount exchanged. The magnitude and significance of all other variables are unchanged.

Table 4: Regression results: Social proximity

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| $A^*_{ij}$ | 0.041*** | 0.041*** | 0.053*** | 0.051*** | 0.050*** |
| | (0.008) | (0.008) | (0.016) | (0.017) | (0.016) |
| Kin tie | 1.558*** | | 1.735*** | 1.560*** | 1.209*** |
| | (0.304) | | (0.286) | (0.260) | (0.274) |
| Kin of head | | 1.442*** | | | |
| | | (0.379) | | | |
| Kin of wife | | 0.982* | | | |
| | | (0.578) | | | |
| Marriage kin | | 0.696 | | | |
| | | (0.472) | | | |
| Neighbour compound | | | | 1.246*** | 0.731* |
| | | | | (0.356) | (0.401) |
| Neighbour plot | | | | | 2.270*** |
| | | | | | (0.409) |
| Add. expl. variables | Yes | Yes | Yes | Yes | Yes |
| Observations | 57,060 | 57,060 | 23,917 | 23,917 | 23,917 |
| Households | 1,625 | 1,625 | 624 | 624 | 624 |

Note: Dependent variable: Amount of land $i$ receives from $j$. $A^*_{ij}$ denotes the predicted household-level efficiency-achieving transaction. All regressions include village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

## Are social ties efficiency enhancing?

As the results above indicate, socially and geographically proximate households do indeed exchange more land. A natural follow-up question to ask is whether the presence of proximity effects is efficiency enhancing. We test this by including interaction terms between the proximity dummies and $A^*_{ij}$. The estimations in Table 5, column (1), show that the additional land that is transacted along kinship lines does not enhance efficiency. This is also true when disaggregating the kinship measure into its components (not shown). The absence of more efficiency-enhancing transactions between kin-related households as compared to transactions between non-kin households may be because it is harder to refuse to transact land with kin than with non-kin, and thus land between kin is exchanged for reasons other than to further increase efficiency.

According to column (2), geographical proximity, measured both in terms of having proximate plots and proximate compounds, increases the probability of conducting efficiency-enhancing transactions: More efficiency-enhancing land transactions take place between geographically proximate households, underlining the importance of trust and/or enforceability when conducting land exchanges.

The results show that not all social ties affect the land market in the same

Table 5: Efficiency and social proximity

| | (1) | | (2) | |
|---|---|---|---|---|
| $A^*_{ij}$ | 0.040*** | (0.010) | 0.027 | (0.021) |
| Kin tie | 1.551*** | (0.312) | 1.185*** | (0.276) |
| Neighbour compound | | | 0.636* | (0.367) |
| Neighbour plot | | | 2.239*** | (0.405) |
| $A^*_{ij}$×Kin tie | 0.002 | (0.009) | | |
| $A^*_{ij}$×Neighbour compound | | | 0.071*** | (0.023) |
| $A^*_{ij}$×Neighbour plot | | | 0.060** | (0.027) |
| Additional expl. variables | Yes | | Yes | |
| Observations | 57,060 | | 23,917 | |
| Households | 1,625 | | 624 | |

Note: Dependent variable: Amount of land i receives from j. $A^*_{ij}$ denotes the predicted household-level efficiency-achieving transaction. All regressions include village fixed effects. Standard errors are clustered on the village level. ***, **, * indicate significance at the 1, 5, and 10 percent levels, respectively.

way. They are consistent with geographical proximity having a cleaner effect on land transactions by lowering market imperfections, while land transactions between family members occur for strategic reasons rather than in order to increase efficiency.

## Robustness checks

In this section, we conduct a series of robustness checks. In relation to the description of the traditional tenure system in The Gambia, one concern is that transactions are driven by social ties to the *Alkalo*. Two underlying mechanisms might drive the results: (1) Households who are related to the *Alkalo* are more likely to receive land from other households, and (2) related households are likely to be favored by the *Alkalo* relative to other households residing in the village. To test the hypotheses explicitly, we include additional regressors on whether the sending or receiving household is related to the village *Alkalo*. The regression results shown in Table 6 indicate that households that are kin-related to the *Alkalo* receive less land, not more as one might have expected. The interaction with $A^*_{ij}$ in column (2) implies that relatives of the *Alkalo* send less land in efficiency-enhancing ways. In addition, the results shown in column (4) imply that two households that are both related to the *Alkalo* exchange less land in efficiency-enhancing ways when using $A^*_{ij}$.[22] Hence, households related to the *Alkalo* are not favored in terms of receiving more land. Furthermore, the

[22]When using $A^{*v}_{ij}$ as the predictor, the interaction effect become insignificant. These results are reported in the Appendix.

coefficients on our predictor $A^*_{ij}$ remain virtually unchanged when controlling for these effects.

A different type of concern is related to the assumptions underlying the theoretical model. All the households in a village are assumed to have the same production function. One concern is that households do not have the same production function if the products produced differ, as different crops may have different optimal land–labor input ratios. To investigate this further, we split the sample depending on the village's agricultural production. We categorize villages into two groups: villages that only cultivate groundnuts, and villages that cultivate at least two types of crops. Since groundnuts cannot be grown in the wet lowland along the Gambian River basin and are solely produced for the market, this group consists of mostly upland cash-crop-producing villages. The columns (5) and (6) in Table 6 show that the predicted efficiency-achieving transaction continues to be statistically significant and positive for both samples. In terms of magnitude, more efficiency-enhancing transactions take place in the lowland villages cultivating at least two types of crops. We take this as supportive evidence suggesting that household production functions do not vary to an extent that invalidates our results. Additional village splits by (i) land usage rights ownership, (ii) population density, and (iii) ethnic diversity are shown in the Appendix.

We also investigate the assumption of homogeneous within-village land quality. We believe that this is not too strong of an assumption by re-estimating equation 7 using a linear probability model (LPM) including household receiver and sender fixed effects. This is a viable strategy due to the dyadic nature of the dataset. The two-way fixed effect, however, wipes out all household-level variation, meaning that only our efficiency-achieving predictor and link-specific variables can be estimated. To increase comparability, the original model is re-estimated using an LPM, and results are reported in Table 6, column (7). The estimation results including sender and receiver fixed effects are shown in column (8). The results are remarkably consistent, though the coefficient estimates are smaller compared to the Table 3. The efficiency-achieving predictor is still positive and statistically significant, suggesting partial equalization of land–labor ratios.

Finally, the tobit model assumes that the same coefficients govern both the decision to transact and the size of the transaction. One possibility is to simply use OLS instead, which is always consistent. Another option is to to relax the assumption while still accounting for the high number of zero observations. One way of doing this would be to estimate a Heckman hurdle model. However, in the absence of variables that affect only the decision to transact and not the size, this model is identified only by the functional form. Instead, we estimate Cragg's (1971) hurdle model. The transaction decision is consistently estimated by probit. However, if the error terms of the two decisions are correlated, the size equation is not. Table A1.5 reports parameter estimates of the main variables. Sign, magnitude, and significance are consistent across the tobit and OLS models (columns (2) and (3)). The results from the probit regression also confirm the sign and significance of the main variables. However, using either a truncated normal hurdle model or a log-normal hurdle model to estimate the size equation yields no significant effects. Whether this is caused by the possible inconsistency of the second-stage regression or whether there are no effects in the second stage is unknown. Overall, the main result is robust to alternative specification techniques.

## Economic impact: Simulation results

In order to investigate the economic significance of the parameters estimated above, we simulate counterfactual land transfer networks by using point estimates reported in Table 5. We generate stochastic error terms and calculate desired transfer sizes for all links based on observed household and link characteristics. We sequentially ensure that all transfers are pairwise stable and add as a condition that households cannot send more land than their endowment plus any land they may have received. If a transfer is not pairwise stable taking all other transfers into account, the transfer value is adjusted. This is done until all links are pairwise stable. For each village, 100 simulations are carried out.[23] As a baseline scenario, the actual social networks are used. Counterfactual simulations where all social ties and where no social ties exist are also carried out. Finally, we simulate a scenario where there is no direct efficiency-enhancing effect, i.e., $A^*_{ij} = 0 \; \forall \, i, j$. Table 7 reports actual Gini indices observed in the data before and after land transactions have taken place alongside mean Gini indices of the simulations. Perfect allocative efficiency is achieved when land–labor ratios are equal across households. This corresponds to a Gini index of 0, while a Gini index of 100 corresponds to a scenario where a single household with a single member cultivates all land in the sample.

Comparing the actual Gini indices before and after transactions (a and b),

[23] The network converges to a pairwise stable equilibrium in 99.75 percent of all village simulations. Non-converging simulations are discarded.

Table 6: Robustness checks

| | The importance of being related to the Alkalo | | | | Village production | | Fixed effects | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Groundnuts | Various crops | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $A^*_{ij}$ | 0.044*** | 0.057*** | 0.044*** | 0.052*** | 0.026*** | 0.044*** | 0.0002** | 0.0003** |
| | (0.008) | (0.014) | (0.008) | (0.010) | (0.009) | (0.011) | (0.000) | (0.000) |
| Kinship ties | 1.612*** | 1.611*** | 1.603*** | 1.606*** | 1.055* | 1.766*** | 0.0104*** | 0.011*** |
| | (0.305) | (0.306) | (0.305) | (0.304) | (0.579) | (0.371) | (0.002) | (0.002) |
| $i$ is family related to *Alkalo* | -0.535** | -0.547** | | | | | | |
| | (0.248) | (0.253) | | | | | | |
| $j$ is family related to *Alkalo* | 0.211 | 0.254 | | | | | | |
| | (0.294) | (0.297) | | | | | | |
| $i$ is family related to *Alkalo* $\times A^*_{ij}$ | | 0.008 | | | | | | |
| | | (0.014) | | | | | | |
| $j$ is family related to *Alkalo* $\times A^*_{ij}$ | | -0.035* | | | | | | |
| | | (0.021) | | | | | | |
| Both family related to *Alkalo* | | | -0.353 | -0.314 | | | | |
| | | | (0.325) | (0.318) | | | | |
| Both family related to *Alkalo*$\times A^*_{ij}$ | | | | -0.025** | | | | |
| | | | | (0.012) | | | | |
| Additional control variables | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No |
| Receiver and sender fixed effects | No | No | No | No | No | No | No | Yes |
| Observations | 55,812 | 55,812 | 55,812 | 55,812 | 17,902 | 38,218 | 57,060 | 57,060 |
| Households | 1,609 | 1,609 | 1,609 | 1,609 | 514 | 1,068 | 1,625 | 1,625 |

Dependent variable: Amount of land $i$ receives from $j$. $A^*_{ij}$ denotes the predicted household-level efficiency-achieving transaction. All regressions include village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table 7: Land Ginis using counterfactual social network structures

| The social network: | | (1) Family | (2) Neighborship |
|---|---|---|---|
| Actual Gini indices | | | |
| - *(a) Before transactions* | | 70.57 | 68.85 |
| - *(b) After transactions* | | 68.06 | 60.58 |
| Mean simulated Gini indices using | | | |
| - *(c) Actual network* | | 67.35 | 59.69 |
| - *(d) No efficiency effect* | $(A^*_{ij} = 0,\ \forall i, j)$ | 67.45 | 59.92 |
| - *(e) All social ties* | $(w_{ij} = 1,\ \forall i, j)$ | 67.30 | 57.33 |
| - *(f) No social ties* | $(w_{ij} = 0,\ \forall i, j)$ | 67.29 | 59.87 |
| Observations (households) | | 1,625 | 624 |

Dependent variable: Simulations on the family and neighbor use parameter estimates from Table 5, column (1) and column (2), respectively. When changing the neighbor network, both neighbor plot and neighbor compound links are changed. $A^*_{ij}$ denotes the predicted household-level efficiency-achieving transaction.

land transactions do move the Gini downward towards a more allocatively efficient allocation, as expected. This effect is larger in the smaller sample, where neighborship link information is available (8.27 point decrease), than in the full sample (2.51 point decrease). The baseline simulations which use the actual social networks (c) are relatively close to the actual Gini coefficients after land transactions have been accounted for. The impact of the efficiency predictor combined with the existing social networks can be found as the difference between the baseline simulations and the counterfactual scenario where there is no effect from the efficiency predictor (d). This yields impacts of 0.1 and 0.23 Gini points, using the family and the neighbor network, respectively.

The land–labor distribution does not respond to allowing all family ties (e) or no family ties (f) to be present. This is consistent with the small interaction effect of kin ties and the efficiency predictor $A^*_{ij}$. Removing all neighborship links increases the Gini index by just 0.18 points (e-c). Constructing all neighborship links decreases the Gini index by 2.36 points (d-c). This is a substantial change. The reason for this is that while the parameter estimate of the interaction between the neighborship variables and the efficiency predictor is positive and much larger than the interaction with the family variable, the existing neighborship network is sparse: Only around 9 percent of all potential links exist. So while the size of the neighborship effect on allocative efficiency is large, the impact of the effect is small due to the sparsity of the neighborship network. However, if the efficiency-enhancing benefits of neighborship could be scaled up to hold between all households in terms of increased trust and ease of contract enforcement, it would have a substantial impact on allocative efficiency.

## The labor market

Land transactions are not the only factor that can be used to equalize factor ratios across households. Another option is to adjust the amount of labor. If the labor market is used to equalize land–labor ratios, one would expect households with low land–labor ratios to take up wage labor, while households with high land–labor ratios would hire workers. One can construct a similar set of predictions for labor transactions where land–labor ratios are based on the amount of land cultivated after all land exchanges have taken place. The reason for this is the sequential nature of land and labor exchanges described earlier, where labor exchange takes place after land allocation. According to Table A1.2 in the Appendix, heads of land-abundant households are more likely to both receive and send labor. This descriptive evidence suggests that mechanisms other than equalization of land–labor ratios may be driving the exchanges of labor.[24]

In order to more formally investigate this, we re-estimate equation 7 substituting the dependent variable to examine the labor market with labor transactions of the household head, measured in workdays. To ease interpretation in the sense that the expected signs correspond to the regression results presented on land exchanges, the dependent variable is now defined as the amount of labor household $i$ *sends* to $j$. The baseline results are presented in Table 8, while the control variables are shown in Table A1.3. Table A1.2 includes social and geographical proximity. The control variables included in all estimations are the same as those included in the land market regressions. Neither predictor is statistically significant at the 5 percent level.

Turning to the impact of preexisting social ties, we find that kin-related households exchange more labor (columns (1) and (6) in Table A1.4). This is in line with the literature on risk sharing where it is found that transfers that are performed in order to offset the impact of shocks often travel along family networks. Disaggregating kinship, we see that the effect of family relations on labor transactions can be attributed to kin effects of the head, the wife (wives), and marriage ties as well (columns (2) and (7)).

In accordance with expectations, geographical proximity of compounds and

[24] We obtain similar results if we use initial land endowment rather than the ex-post amount of land cultivated. It should be noted that due to data limitations, we do not consider labor hired from outside the village.

Table 8: Baseline regression results: Labor

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| $A^{*}_{ij}$ | -0.000 | -0.000 | | |
| | (0.000) | (0.000) | | |
| $A^{*v}_{ij}$ | | | 0.001* | 0.000 |
| | | | (0.001) | (0.001) |
| Same ethnicity | | 0.023*** | | 0.023*** |
| | | (0.005) | | (0.005) |
| Additional expl. variabels | No | Yes | No | Yes |
| Observations | 57,060 | 57,060 | 57,060 | 57,060 |

Dependent variable: Amount of labor $i$ sends to $j$. $A^{*}_{ij}$ denotes the predicted household-level efficiency-achieving transaction. $A^{*v}_{ij}$ denotes the predicted village-level efficiency-achieving transaction. All regressions include village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

agricultural plots decreases the monitoring and enforcement costs associated with labor: Households located close to each other are more likely to be linked through labor exchange. However, while preexisting social ties increase labor market access, these transactions do not flow in such a way as to increase allocative efficiency. This is evident from the insignificant interaction terms between our predictors of allocative efficiency and the link-specific characteristics reported in Table A1.4.

# 6 Conclusion

In this paper, we tested the importance of social ties in land markets using a dataset of social networks and land transactions. We provided a theoretical framework that yields testable predictions of whether preexisting social networks increase allocative efficiency or whether they lead to market segmentation where some households get preferential access to land. We operationalized the concept of allocative efficiency and tested our predictions using a dyad-level analysis. The empirical analysis was based on complete network data covering farmers in 52 rural villages in The Gambia and yields three main findings:

First, land markets in The Gambia do not fully equalize land–labor ratios due to costs and risks associated with land transfers. However, land transactions do flow in the predicted efficiency-enhancing direction. Second, interpersonal relations in terms of social and geographical proximity are found to increase the amount of land transacted. This results in market segmentation, where land follows the paths where the costs are lowest. but not necessarily where it improves allocative efficiency the most. Third, whereas both geographical ties as well as family ties give rise to market segmentation, only geographically proximate households conduct more efficiency-enhancing transactions, whereas transactions between family-related households do not appear to be efficiency-improving beyond transactions between non-kin. We interpret this finding as evidence that other aspects also influence land exchanges, particularly between households with family ties. We attribute this to strategic transactions with kin-related households. This leads to more inefficient transactions between kin as compared to transactions conducted between geographically close households.

Simulations of the network structure revealed that the efficiency-enhancing impact of neighborship was limited. This is found to be due to the sparsity of the existing network. There would be substantial allocative efficiency benefits if one could extend the benefits that neighborship brings to all links. This would require increased trust or reduced enforcement and monitoring costs as these are proposed channels through which neighborship induces efficiency-improving transactions. Reducing these costs through strengthening the institutional and judicial framework would improve efficiency, even in the absence of land redistribution.

The results show that there are substantial imperfections in the rural Gambian land market. Furthermore, they suggest that the preexisting social networks have limited consequences for land market efficiency. This does not mean that social ties are unimportant for land transfers. In particular, the combination of unequal land endowments and the indigenous tenure system means that kinship ties create a land exchange network where some households have easier access to land than others and are more likely to receive land, even though these transactions are not necessarily efficiency enhancing.

To investigate the strength of the proposed approach and main findings, the model was re-estimated for the labor market. We find that labor market transactions do not contribute to equalization of land–labor ratios. This is consistent with other evidence that labor transactions are used to smooth supply over shorter periods, and that the rural labor market is even less well-functioning than the land market and therefore ill-suited to correct land–labor imbalances between households.

# Appendix

This appendix contains supplementary theoretical and empirical results, including network illustrations and evidence related to critical assumptions, to accompany the paper *"Efficiency of land markets: Network level evidence of the importance of social ties"*.

## Overview

This appendix presents the following materials to supplement the paper:

- Additional tabels to directly supplement the main text
- Illustration of village location in The Gambia
- Illustrations of kinship and land networks for a single village
- Empirical estimation results for the village-level efficiency-achieving predictor
- Household-level test of allocative efficiency (Skoufias, 1995)
- Robustness checks:
  - The importance of being related to the *Alkalo*: $A_{ij}^{v*}$
  - Analysis of the sensitivity of the results: sample splits based on village characteristics
  - The extent of technical inefficiency if allocative efficiency holds
  - Estimation results using different estimation techniques

# A1 Additional tabels

Table A1.1: Descriptive statistics: Dyad-level

| | Mean | Std. Dev. | Min. | Max. | Obs. |
|---|---|---|---|---|---|
| Receives land dummy | 0.010 | 0.100 | 0 | 1 | 57,060 |
| Send labour dummy | 0.015 | 0.122 | 0 | 1 | 57,060 |
| $A^*_{ij}$ | 2.747 | 9.856 | 0 | 237.2 | 57,060 |
| $A^{*v}_{ij}$ | 0.893 | 3.835 | 0 | 133.6 | 57,060 |
| Kinship tie | 0.132 | 0.338 | 0 | 1 | 57,060 |
| Kin of head | 0.059 | 0.235 | 0 | 1 | 57,060 |
| Kin of wife | 0.027 | 0.161 | 0 | 1 | 57,060 |
| Marriage kin | 0.022 | 0.147 | 0 | 1 | 57,060 |
| Neighbour compound | 0.093 | 0.293 | 0 | 1 | 23,917 |
| Neighbour plot | 0.094 | 0.292 | 0 | 1 | 23,917 |
| Same ethnicity | 0.715 | 0.451 | 0 | 1 | 57,060 |

Note: $A^*_{ij}$ denotes the predicted household-level efficiency-achieving transaction. $A^{*v}_{ij}$ denotes the predicted village-level efficiency-achieving transaction. Information on neighboring compound and neighboring plot is only available for a subsample of the data, corresponding to at least 20 villages.

Table A1.2: Labor market participation rates by land–labor ratios (ex-post land transactions)

| | All | Landless | 0.1-0.6 ha/w | 0.6-1.6 ha/w | 1.6-3.0 ha/w | > 3.0 ha/w |
|---|---|---|---|---|---|---|
| % in land market | 54.4 | 42.5 | 49.8 | 56.6 | 64.3 | 57.5 |
| % Labor sender | 36.9 | 30.6 | 34.1 | 39.5 | 43.6 | 34.7 |
| % Labor receiver | 32.3 | 23.8 | 27.5 | 30.6 | 41.0 | 39.8 |
| Observations | 1,625 | 294 | 255 | 516 | 266 | 294 |

Note: ha/w correspond to the number of hectares per working adult.

Table A1.3: Regression results labour: control variables

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| $A^{*}_{ij}$ | -0.000 | -0.000 | | |
| | (0.000) | (0.000) | | |
| $A^{*v}_{ij}$ | | | 0.001* | 0.000 |
| | | | (0.001) | (0.001) |
| Same ethnicity | | 0.023*** | | 0.023*** |
| | | (0.005) | | (0.005) |
| *Sender characteristics (i):* | | | | |
| Household size | | 0.000 | | 0.000 |
| | | (0.000) | | (0.000) |
| Age of head | | -0.000* | | -0.000* |
| | | (0.000) | | (0.000) |
| Illiterate | | -0.006* | | -0.006* |
| | | (0.003) | | (0.003) |
| Formal schooling | | 0.002 | | 0.002 |
| | | (0.004) | | (0.004) |
| Female head | | 0.009 | | 0.009 |
| | | (0.007) | | (0.008) |
| Receive remittances (dummy) | | -0.003 | | -0.003 |
| | | (0.003) | | (0.003) |
| Wealth level | | 0.002 | | 0.002 |
| | | (0.002) | | (0.002) |
| *Receiver characteristics (j):* | | | | |
| Household size | | 0.000*** | | 0.000*** |
| | | (0.000) | | (0.000) |
| Age of head | | 0.000*** | | 0.000*** |
| | | (0.000) | | (0.000) |
| Illiterate | | -0.005 | | -0.005 |
| | | (0.003) | | (0.003) |
| Formal schooling | | -0.018*** | | -0.018*** |
| | | (0.006) | | (0.006) |
| Female head | | -0.016** | | -0.016** |
| | | (0.008) | | (0.008) |
| Receive remittances (dummy) | | 0.004 | | 0.004 |
| | | (0.003) | | (0.003) |
| Wealth level | | 0.004* | | 0.004* |
| | | (0.002) | | (0.002) |
| Observations | 57,060 | 57,060 | 57,060 | 57,060 |

Note: Dependent variable: Amount of labor $i$ sends to $j$. Equivalent to Table 6 in the main text including control variables. All regressions include village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table A1.4: Labor: Social ties

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| $A^*_{ij}$ | -0.000 | 0.000 | -0.000 | -0.000 | -0.000 | -0.000 | -0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Kin tie | 0.046*** | | 0.037*** | 0.035*** | 0.030*** | 0.046*** | |
| | (0.008) | | (0.009) | (0.008) | (0.007) | (0.008) | |
| Kin of head | | 0.047*** | | | | | 0.030*** |
| | | (0.009) | | | | | (0.007) |
| Kin of wife | | 0.022*** | | | | | 0.009*** |
| | | (0.007) | | | | | (0.003) |
| Marriage kin | | 0.012** | | | | | 0.025*** |
| | | (0.006) | | | | | (0.007) |
| Neighbour compound | | | | 0.014*** | 0.009*** | | |
| | | | | (0.004) | (0.003) | | |
| Neighbour plot | | | | | 0.026*** | | |
| | | | | | (0.007) | | |
| $A^*_{ij}\times$Kin tie | | | | | | 0.000 | |
| | | | | | | (0.001) | |
| $A^*_{ij}\times$Neighbour compound | | | | | | | -0.000 |
| | | | | | | | (0.001) |
| $A^*_{ij}\times$Neighbour plot | | | | | | | -0.000 |
| | | | | | | | (0.000) |
| Add. expl. variables | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 57,060 | 57,060 | 23,917 | 23,917 | 23,917 | 57,060 | 23,917 |

Note: Dependent variable: Amount of labor $i$ sends to $j$. $A^*_{ij}$ denotes the predicted household-level efficiency-achieving transaction. All regressions include village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

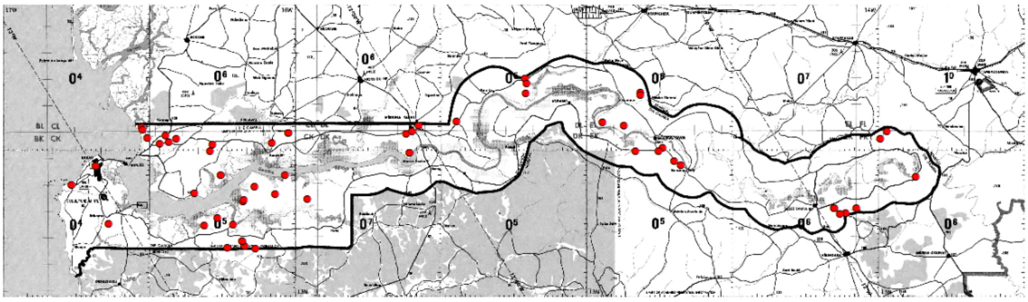
Table A1.5: Robustness checks: Estimation techniques

| | (1) | (2) | (3) | (4) | | (5) | |
|---|---|---|---|---|---|---|---|
| | Tobit | Tobit | OLS | Truncated normal hurdle | | Log-normal hurdle | |
| | | Partial effects | | Probit | Truncated regression | Probit | OLS (log) |
| $A^*_{ij}$ | 0.041*** | 0.000*** | 0.001** | 0.008*** | 0.010 | 0.008*** | -0.000 |
| | (0.008) | (0.000) | (0.000) | (0.001) | (0.015) | (0.001) | (0.004) |
| Kinship tie | 1.558*** | 0.016*** | 0.020*** | 0.304*** | -0.160 | 0.304*** | 0.052 |
| | (0.304) | (0.000) | (0.005) | (0.054) | (0.365) | (0.054) | (0.070) |
| | (6) | (7) | (8) | (9) | | (10) | |
| $A^{*v}_{ij}$ | 0.073*** | 0.001*** | 0.002** | 0.014*** | 0.013 | 0.014*** | 0.001 |
| | (0.020) | (0.000) | (0.001) | (0.004) | (0.034) | (0.004) | (0.006) |
| Kinship tie | 1.566*** | 0.016*** | 0.020*** | 0.305*** | -0.151 | 0.305*** | 0.051 |
| | (0.305) | (0.000) | (0.005) | (0.054) | (0.375) | (0.054) | (0.071) |
| Observations | 57,060 | 57,060 | 57,060 | 57,060 | | 57,060 | 579 |

Note: Dependent variable: Amount of labor $i$ sends to $j$, except for the truncated (log-)normal hurdle estimations where the dependent variable is equal to 1 if $i$ receives land from $j$, and 0 otherwise. $A^*_{ij}$ denotes the predicted household-level efficiency-achieving transaction. $A^{*v}_{ij}$ denotes the predicted village-level efficiency-achieving transaction. All regressions include additional explanatory variables and village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

# A2 Location of the surveyed villages in The Gambia

Figure A2.1: Location of surveyed villages



# A3 Kinship and land network characteristics in a single village

To give an idea of the nature of the data, Figures A3.1 and A3.3 illustrate the family and land networks in a specific village. Figure A3.1 shows all family links. Figure A3.2 shows all land transactions.

Figure A3.1: The kinship network in a single village

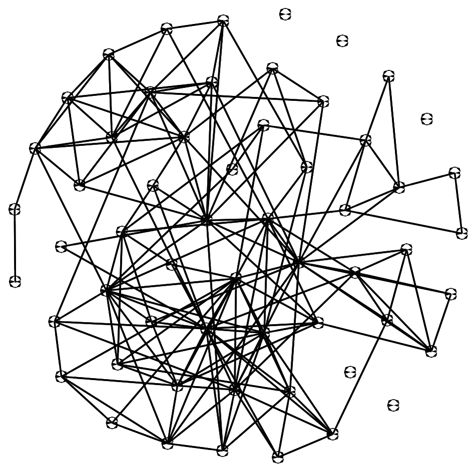


Figure A3.2: The land network in a single village

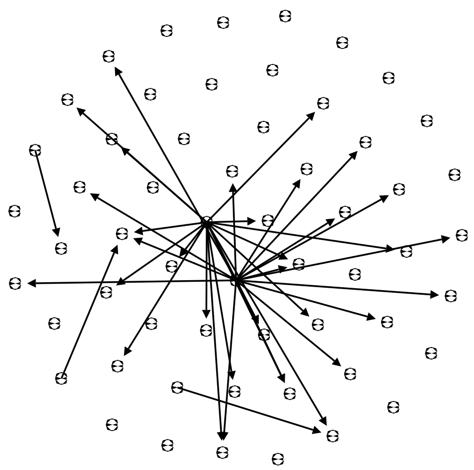
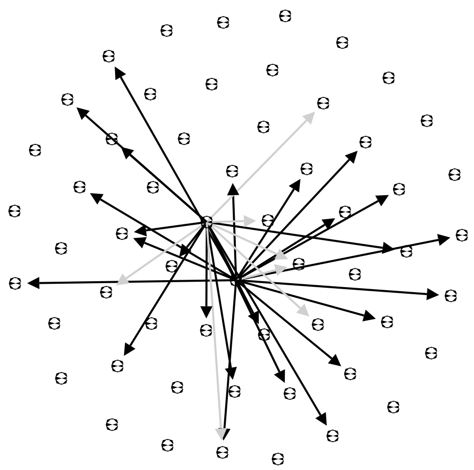


Figure A3.3: Land and family village network of land senders



Black links are land transactions. Gray links are where both a family tie exists and a land transaction takes place.

Two things are worth noting: First, very few land transactions do not originate from one of the two main senders. This corresponds well to the stylized description of the first settlers and the *Alkalo* being the ones who control primary usage rights and distribute secondary usage rights. Second, the family network is quite dense in the village: Almost all households are connected to other households through the family network (blue and red links). Figure A3.3 shows only the network of those connected to the two main land-sending households. This figure illustrates that the land-sending households send land to a relatively large subset of village households, but not to all. Few households receive land from both of the two main sender households. There is some overlap between the households that receive land and the family network of senders. However, this overlap is not complete: Many households are related through family ties to the two main sender households yet receive no land, and many others are not related to the sender households but do receive land.

# A4 Empirical results: Village level efficiency-achieving transactions

Table A4.1 shows market segmentation and whether transaction costs decrease in social and geographical proximity using the village-level efficiency-achieving predictor, $A^{*v}_{ij}$. The results are similar to the regressions in the main paper

Table A4.1: Market segmentation: proximity

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| $A^{*v}_{ij}$ | 0.073*** | 0.075*** | 0.099** | 0.095** | 0.098*** |
| | (0.020) | (0.021) | (0.040) | (0.037) | (0.037) |
| Kin tie | 1.566*** | | 1.770*** | 1.592*** | 1.236*** |
| | (0.305) | | (0.295) | (0.267) | (0.278) |
| Kin of head | | 1.457*** | | | |
| | | (0.382) | | | |
| Kin of wife | | 0.988* | | | |
| | | (0.584) | | | |
| Marriage kin | | 0.718 | | | |
| | | (0.474) | | | |
| Neighbor comp. | | | | 1.261*** | 0.740* |
| | | | | (0.354) | (0.400) |
| Neighbor plot | | | | | 2.291*** |
| | | | | | (0.412) |
| Add. expl. variables | Yes | Yes | Yes | Yes | Yes |
| Observations | 57,060 | 57,060 | 23,917 | 23,917 | 23,917 |

Note: Dependent variable: Amount of land $i$ receives from $j$. $A^{*}v_{ij}$ denotes the predicted household-level efficiency-achieving transaction. All regressions include village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

using $A^{*}_{ij}$: Kin-related households, as well as neighboring households, have easier access to land. Table A4.2 demonstrates whether the presence of proximity effects can be efficiency enhancing by increasing the number of feasible efficiency-enhancing transactions. Again, the results are similar to those obtained when using $A^{*}_{ij}$: Only geographical proximity increases the probability of conducting efficiency-enhancing transactions.

# A5 Skoufias's (1995) test of allocative efficiency

As noted in the paper, a standard test by Skoufias (1995) has been extensively employed in testing the efficiency of land transaction sizes at the household level.

The idea of this test can be summarized as follows: Due to market imperfections, the (unobserved) net demand for land in the absence of land market imperfections ($y^{*}$) will differ from the observed amount of land that is used after transactions ($y$). We can express this as $y = h(y^{*})$. If the slope of $h$ is equal to 1, an increase in the desired area of production by 1 will result in

Table A4.2: Efficiency and social proximity

| | (1) | | (2) | |
|---|---|---|---|---|
| $A_{ij}^{*V}$ | 0.075*** | (0.028) | 0.060* | (0.034) |
| Kin tie | 1.572*** | (0.312) | 1.207*** | (0.283) |
| Neighbor compound | | | 0.596* | (0.358) |
| Neighbor plot | | | 2.322*** | (0.417) |
| $A_{ij}^{*V}$ ×Kin tie | -0.006 | (0.026) | | |
| $A_{ij}^{*V}$ ×Neighbor compound | | | 0.269*** | (0.073) |
| $A_{ij}^{*V}$ ×Neighbor plot | | | -0.021 | (0.105) |
| Additional expl. variables | Yes | | Yes | |
| Observations | 57,060 | | 23,917 | |

Note: Dependent variable: Amount of land $i$ receives from $j$. $A^{*}v_{ij}$ denotes the predicted household-level efficiency-achieving transaction. All regressions include village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

an increase in the cultivated area by 1. Furthermore, the intercept of function $h()$ may be positive due to the presence of fixed costs: A certain level of land must be desired before any transaction takes place. The desired area that a household wants to cultivate in the absence of market imperfections (DCA), together with the land endowment (LAND), determines the unobserved ideal net land transaction, $y^{*}$: $y^{*} = DCA - LAND$. The DCA is determined by the levels of the other production factors. For simplicity of exposition, assume that the only other production factor is labor: $DCA = f(L)$. Using these three expressions and carrying out a first-order Taylor expansion, we get

$$y = h' * f'' * L - h' * LAND \tag{A.1}$$

As transaction costs may differ between the supply and demand sides of the market, the land coefficient is allowed to vary between senders ($n$) and receivers ($p$). The model that is estimated is then (allowing for variables other than labor to enter into the $DCA$)

$$Y_{ik} = \begin{cases} \alpha_{nk} + \lambda_n * land_i + \beta_n^{'} Z_i + \epsilon_i & \text{if } \epsilon_i < \alpha_{nk} - \lambda_n * land_i - \beta_n^{'} Z_i \\ 0 & \text{if } \alpha_{nk} - \lambda_n * land_i - \beta_n^{'} Z_i \leq \epsilon_i \\ & \leq \alpha_p - \lambda_p * land_i - \beta_p^{'} Z_i \\ \alpha_{pk} + \lambda_p * land_i + \beta_p^{'} Z_i + \epsilon_i & \text{if } \epsilon_i > \alpha_{pk} - \lambda_p * land_i - \beta_p^{'} Z_i \end{cases} \tag{A.2}$$

where $i$ denotes the household in village $k$, and $Z_i$ is a vector of household characteristics, including the household's physical endowment, human capital endowment, social network, and status variables. (Skoufias, 1995). The physical endowment of the household includes the land area (ha.) owned with usage rights, household member composition, and whether the household owns cattle.[25] To account for the human capital endowment of households, which is likely both to impact farm productivity and non-farm sources of income, a dummy for whether the head of the household is illiterate, as well as the age and gender of the household head, is included. Since household status in rural Gambia is highly correlated with household wealth, the four proxy variables included for household status should also be seen as imperfect proxies for the household's wealth-position in the village. Two variables cover marital status: One indicates whether the household head is unmarried and another indicates polygamy. The third variable included to take into account household status and level of wealth is the number of corrugated huts. Fourth, a dummy variable for whether the household head is the *Alkalo* is included. In an expanded regression, we can see whether having family ties affects the amount of land that the household receives, by including a dummy that is equal to 1 if the household has a family tie in the village. Finally, village-level dummies are included to control for village-specific unobserved differences across villages, such as agro-climatic factors like soil quality affecting farm productivity, relative prices, access to markets, density of village population,[26] off-farm sources of income, and village-specific lease customs. Under the assumption that $\epsilon \sim N(0, \sigma^2)$, this can be thought of as a two-sided tobit model and can be estimated using maximum likelihood. Testing $\lambda_n = \lambda_p = -1$ provides a direct test of whether $h'$ is equal to 1, and thus whether there is allocative efficiency. The results are reported in A5.1, where the signs of the sender equation have been switched in order to ease interpretation.

The coefficients on the land variable have the expected sign, but are significantly different from -1. Thus, this standard test rejects allocative efficiency, which lends credibility to the estimation in the main text. A household is more likely to send, and will on average send more, if it has a large land endowment, and is more likely to receive land if it has a small land endowment. In line

[25]Unfortunately, the data at hand do not include information on the household's endowment of draft animal power, and we therefore apply a (somewhat imperfect) proxy for whether the household pays cattle tax under the presumption that households that pay cattle tax are also more likely to own bullocks.

[26]Jin and Jayne (2013) find that rental market activity varies considerably across regions.

Table A5.1: Net land leased in

| | (1) | | (2) | |
|---|---|---|---|---|
| | Send | Receive | Send | Receive |
| Land endowment (hac) | 0.056*** | -0.033* | 0.055*** | -0.032* |
| | (0.017) | (0.018) | (0.017) | (0.018) |
| Labor (active adult members) | -0.019* | 0.019 | -0.019* | 0.018 |
| | (0.011) | (0.017) | (0.011) | (0.017) |
| Kin of head | | | 4.018*** | -0.984* |
| | | | (1.317) | (0.585) |
| No of nonworking members | 0.135*** | 0.062** | 0.130*** | 0.064** |
| | (0.031) | (0.028) | (0.032) | (0.028) |
| No long-term sick members | -0.799** | 0.492 | -0.717** | 0.475 |
| | (0.336) | (0.322) | (0.335) | (0.318) |
| Female headed household | 0.016 | -1.547** | 0.147 | -1.521** |
| | (0.973) | (0.647) | (1.007) | (0.639) |
| Corrugated hut | 0.452*** | -0.022 | 0.425*** | -0.012 |
| | (0.161) | (0.134) | (0.158) | (0.134) |
| *Alkalo* (village chief) | 7.417*** | -1.510 | 7.414*** | -1.593 |
| | (1.454) | (1.682) | (1.443) | (1.712) |
| Age of Head | 0.022 | -0.012 | 0.024* | -0.012 |
| | (0.014) | (0.010) | (0.014) | (0.010) |
| Illiterate | -0.967* | 0.015 | -0.942* | 0.000 |
| | (0.536) | (0.409) | (0.537) | (0.409) |
| Formal schooling | -0.048 | -0.644 | -0.073 | -0.641 |
| | (0.786) | (0.624) | (0.768) | (0.616) |
| Unmarried | 1.007 | 1.311 | 0.937 | 1.321 |
| | (1.321) | (0.902) | (1.281) | (0.916) |
| Polygamous (>1 wife) | -0.731 | 0.309 | -0.809 | 0.327 |
| | (0.505) | (0.321) | (0.507) | (0.319) |
| Ethnicity: Fula | 0.924 | -0.299 | 0.693 | -0.275 |
| | (1.243) | (1.061) | (1.226) | (1.034) |
| Ethnicity: Wollof | -1.250 | 0.509 | -1.329 | 0.419 |
| | (1.174) | (0.773) | (1.150) | (0.737) |
| Village fixed effects | Yes | Yes | Yes | Yes |
| Observations | 1,625 | | 1,625 | |

Note: Maximum likelihood estimates. Dependent variable: Net land rented in. To ease interpretation, all the coefficients in the sending regressions are multiplied by -1. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

with the description of rural Gambian society, the *Alkalo* households send significantly more land. Regression (2) shows that even at the household level, there is evidence that family links matter: Households are more likely to both send land and receive land (significant only at the 10 percent level) if they have family ties in the village. While it is not possible to dig deeper into this effect using household-level data, it is consistent with the main results using dyad-level data.

# A6 Robustness checks

This section reports additional results related to the robustness checks in the main text. Section A6 discusses estimation results using the village-level efficiency-achieving predictor. The comparable results for the household efficiency-achieving predictor are reported as robustness checks in the main text. Next, Section A6 develops a measure for the extent of technical inefficiency under allocative efficiency. Finally, Section A6 investigates the sensitivity of the results by splitting the sample along different dimensions, including land usage rights ownership, village production, population density, and ethnic diversity.

## The importance of being related to the *Alkalo:* Village level efficiency-achieving transactions

Table A6.1 reports estimation results using the village-level efficiency-achieving predictor. The regression results shown in column (1) indicate that households that are kin related to the *Alkalo* receive less land, not more as one might have expected. The interaction $A_{ij}^{*v}$ in column (2) implies that relatives of the *Alkalo* send less land in efficiency-enhancing ways. In contrast to the result for $A_{ij}^{*}$ presented in the main text, both households being related to the *Alkalo* has no impact on the probability to exchange land. Hence, households related to the *Alkalo* are not favored in terms of receiving more land. Furthermore, the coefficients on our predictors $A_{ij}^{*v}$ remain virtually unchanged when controlling for these effects.

## Sensitivity analysis using sample splits

In the subsequent analysis, we investigate the sensitivity of the results by splitting the sample. A truncated normal hurdle model is estimated; however,

Table A6.1: The importance of being related to the *Alkalo*

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| $A^{*v}_{ij}$ | 0.081*** | 0.095*** | 0.082*** | 0.097*** |
| | (0.022) | (0.032) | (0.022) | (0.032) |
| Kinship ties | 1.614*** | 1.617*** | 1.605*** | 1.609*** |
| | (0.307) | (0.307) | (0.306) | (0.307) |
| $i$ is family related to Alkalo | -0.524** | -0.538** | | |
| | (0.247) | (0.250) | | |
| $j$ is family related to Alkalo | 0.233 | 0.264 | | |
| | (0.287) | (0.289) | | |
| $i$ is family related to Alkalo $\times A^{*v}_{ij}$ | | 0.028 | | |
| | | (0.036) | | |
| $j$ is family related to Alkalo $\times A^{*v}_{ij}$ | | -0.063* | | |
| | | (0.033) | | |
| Both family related to Alkalo | | | -0.321 | -0.292 |
| | | | (0.321) | (0.315) |
| Both family related to Alkalo$\times A^{*v}_{ij}$ | | | | -0.068 |
| | | | | (0.044) |
| Additional expl. variables | Yes | Yes | Yes | Yes |
| Observations | 55,812 | 55,812 | 55,812 | 55,812 |

Note: Truncated normal hurdle model. Only the first stage (probit) is reported. Dependent variable: Equal to 1 if $i$ receives land from $j$, and 0 otherwise. $A^{*v}_{ij}$ denotes the predicted village-level efficiency-achieving transaction. Due to missing data on whether a household is related to the *Alkalo*, the sample is reduced to 1,609 households. All estimations include village fixed effects. Standard errors are clustered on the village level. ***, **, * indicate significance at the 1, 5, and 10 percent levels, respectively.

only the first stage of the model is reported in the Table A6.2 as the efficiency-achieving predictors are insignificant in the second stage. All estimations are run on the whole sample to ensure a sufficient sample size, and thus power, of the estimates.

### Majority owners

The complex indigenous land tenure system creates a highly unequal allocation of land rights within rural Gambian villages and is a main reason why the texture of land reallocation is essential. To investigate whether social proximity alleviates transaction costs more in villages characterized by high land ownership inequality, we split the sample into two categories: villages where land belong to the first settlers and villages where land ownership is mixed. This also allows us to examine whether greater inequality in land ownership leads to more efficiency-enhancing transactions compared to mixed ownership where a larger number of villagers own land. The estimation results are shown in Table A6.2, column (A). There is only weak evidence that land flows in efficiency-enhancing directions in villages where land ownership is mixed (the predictors are only significant at the 10 percent level). Furthermore, kinship ties do not have a significant effect in these villages, indicating that the rural ownership structure does influence the role of social networks in land allocation. However, these findings could be due to the relatively few villages in the sample that have mixed ownership of land. This indicates that the results in the main text are most relevant for villages in which the traditional tenure system is still in place, which is the majority of the analyzed villages.

### Village production

In the theoretical model underlying the empirical analysis, it is assumed that all households have the same production function. This is not necessarily the case if the products produced differ, as different crops may have different optimal land–labor input ratios. Therefore, we split the sample depending on the village's agricultural production. We categorize villages into two groups: villages that only cultivate groundnuts, and villages that cultivate at least two types of crops. Column (B) in Table A6.2 shows that the predicted efficiency-achieving transaction continues to be statistically significant and positive for both samples. In terms of magnitude, more efficiency-enhancing transactions take place in the lowland villages cultivating at least two types of crops. Interestingly, kinship ties are only weakly significant as a determinant of land

Table A6.2: Land transactions by land ownership status

| | (A) Land ownership | | (B) Village production | | (C) Population density | | (D) Ethnic diversity | |
|---|---|---|---|---|---|---|---|---|
| | First setlers | Mixed | Ground-nuts | Various crops | Low | High | Hetero-geneous | Homo-genous |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $A^*_{ij}$ | 0.046*** | 0.030* | 0.026*** | 0.044*** | 0.068** | 0.040*** | 0.042*** | 0.043*** |
| | (0.011) | (0.017) | (0.009) | (0.011) | (0.035) | (0.008) | (0.009) | (0.012) |
| Kin tie | 1.721*** | 0.787 | 1.055* | 1.766*** | 0.602** | 2.058*** | 1.503*** | 1.615*** |
| | (0.321) | (0.944) | (0.579) | (0.371) | (0.260) | (0.414) | (0.472) | (0.397) |
| | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) |
| $A^{*v}_{ij}$ | 0.065*** | 0.120* | 0.037*** | 0.109*** | 0.097* | 0.076*** | 0.103** | 0.069*** |
| | (0.019) | (0.062) | (0.010) | (0.034) | (0.058) | (0.023) | (0.046) | (0.020) |
| Kin tie | 1.742*** | 0.780 | 1.056* | 1.781*** | 0.588** | 2.068*** | 1.503*** | 1.638*** |
| Additional expl. variables | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 45,884 | 9,616 | 17,902 | 38,218 | 17,397 | 39,663 | 25,289 | 31,771 |

Note: Truncated normal hurdle model. Only the first stage (probit) is reported. Dependent variable: Equal to 1 if $i$ receives land from $j$, and 0 otherwise. $A^*_{ij}$ denotes the predicted household-level efficiency-achieving transaction. $A^{*v}_{ij}$ denotes the predicted village-level efficiency-achieving transaction. All estimations include village fixed effects. Standard errors are clustered on the village level. ***, **, * indicate significance at the 1, 5, and 10 percent levels, respectively.

allocation in groundnut-producing villages.

### Population density

Land rental markets in African countries are argued to be more active in densely populated areas. However, the impact on efficiency is largely unknown (Holden, Keijiro and Place, 2009; Jin and Jayne, 2013).[27] We split the sample by arable land per villager to examine whether areas with a higher population density are characterized by more efficiency-enhancing transactions. The cutoff chosen is the median, corresponding to 1.87 ha per person.[28] Table A6.2, column (C) reports the estimation results by population density. The coefficient estimates indicate that transactions are only significantly efficiency-enhancing at the 10 percent level in low-population-density villages, but are significantly so at the 1 percent level in high-density villages. The magnitudes of the coefficients correspond to previous findings. Furthermore, kinship ties are found to be of importance in both kinds of villages.

### Ethnic diversity

If there is less trust between different ethnic groups, and fewer social and economic links are created in ethnically diverse societies, as suggested by Easterly (2001), fewer land transactions should take place in ethnically heterogeneous villages. We therefore split the sample into villages with high and low levels of ethnic diversity. The cutoff is the median of the within-village Herfindahl concentration indices. Contrary to expectations, the point estimates of column (D) of Table A6.2 indicate that more efficiency-enhancing land transactions take place in ethnically heterogeneous villages. We cannot, however, reject that the coefficients are equal across heterogeneous and homogeneous villages. In both types of villages, households exchange land based on family ties.

## The extent of technical inefficiency if allocative efficiency holds

Under allocative efficiency, the marginal rates of technical substitution must be equal between households. If this is not the case, it means that we can increase production by re-allocating inputs, thus achieving a more efficient outcome. To

[27] Jin and Jayne (2013) find that village population density has no determining impact on the net land rented in or out in Kenya.

[28] The estimation results are not sensitive to cutoffs in the same range.

illustrate, we assume a simple Cobb–Douglas production function where $i$ and $j$ are two households in the same village, $T_i$ is the total factor productivity, and $\theta_i^A$ and $\theta_i^l$ are individual specific land- and labor-enhancing parameters. $\theta_i^A$ can be seen as an indicator of the quality of household $i$'s land, and $\theta_i^l$ can be seen as a measure of human capital. Both of these are unobserved:

$$y_i = T_i(\theta_i^A A_i)^\alpha (\theta_i^l L_i)^{1-\alpha} \tag{A.3}$$

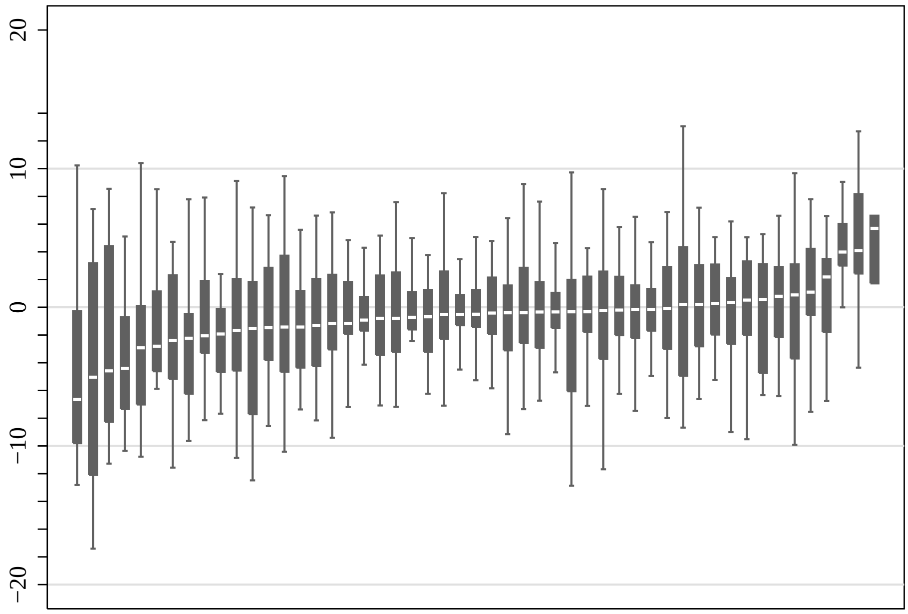
If there is allocative efficiency, the marginal rates of substitution between all households are equal to each other; i.e., for two households $i$ and $j$, we must have that

$$\frac{\theta_i^L}{\theta_i^A}\frac{A_i}{L_i} = \theta_i \frac{A_i}{L_i} = \theta_j \frac{A_j}{L_j} \tag{A.4}$$

where $\theta_i = \theta_i^A/\theta_i^L$ is the relative level of human capital to land quality for household $i$. We can calculate the relative $\theta_i$ for each household $\theta_i = (\frac{A_i}{L_i})/(\frac{A}{L})^v$, where $(\frac{A}{L})^v$ is the village-level average land–labor ratio of households in the baseline estimation sample. This gives us a measure of how different the relative skill-to-land-quality levels must be if the allocation after transactions have taken place is efficient. Note that in this model, land–labor ratios should only differ if a household has high human capital *and* low land quality or vice versa. Figure A6.1 shows the relative $\theta$ by village as a boxplot on a logarithmic scale. $\theta_i$ is only calculated for households who have a positive amount of land available after transactions have taken place, and the logarithmic transformation means that only households who have a positive amount of labor are shown in the figure.

The figure shows that in most villages, the majority of households has a lower land–labor ratio than the village average. Furthermore, the land–labor ratios exhibit large inter-village variation: In the median village, $\theta_i$'s differs by a factor of 3.5 between the 25th and the 75th percentile. This is substantial variation: For two households whose land quality is the same, it means that the labor from the more efficient household is equivalent to 3.5 times a unit of labor from the less efficient household. Taking into account that land quality differences are likely to be small within villages, as the distances between the plots are small, such large variation appears unrealistic. Expanding the hypothesis of efficiency to all households in the baseline estimation sample, households' relative $\theta$'s should vary by a factor of over 42. In the small, demarcated rural communities considered here, these differences in relative

Figure A6.1: Relative land-labor ratios by village (base-10 logarithmic scale)




Note: Boxes indicate 25th and 75th percentiles of distribution; whiskers indicate upper and lower values.

land quality and human capital are unreasonably large, especially considering that there is no reason to believe a priori that land and human capital should be negatively correlated. Hence, heterogeneity in household-level parameters, such as managerial abilities and soil quality, are unable to explain the large differences within-village land–labor ratios.

## References


Arcand, J.-L. and D. Jaimovich (2010): "Social and Economic Networks in Rural Gambia," Working paper.

Arcand, J.-L. and D. Jaimovich (2012): "Does ethnic diversity decrease economic interactions? Evidence from exchange networks in rural Gambia," .

Bank, A. D. (2005): *Appraisal report: farmer managed rice project.* Agricultural and Rural Development Department, Central and West Regions, African Development Bank Group. Tunis, Tunisia.

Bourlès, R. and Y. Bramoullé (2013): "Altruism in Networks," Working Paper Nr 56.

Chavas, J.-P., R. Petrie and M. Roth (2005): "Farm Household Production Efficiency: Evidence from The Gambia," *American Journal of Agricultural Economics*, 87, 160–179.

Comola, M. and M. Fafchamps (Forthcoming 2014): "Testing Unilateral and Bilateral Link Formation," *Economic Journal*.

Cox, D. and M. Fafchamps (2007): "Extended Family and Kinship Networks: Economic Insights and Evolutionary Directions," *Handbook of Development Economics, Chapter 58*, 4, 3711–3784.

Cragg, J. G. (1971): "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," *Econometrica*, 39(5), pp. 829–844.

Deininger, K., D. A. Ali and T. Alemu (2008): "Assessing the Functioning of Land Rental Markets in Ethiopia," *Economic Development and Cultural Change*, 57(1), 67–100.

Deininger, K., D. A. Ali and T. Alemu (2009): *Land rental markets: Transaction costs and tenure insecuity in rural Ethiopia*chap. 3 in The emergence of land markets in Africa: Impacts on poverty, equity, and efficiency. Routledge; First edition.

Deininger, K. and G. Feder (2001): "Land institutions and land markets," *Handbook of agricultural economics*, 1, 288–331.

Easterly, W. (2001): "Can institutions resolve ethnic conflict?," *Economic Development and Cultural Change*, 49, 687–706.

Fafchamps, M. and F. Gubert (2007): "The formation of risk sharing networks," *Journal of Development Economics*, 83, 326–350.

Fafchamps, M. and M. Söderbom (2014): "Network Proximity and Business Practices in African Manufacturing," *World Bank Economic Review*, 28, 99–129.

Freudenberger, M. S. (2000): "Tenure and Natural Resources in The Gambia: Summary of Research Findings and Policy Options," Land Tenure Center. University of Wisconsin-Madison. Working paper no. 40.

Gajigo, O. and A. Saine (2011): "The effects of government policies on cereal consumption pattern change in the Gambia," *Review of African Political Economy*, 38(130), 517–536.

Gambia Bureau of Statistics (2011): "Integrated Household Survey. Income and Expenditure. Poverty Assessment - 2010," Discussion paper.

Ghebru, H. and S. T. Holden (2009): *Factor market imperfections and rural land rental markets in Northern Ethiopian Highlands*chap. 4 in The emergence of land markets in Africa: Impacts on poverty, equity, and efficiency. Routledge; First edition.

Hayami, Y. and K. Otsuka (1993): *The Economics of Contract Choice: An Agrarian Perspective*. Oxford, UK: Oxford University Press.

Holden, S. and H. Ghebru (2006): "Kinship transaction costs and land rental market participation," Department of Economics and Resource Management. Norwegian University of Life Sciences.

Holden, S. T., O. Keijiro and F. Place (eds.) (2009): *The Emergence of Land Markets in Africa: Impacts on Poverty, Equity, and Efficiency.* Washington, DC: Resources for the Future.

IMF (2007): "The Gambia: Poverty strategy reduction paper," IMF country report no. 07/308. Washington, DC: IMF.

Jackson, M. O. and A. Wolinsky (1996): "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71(0108), 44–74.

Jaimovich, D. (2011): "Macrostructure and microstructure: Evidence from overlapping village networks in The Gambia," *Working Paper*.

——— (2013): "Missing links, missing markets: Internal exchanges, reciprocity and external connections in the economic networks of Gambian villages," MPRA Paper No. 44080.

Jin, S. and T. S. Jayne (2013): "Land Rental Markets in Kenya: Implications for Efficiency, Equity, Household Income, and Poverty," *Land Economics*, 89(2), 246–271.

Kevane, M. (1997): "Land tenure and rental in Western Sudan," *Land use Policy*, 14(4), 295–310.

Laferrère, A. and F.-C. Wolff (2006): "Microeconomic models of family transfers," *Handbook of the economics of giving, altruism and reciprocity*, 2, 889–969.

Macours, K., A. d. Janvry and E. Sadoulet (2010): “Insecurity of property rights and social matching in the tenancy market,” *European Economic Review*, 54(7), 880–899.

Otsuka, K. (2007): “Efficiency and Equity Effects of Land Markets,” *Handbook of Agricultural Economics*, 3, 2671–2703.

Pamela, K. (2010): *Land, labour and entrustment : West African female farmers and the politics of difference*. Leiden, Boston.

Rauch, J. E. (2001): “Business and Social Networks in International Trade,” *Journal of Economic Literature*, 39(4), pp. 1177–1203.

Sadoulet, E., A. de Janvry and S. Fukui (1997): “The meaning of kinship in sharecropping contracts,” *Amrical Journal of Agricultural Economics*, 79(2), 394–407.

Sadoulet, E., R. Murgai and A. D. Janvry (2001): *Access to Land, Rural Poverty, and Public Action*Ch. 8, Oxford University Press.

Skoufias, E. (1995): “Household Resources, Transaction Costs, and Adjustment through Land Tenancy,” *Land Economics*, 71(1), 42–56.

Swindell, K. (1987): “Family Farms and Migrant Labour: The Strange Farmers of the Gambia,” *Canadian Journal of African Studies*, 12(1), 3–17.

Teklu, T. and A. Lemi (2004): “Factors affecting entry and intensity in informal rental land markets in Southern Ethiopian highlands,” *Agricultural Economies*, 30, 117–128.

Topa, G. (2001): “Social Interactions, Local Spillovers and Unemployment,” *The Review of Economic Studies*, 68(2), pp. 261–295.

Udry, C. (1996): “Gender, agricultural production and the theory of the household,” *Journal of Political Economy*, 5, 1010–1046.

Udry, C. and T. Conley (2004): “Social Networks in Ghana,” Working paper.

Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, vol. 1 of *MIT Press Books*. The MIT Press.

# Chapter 3

# Are inter-household land transactions pro-poor?

Ulrik Richardt Beck and Benedikte Alkjærsig Bjerge*

**Abstract**

In many African countries, the poor rely on community-based systems of social security. One such social institution aimed at the reduction of food insecurity is norm-based access rules for vital resources. We use a network-level dataset to investigate patterns of inter-household land transactions in rural Gambia. The dataset allows us to address problems of omitted variable bias and key endogeneity issues. Furthermore, we investigate patterns of land transactions and reciprocating transactions of labor and other inputs. Land transactions tend to be pro-poor: Households in the first income quartile (low-income households) receive more land than other households, and households in the fourth income quartile send more land. Results are consistent with inter-household land transactions being partly driven by social security considerations. This result is driven by less densely populated and less ethnically diverse villages. Even though poor and landless households receive more land, transactions come with an obligation: We find evidence in support of direct reciprocity of land transactions in the labor market. Patterns of reciprocity are not affected by population density or ethnic diversity.

**Keywords:** community social security systems, inter-household transactions, reciprocity

**JEL classifications:** D85, O12, O55

*Ulrik Richardt Beck (ulrik.richardt.beck@econ.ku.dk) and Benedikte Alkjærsig Bjerge (benedikte.bjerge@econ.ku.dk) are both employed at the Department of Economics, University of Copenhagen. The authors thank John Rand, Finn Tarp, Theday Jemaneh Mekasha, and the participants at the Development Economic Research seminar at University of Copenhagen for their helpful comments. We also thank Dany E. Jaimovich and the research team, including principal investigator Jean-Louis Arcand, investigators Yiu Por Chen and Yong He, and finally, field coordinators Cheikh Ibrahima Fall Diop, Eric Djimeu Wouabe, Malek Garbouj and Slavica Zec for making the household survey available to us.

# 1 Introduction

Many households in West Africa depend on rain-fed agricultural farming for survival. For such households, access to land is crucial. Anthropologists and sociologists argue that traditionally, village-based communities in this setting have relied on a norm-based access rule to land, guided by social security considerations (Dey, 1982; Eastman, 1990; Platteau, 1991; Freudenberger, 2000; Platteau, 2002; Pamela, 2010). Under such a rule, membership in the rural community ensures access to sufficient land to provide for the livelihood of the household members.[1] This rule can, if effective and functional, ensure that all households who need it have access to agricultural land. An immediate implication of such a membership-based access rule is that households that own little or no land can obtain usage rights to land, even in the absence of external land rights redistribution. However, this traditional system faces at least two challenges. First, it has been argued that as population density increases and high-quality agricultural land becomes increasingly scarce, land access will be based on increasingly market-oriented terms as the previously important norm-based land-access mechanism is weakened (Platteau, 2002). Second, a high level of ethnic diversity can be thought to affect the level of trust and sense of community solidarity and, in turn, the strength of the norm-based access rule. This paper investigates whether land transfers are motivated by differences in income, consistent with the presence of a norm-based access rule. Furthermore, it investigates whether village-level differences in population density and ethnic heterogeneity can explain differences in the strength of the norm-based access rule.

Within economics, the literature on gift-giving has long recognized reciprocal gestures that take the same tangible form as the gift that triggers them (Cox, 1987). In contrast, anthropologists and sociologists are often concerned with reverse transfers that take a different form compared to the original medium of the favor. These may be tangible goods or more abstract concepts such as loyalty or social status for the donor (for a discussion see Platteau and Sekeris, 2010). We investigate whether land transfers are reciprocated in the form of labor or agricultural inputs. Moreover, experimental economists have gathered evidence in recent years that reciprocity need not be restricted to pairs of individuals (Seinen and Schram, 2006). Instead, so-called indirect

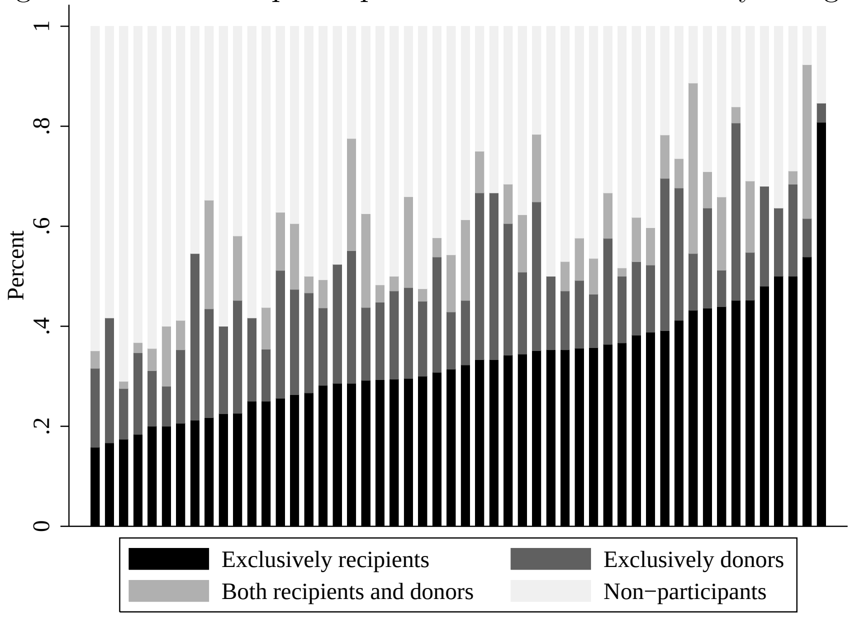
[1] The term "member" includes not only those who can claim descent from the founding lineages but also strangers and migrants who have been accepted as members of the village community (Platteau, 2002).

reciprocity involving a third party are likely to occur. To accommodate the possibility that land transfers may be returned to the donor by a non-recipient third party, we extend the traditional analysis, which only involves direct reciprocity, to also include the various measures of indirect reciprocity suggested in the literature. We therefore empirically investigate whether reciprocating norms exist and whether they depend on the village-level characteristics of population density and ethnic heterogeneity.

Rural Gambia is an illustrative case for studying traditional social security systems. According to previous sociological studies, no important feature of traditional Gambian land tenure is unusual compared to other African countries (Eastman, 1990). According to the World Development Indicators database, 48.4 percent of the population and 74 percent of the rural population lived below the national poverty line in 2010. A significant fraction of the population depends on subsistence farming (Gajigo and Saine, 2011), and while the land system in urban areas has been subject to land reform, an indigenous system of land rights is still in effect in rural areas (Freudenberger, 2000; Chavas, Petrie and Roth, 2005; Pamela, 2010). A World Bank report states, "In the absence of any state-supported welfare programs, social safety nets in The Gambia are based on social and religious traditions. [...] sharing of income and work continues to act as an effective safety net [...] the obverse implication is that anyone with above average earnings is expected to support near relatives and friends with lower income levels" (World Bank, 1993). Traditionally, descendants of the first settlers, that is, those who possess surplus land, have a moral obligation, along with the village chief, to assign land to those in need. Land access and the quality of land that is assigned are, however, dependent on community membership status. Importantly, even though there is no market support for a transaction, it may still occur. Indeed, such allocations of temporary usage rights are often non-monetary in nature. This means that donors of land rarely receive monetary payment for the land they lend to other farmers (Eastman, 1990). However, the lending of land itself may create ties that can be called upon when needed (Cashdan, 1985). Thus, in periods of labor shortage, often before and during the rainy season in relation to weeding and harvesting, donors of land may receive labor or other inputs from land-receiving households.

The survey data reported in Figure 1 confirm that transfers are frequent in rural Gambia. Each bar in the figure represents a village; the top portion represents the fraction of households that do not participate in the land mar-

Figure 1: Household participation in the land market by village




ket, while the bottom two portions represent households that are exclusively recipients and donors, respectively. More than 50 percent of the households surveyed participate actively in the land market either as donors or recipients. A highly unequal land distribution of land ownership rights means that some households transfer land to more than one household. Around 40 percent of households are classified as land recipients.

This paper contributes to the existing literature in three ways. The first contribution is to the empirical literature on inter-household gift transactions (i.e. Coxa, Hansenb and Jimenez, 2004; Kazianga, 2006; Mitrut and Nordblom, 2010). Using a dyad-level network dataset, we are able to explicitly account for several identification issues of previous studies. First, it is possible to correct for potential omitted variable bias by including both recipient and donor income. Second, we check for the possible endogeneity of monetary income using information about households' pre-transfer income. A second contribution is that the paper provides empirical evidence of the existence of a norm-based access rule regarding vital resources in rural communities, but also shows how the importance of the access rule depends on village characteristics.. Finally, the paper contributes to the literature concerned with indirect reciprocity and provides a simple test of indirect reciprocity using observational data.

We find that inter-household transfers of land are motivated by social security considerations. However, this is only the case in the relatively less population-dense villages as well as villages that are more ethnically homogeneous. In fact, inter-household land transfers are pro-poor: (i) The poorest households with no or little monetary income per capita receive more land, and (ii) landless households are allocated more land. This pattern contrasts previous studies on gift giving that suggest that "transfers occur through social networks and poor individuals are excluded from these networks" (Devereux, 2001; Kazianga, 2006). However, this effect is only found in low-density villages, indicating that population pressure may affect the functioning of the norm-based access rules. We further find strong evidence of reciprocating behavior in the labor market. However, we are unable to confirm the hypothesis that behavior is rewarded by a third party through indirect reciprocity (Seinen and Schram, 2006; Kolm, 2006), suggesting that the underlying network structures are less complicated than one might expect.

Even if the premise of mostly non-monetary land transfers is accepted, there may still be multiple motives for land transfers. In fact, using the same dataset, Beck and Bjerge (2014) found that the land transfers increase allocative efficiency by transferring land from households with high land–labor ratios to households with low land–labor ratios. The present paper investigates another motive for land transfers, namely whether land transfers are related to differences in income of households, which would provide evidence for the existence of the norm-based access rule.

While the findings provide insights about the functioning of community-based social security schemes, they also have immediate policy implications. Throughout the post-colonial period, improvement in the asset base of the poor has been viewed as a central strategy to mitigate poverty. In a poor agrarian economy, this entails improving the terms on which the poor have access to land (Besley and Burgess, 2000). Land reforms in West Africa have typically been concerned with increased land tenure security, though the effect has often failed to materialize into increased investments (for a review for West Africa see  Fenske, 2011). A recent paper shows how this can potentially be caused by altruistic effects in a network setting (Bourlès and Bramoullé, 2013). The authors show how a Pigou–Dalton redistribution from rich to poor can end up increasing inequality if the redistribution removes resources from an agent that playsan important supporting role in the local neighborhood. Hence, additional care must to be taken in the design of public policies in cases where network effects of altruism are likely to play a role. The results from this paper are in line with the increasing recognition of the "efficiency and dynamism" of indigenous systems (Stamm, 2004). However, they also provide credibility to the argument that such systems function less well in high-density and more ethnically diverse areas. Therefore, further population growth may reduce the effectiveness of such traditional systems, and reforms may therefore be needed. Our results suggest that such future reforms should be concerned with potential crowding-out effects. To the extent that public land reforms crowd out private voluntary transfers to landless households, the potential effect of land redistribution is likely to be considerably smaller than often anticipated. This result may help explain the lack of effect in the wake of previous land reforms targeting tenure security.

The remainder of this paper is organized as follows. The next section discusses the underlying idea of community social security schemes in more detail. The third section presents the empirical estimation strategy and discusses endogeneity of our income variable. The fourth section presents the data, while section five discusses the estimation results. Section six concludes this paper.

# 2 Traditional land-access norms and reciprocity

## Norm based land access rules

One can distinguish between two types of social institutions aimed at the reduction of food insecurity (Platteau, 1991, 2002). These are informal mutual insurance arrangements and norm-based access rules for vital resources. The existing empirical literature has focused almost exclusively on informal mutual insurance (Morduch, 1995; Fafchamps and Lund, 2003; Dercon and Weerdt, 2006; Mazzocco and Saini, 2012). This paper instead focuses on the mechanics of norm-based access rules to land.[2] While informal mutual insurance arrangements normally kick in after a shock has occurred, and can therefore be regarded as an ex-post insurance mechanism, the norm-based access rule to vital resources is an ex-ante mechanism used to secure livelihoods on an annual basis independent of a shock. Land transfers, due to the substantial delay between planting and harvesting, are not well suited for insurance purposes

[2]The former mechanism typically kicks in following a bad harvest or some other negative shock to welfare. Development economists have extensively tested the informal mutual insurance mechanism focusing on reciprocal state-contingent transfers. The main insight from this literature is that voluntary transfers in response to shocks are enforced by expected future reciprocation: If household $j$ expects to receive future insurance benefits from sharing risk with household $i$, $j$ will give something to $i$ today if $i$ is hit by a shock.

since immediate relief is often necessary following a negative shock. However, in a setting of subsistence agriculture, access to land is vital for the welfare of individual households. The network of land transfers is therefore well suited to investigate the mechanics of norm-based access rules.

Social security motives in rural communities are rooted in a shared belief of how to behave towards other community members. In The Gambia, feelings of duty, cooperation, trust, and obligation towards kin, friends, and strangers form the basis of the local moral economy (Platteau, 2006; Pamela, 2010: 6). Such feelings are central to the concept referred to as *badingya* by the largest ethnic group in The Gambia, the Mandinkas. *Badingya* represents cooperation, obligation, harmony, and productivity (von Braun and Webb, 1989; Freudenberger, 1993). This principle binds relatives and communities together as it shapes notions of social justice. In contrast, the concept of *fadingya* refers to the negative traits of individual selfish ambitions and competition. These twin concepts impose a limit on accumulation of private productive assets, thereby working to ensure social stability by avoiding detachment between individual actions and the best interest of the social group (von Braun and Webb, 1989; Platteau, 2006).

While district-level authorities administer some land for the benefit of district inhabitants, they do not have the legal capacity to interfere with the allocation of land by individual families (Freudenberger, 1993). Hence, access to vital resources is not governed by explicit agreements or external enforcement but instead by social norms, such as those embedded in the concepts of *fadingya* and *badingya*. Even in the presence of strong social norms, there may be other reasons why richer households choose to help poorer households. One candidate is pure self-interest: It may be the case that in the attempt to achieve a higher social status in the village or gain political power, loyal allies are "bought" through a land transfer. Indeed, acting in accordance with social norms can also be seen as an expression of self-interest since households may choose to do so in order to protect themselves from community sanctions. Another possibility is that households are altruistic and sincerely care about the welfare of other community members.

Two principal types of usage rights exist in The Gambia. They are referred to as primary and secondary rights. Primary usage rights are obtained by clearing bush land. Primary rights are similar, although not equivalent, to the Western concept of land ownership: The household can decide which crops to grow and whether to lend some of the land to other farmers. Under the indigenous tenure system, leasing and selling land has traditionally been prohibited (i.e. Freudenberger, 1993). There is evidence that increasing land pressure is leading to a breakdown of the prohibition on renting land (Dey, 1982; Freudenberger, 2000). However, an anthropological study based on key informants in a countrywide sample of 52 villages found that "no respondent reported a case of land being rented for money or a fixed share of the crop" (Eastman, 1990). The descendants of those who first settled and cultivated the land, as well as the village chief, called the *Alkalo*, retain a considerable proportion of the primary usage rights in the village (Dey, 1983; Freudenberger, 2000; Pamela, 2010). This creates a highly unequal distribution of primary land rights between households, where newcomers have very little land or none at all (Dey, 1983). As a result of the inequalities of landholdings and the opposition to the sale of land to non-family or non-residents of the community, allocation of secondary land rights from primary rights holders is common. Secondary tenure secured from primary rights holders is thus an important mode of access to land and other natural resources for non-lineage members. Descendants of the first settlers who possess surplus land have a moral obligation to supply secondary rights to those in need.[3] If the *Alkalo* is among the major landholders in the community, he will also be able to allocate some land of his own to poorer village members.[4] The transfer of secondary rights must often be renewed on an annual basis.[5] It should be noted that, to the extent that land rights can be leased out legally, secondary rights can also be obtained through normal market-based transactions.[6]

[3]In a field study in Dumbutu, it was found that the principle that all residents should have access to land if they have the means to either cultivate or to build compounds is reinforced by the fact that many land-rich compounds do not have enough labor to use all their land. This is partly explained by the considerable urban migration of youth (Freudenberger, 2000).

[4]In Fula-dominated villages, social status is usually manifested in the value attached to the position that the individual holds, and thus the *Alkalo* is often the most powerful village member. Hence, the village *Alkalo* is therefore also the one who allocates land for settlement and cultivation. However, if the land is used for agricultural purposes, the *Alkalo* first needs the consent of the landholder (Freudenberger, 2000).

[5]Previously, land was often lent for several seasons at a time. To prevent people from borrowing land over long periods of time to claim land ownership rights, land owners tend to insist on seasonal loans in order to maintain control (Dey, 1983: 388). It should also be noted that borrowers are not necessarily allocated the same fields every year (Freudenberger, 2000: 82).

[6]In the sample of Gambian rural households studied by Chavas, Petrie and Roth (2005), some 28 percent of all household land was borrowed.

## Reciprocity

Even though norm-based land transfers are traditionally non-monetary transfers (Dey, 1982; Eastman, 1990), recipients may be socially inclined to reciprocate the kindness. If social relations in neighborhoods, families, and workplaces are governed by social norms rather than explicit agreements, reciprocity-based transfers can be regarded as a norm enforcement device (Fehr and Gašchter, 2000; Sacco, Vanin and Zamagni, 2006). The reciprocity norm is a common social expectation that helps sustain balance between donors and recipients: A donor can therefore often expect a reverse transfer of some kind. Failing to repay kind favors brings feelings of guilt and potentially social exclusion, which help to enforce and uphold the norm of reciprocating. Fehr and Schmidt (2006) discuss experimental evidence showing the relevance of reciprocity. Fong, Bowles and Gintis (2006) argue that reciprocity norms are important determinants of support for re-distributive systems of the welfare state, and Platteau (2006) argue that reciprocity norms are relevant enforcement mechanisms underlying risk-sharing arrangements in village societies. With the exception of experiments, studies of reciprocity have in general been limited to investigating direct reciprocity. Direct reciprocity occurs when a gift by a donor is reciprocated to the recipient. In contrast to the concept of altruism, which can be defined as an unconditional kindness or "cooperative" behavior where actors do not expect future material benefits from their actions, reciprocity is an in-kind response that take place even if no material gains are expected (Fehr and Gašchter, 2000; Fehr and Falk, 2002).[7] Experiments suggest that direct reciprocity leads to high levels of continued interaction between fixed partners. This is in accordance with general beliefs that people behave more nicely to people who were nice to them.

An insight from evolutionary biology is that reciprocity need not be restricted to dyads of interacting individuals (Trivers, 1971; Alexander, 1985). Instead, instances where the transfer is returned to the donor by a non-recipient third party are referred to as indirect reciprocity. Two central measures of indirect reciprocity, summarized by Kolm (2006), are generalized and reverse reciprocity.[8] Reverse reciprocity, where $i$, $j$, and $k$ denote households, takes

[7]Therefore, reciprocal behavior in one-shot interactions is often called "strong reciprocity", as opposed to "weak reciprocity," which is motivated by long-term self-interest in repeated interactions.

[8]Kolm (2006) refers to indirect reciprocity as "extended reciprocity." Apart from generalized and reverse reciprocity, he also discusses chain reciprocity ($j$ gives to $i$, $k$ gives to $m$). We do not regard chain reciprocity as important in the setting analyzed here.

place when a transfer from $j$ to $i$ induces a third party $k$ to transfer to $j$. Related to this concept is generalized reciprocity, where the transfer from $j$ to $i$ entails $i$ transferring to a third party $k$. Using extensive computer simulations, Nowak and Sigmund (1998) were the first authors to recognize and test the importance of indirect reciprocity. Since then an increasing number of experimental studies have tested the hypothesis of indirect reciprocity (i.e. Bolton, Katok and Ockenfels, 2005; Seinen and Schram, 2006). For instance, Seinen and Schram use a repeated helping game to test whether people behave more nicely towards people who behave nicely towards others (i.e., generalized reciprocity). They show that indirect reciprocity is an important phenomenon in the laboratory: 48 percent of the decision makers base their strategy at least partly on the social status of the person they are matched with. The concept of reverse reciprocity is less thoroughly tested, possibly because it is regarded as less important compared to the possibility of generalized reciprocity (Kolm, 2006).

In this paper, we focus on patterns of direct and indirect reciprocity to land transfers within villages. We do this by investigating how land transfers are correlated with patterns of transfers in other media, namely labor and other production inputs, such as fertilizer or tools.

In The Gambia, receivers of land usage rights motivated by norm-based access rules have the responsibility to be good neighbors (Freudenberger, 2000), and sometimes land access will be reciprocated by a symbolic payment in kola nuts, cash, or labor services (Freudenberger, 1993; Pamela, 2010). In the words of Freudenberger (1993, p.19), based on a case study of a single district in The Gambia, "Land and other productive resources are in general not movable or exchangeable in traditional societies. This is especially true for land. 'Gifts' or 'tokens' may be given in 'exchange' for the right to cultivate. [...] Today 'kola-money' is given instead. This is not a payment for the land itself, but a token that defines specific user rights. Gifts and tokens are used for "payment" when the remittance does not reflect a market price for the object in question. The term "kola-money" reflects an attempt by lenders and borrowers to maintain the pre-market conception of land." The dataset does not allow for distinguishing between barter transactions and land usage rights donations that are reciprocated through a transfer of labor or input. Thus, in line with Seinen and Schram (2006), we use the term "reciprocity" to denote a certain pattern of conditional behavior and not to describe a type of motivation or a preference.

Even though sale of land is prohibited under the traditional tenure system as previously described, there is some evidence that norm-based access supported by reciprocal behavior and fixed-rent contracts coexist to some degree in Gambia today: Using the same dataset as the present paper, Jaimovich (2013) finds that households that have links with households outside the village are less likely to engage in directly reciprocated transfers within the village. This is consistent with a scenario where outside-village links are more likely to be market-based and monetarily reciprocated, and less likely to be reciprocated by a reverse transfer of goods. The focus of the present paper on within-village land transactions limits this concern to some degree.

## Threats to traditional norm based transfer systems

The traditional tenure systems, norm-based access rules, and norms of reciprocity were all developed under conditions of land abundance. However, it has been recognized, at least since the seminal work of Esther Boserup (1965), that population increase can affect the prevailing land access institutions. A basic prerequisite for the effectiveness of asset-sharing mechanisms as food security devices is that productive resources are sufficiently plentiful compared to the size of the population. As land becomes an increasingly scarce resource and patterns of land use shift from extensive to intensive production, these systems inevitably undergo important transformations (Platteau, 2002, 2006). Moreover, a shorter fallow period as a consequence of population pressure means that less land is available for redistribution to the poorest households (Platteau, 2002). Indeed, there is an understanding that under the impact of population growth and increased commercialization, markets emerge as a means to allocate scarce production resources. These consequences need not be uniquely related to customary tenure systems, but are expected to occur in all situations of competition for land and investment in increasing agricultural productivity (Woodhouse, 2003). In such a setting of increasing scarceness of land and market-based allocation of resources, traditional social norms have less influence over actions than they do in the conventional village societies in which they emerged. This has the potential to reduce the food and income security of many individuals or social groups. The Gambia has experienced high rates of population growth in the last few decades, resulting in a significantly higher population density compared to other regions in Sub-Saharan Africa.[9]

[9]According to the World Bank Indicators, the total population in The Gambia was last recorded at 1.8 million people in 2012, up from 0.4 million in 1960, an increase of 384 percent

It is therefore possible that villages with higher population density are experiencing the deteriorating effect of land scarcity on the strength of norm-based access rules.

Population growth and subsequent market formalization are not the only factors that may pose limitations to traditional practices of hunger insurance. The informal social security scheme is enforced through social norms which are themselves dependent on solidarity and group feelings in the community. The large majority of interactions in people's lives take place based on family relations and ethnicity and between neighbors, and thus feelings of solidarity and trust may be larger among certain groups than among others. According to Zak and Knack (2001), the main argument for this inverse relationship between interpersonal trust and social distance is that when people share the same ethnic background, their social distance is reduced, and thus trust is strengthened. A seminal paper by Alesina, Baqir and Easterly (1999) found that ethnic heterogeneity leads to poorer provision of public goods. While the robustness of this result has been questioned by Gisselquist (2013), it is worth investigating whether high levels of ethnic heterogeneity could also affect the level of social security provided by village institutions through the level of trust between community members. Using the same dataset as the present paper, Jaimovich (2011) reports that high ethnic diversity is related to a higher density of links and more clustering in the land network at the village level. Furthermore, members of minority ethnic groups do not receive less land, and senders and recipients belonging to the same ethnic group is not a predictor of land transfer. While this appears to show that ethnicity is not important for access to land, it is possible that ethnicity does matter for the functioning of the social security mechanism.

# 3 Method

## Testing pro-poorness of land transactions

The empirical strategy is similar to the method used to test for altruism in gift behavior in a number of previous studies (for example Cox, 1987; Cox and Jakubson, 1995; Kazianga, 2006; Mitrut and Nordblom, 2010). This approach

during the last 50 years. The population density has thus increased from 40 people per sq. km in 1960 to 171 in 2011. This compares to an average population density in Sub-Saharan Africa of 38 people per sq. km in 2011. In comparison to the large population growth in The Gambia, hectares of arable land per person has increased slightly in recent years (from 0.23 in 2000 to 0.26 in 2011). However, this does not say anything about land quality.

investigates the effect of recipient income on gift-giving behavior, excluding all donor characteristics. A negative relationship in these studies indicates that lower income induces gift giving, consistent with presence of altruism. In contrast to the literature on altruism, we use land transactions as the dependent variable, which enables the interpretation of the effect as caused by norm-based access rules. The dyad-level data allow us to extend the standard method in two ways. First, in contrast to previous studies (Cox, 1987; Kazianga, 2006; Mitrut and Nordblom, 2010), we control for both donor and recipient characteristics explicitly and thereby overcome problems of omitted variable bias. Second, since all households have many potential partners, it is possible to estimate a model which includes either donor or recipient fixed effects.

The baseline regression takes the following form:

$$A_{ij} = \alpha + \gamma_1 y_i + \gamma_2 y_j + \gamma_3 w_{ij} + \beta_1 z_i + \beta_2 z_j + \sum_k a_k + \epsilon_{ij} \tag{1}$$

$A_{ij}$ is the amount of land household $i$ receives from $j$, measured in hectares. Thus, $i$ denotes the recipient, while $j$ denotes the sender (i.e., the donor). $y_i$ and $y_j$ are $i$ and $j$'s log-transformed monetary income per capita. The income variable measures cash-in-hand earned from off-farm employment and cash crop production. The income variable is included as total income per capita in the baseline specification.[10] Subsequent regressions split income into agricultural and non-agricultural income per capita. The vectors $z_i$ and $z_j$ are household-specific attributes. $w_{ij}$ are link-specific characteristics, and $a_k$ is a village fixed effect for village $k$.

If recipients with a lower income receive more land, then we expect $\gamma_1 < 0$. A test of $\gamma_1 < 0$ is therefore evidence that land transactions contain some element of norm-based access rules in that poorer households are allocated more land. If $\gamma_2 > 0$, land is transferred from households with higher income. Even if the social security effect is present, it need not be linear in log-income. It is possible that the transaction motive changes depending on the level of income per capita of the recipient (Coxa, Hansenb and Jimenez, 2004). For

[10] One possibility is that the self-reported income variable does not reflect true household income due to mismeasurement and unwillingness of respondents to reveal their true income. If errors are random, this leads to attenuation bias, and results can therefore be seen as lower bounds on the true absolute effect size. However, underreporting of income is likely to occur – especially since household information was collected in groups where multiple households were present. This will again attenuate results. There is, however, an alternative option: It is possible that donors also do not know potential recipient households' true income, meaning that the potentially underreported income measure in the dataset may be the relevant decision variable.

instance, poor households were found to be excluded from inter-household gift exchange in Burkina Faso (Kazianga, 2006). It is possible that social security only insures the poorest households. If this is the case, only recipients with the lowest incomes should receive land based on their income. Moreover, if villagers care about the general level of equity in the village, one would expect the richest households to transfer more land. To allow for this, we let the effect of income of the recipient and the donor vary over different quartiles of income distribution.[11] We investigate this using a spline regression approach (Kazianga, 2006).

To the extent that other variables are correlated with the actual amount of land transacted and household income, it is essential to control for these in order to obtain the true underlying motive for inter-household transactions. The baseline specification therefore includes controls for characteristics of the households (including the number of household members and whether the household receives remittances) and of the household heads (including age, gender, and educational level). Summary statistics are reported in Table 1. The link-specific characteristics include a variable for whether household $i$ and $j$ are kin related, either through the household head, the wife(s) of the household head, or marriage ties. Dyad-specific summary statistics are shown in Table A3. The value of land and the amount of arable land vary across villages and are likely to affect which transfers take place as well as the size of the inter-household transfer. Hence, village fixed effects are included in all estimations.

Due to a large number of zero observations, equation 1 is estimated using a tobit model. Qualitatively, results are unchanged when estimating using OLS or a log-normal hurdle model (reported in the Appendix, Table B1). Residuals from dyadic observations involving the same individual $i$ are likely to be correlated. To allow for this sort of interdependence between households inside the same village, all standard errors are clustered at the village level. This approach is conservative in the sense that we do not assume anything about the dependency of dyadic observations inside the villages (Fafchamps and Söderbom, 2014).

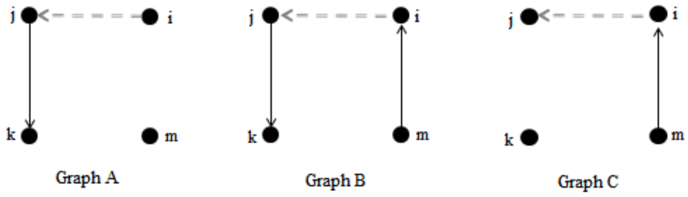
[11] The quartiles are defined using spline regression over the entire sample. One concern is that it is instead the village-specific quartiles that are relevant. Unfortunately, there is no straightforward way of estimating the model using village-level quartiles. Out of 52 villages, for all four quartiles, there are 6 or fewer villages that do not have any households in a given quartile. Therefore, this is not likely to be a substantial problem.

## Testing direct and indirect reciprocity

We focus here on the form of reciprocity behavior and the implications for the interaction between individuals in rural village societies. As previously described, reciprocating transfers are likely to occur in other markets. We consider two alternative markets: the market for labor and that for other agricultural inputs, such as tools and fertilizer. Apart from direct reciprocity, we also investigate the presence of indirect reciprocity. In order to do this, we develop different reciprocity predictors, which are then tested using the labor and input markets, respectively. The reciprocity predictors for the labor market below are equivalent to those for the input market. For the sake of clarity, the following section describes the method using the labor network as an example. The same method is used to investigate reciprocity through the network of other agricultural inputs.

The most basic test of direct reciprocity is to investigate whether a labor transfer from $i$ to $j$ ($L_{ji}$) correlates with a land transfer from $j$ to $i$ ($A_{ij}$). Therefore, a predictor of direct reciprocity from $i$ to $j$ is simply an indicator variable equal to one if there is a land transfer from $j$ to $i$, and zero otherwise. Indirect reciprocity implies that households are more inclined to give to those that have been generous towards others. Such an effect would induce third-party households to send labor to land-sending households. In order to capture the different measures of indirect reciprocity, three predictors, collected in $B_{ij}$, are constructed. The three predictors are illustrated in Figure 3. The illustrations focus on the predicted transfers from household $i$ to household $j$. The solid lines indicate land transfers, whereas the dotted lines indicate reciprocation in the form of labor or input. The first predictor, shown in Graph A, is meant to capture reverse reciprocity. The predictor takes the value one if household $j$ is reciprocated by a third-party household ($i$) not receiving land ($B_{ij} = 1$ if $A_{kj} = 1$, $k \neq i$ and $A_{im} = 0$). The second predictor, shown in graph B, is meant to capture conditional reverse reciprocity. It is similar to the first predictor, but in addition requires that household $i$ receives land, though not from household $j$ ($B_{ij} = 1$ if $A_{kj} = 1$, $k \neq i$ and $A_{im} = 1$, $m \neq j$). Finally, we create a predictor to capture any generalized reciprocity effects (Graph C). This third predictor takes the value one if land-receiving household $i$ sends labor or input to household $j$, though $j$ neither receives nor sends any land ($B_{ij} = 1$ if $A_{ik} = 1$, $k \neq j$, $A_{jk} = 0$ and $A_{kj} = 0$). Table A3 includes summary statistics on the dyad-level reciprocity variables.

Figure 2: Types of indirect reciprocity




Note: Solid lines denote land transfers, and dotted lines denote either labor or input transfers.

By regressing these predictors on the actual labor transaction network, it is possible to investigate the importance of direct and indirect reciprocity patterns. The regression takes the following form:

$$L_{ji} = \eta + \omega A_{ij} + \psi B_{ij} + \lambda w_{ij} + \delta_1 z_i + \delta_2 z_j + e_{ij} \tag{2}$$

where $L_{ji}$ is the amount of labor household $i$ sends to $j$, $A_{ij}$ is the predictor for direct reciprocity, and $B_{ij}$ contains the indirect reciprocity predictors outlined above. The control variables included in $z$ are the same as the ones included in the estimation of equation 1. A test of $\omega > 0$ implies direct reciprocity, while a test of $\psi > 0$ suggests the presence of indirect reciprocity. Equation 2 is estimated using a probit model including village fixed effects, and standard errors are clustered at the village level. Since we do not know the sequence of the transfers, estimates of direct and indirect reciprocity should be interpreted as correlations.

A potential concern is that directly reciprocating labor transfers need not be driven by community norms. It is also possible that the transaction is market based: Land-sending households tend to own more land than land-receiving households both before and after land transfers occur. It would therefore make sense that households with surplus land rent in labor. In order to ensure that this is not driving results, we include land endowments of both donors and recipients.

## Effects of population density and ethnic diversity

To test whether village characteristics influence inter-household transaction motives, we introduce interaction terms between village-specific characteristics $v_k$ and variables of interest. Two village-specific measures are considered. First, to investigate the impact of population density, a dummy variable for village density was constructed. This variable is equal to one if the village has a population density measured as inhabitants per square kilometer above the median village, and zero otherwise. Second, to investigate the impact of ethnic heterogeneity, a dummy for ethnic diversity was constructed. This measure is equal to one if the level of ethnic heterogeneity, as measured by a Herfindahl fractionalization index, in the village is above the village median, and zero otherwise.

In order to investigate how access rules for land differ depending on village-specific characteristics, demeaned donor and recipient incomes are interacted with the village-specific characteristics. The model can be written as

$$A_{ij} = \alpha + \gamma_1 \bar{y}_i + \gamma_2 \bar{y}_j + \gamma_3 w_{ij} + \bar{\gamma}_4 y_i v_k + \gamma_5 \bar{y}_j v_k + \beta_1 z_i + \beta_2 z_j + \sum_k a_k + \epsilon_{ij} \quad (3)$$

where the notation is the same as in Section 3 and $\bar{y}_i$ and $\bar{y}_j$ denote the demeaned incomes of households $i$ and $j$. The uninteracted village characteristic is not included in the estimation as it is swept away by the village fixed effects. A joint test of $\gamma_1 + \gamma_4 < 0$ is a test of whether the norm-based access rule guides transfer motives in high-density (ethnically diverse) villages, while a test of $\gamma_1 < 0$ is a test of whether the norm-based access rule guides transfer motives in low-density (ethnically homogeneous) villages.

In order to investigate how patterns of direct reciprocity differ depending on village-specific characteristics, a demeaned version of the labor and input network predictor is interacted with village-specific characteristics. The model can be written as:

$$L_{ji} = \eta + \omega_1 \bar{A}_{ij} + \omega_2 \bar{A}_{ij} v_k + \lambda w_{ij} + \delta_1 z_i + \delta_2 z_j + e_{ij} \quad (4)$$

where the notation is the same as in Section 3 and $\bar{A}_{ij}$ denotes the demeaned predictor of a link from $j$ to $i$. A joint test of $\omega_1 + \omega_2 > 0$ is a test of whether labor or agricultural input transactions follow a pattern of reciprocity in high-density (ethnically diverse) villages, while a test of $\omega_1 < 0$ is a test of whether this is the case in low-density (ethnically homogeneous) villages.

## Endogeneity issues and omitted variables

In this section, we address two potential estimation issues related to endogeneity of the income variable in the land transfer model. Next, we turn to the possibility of omitted variable bias due to unobserved household heterogeneity in both models. Finally, we ensure that specific definitions of village-level characteristics are not driving results.

First, if households take account of the norm-based land access rule by expecting land transfers if income is sufficiently low, it is possible that households lower their monetary income endogenously. This could, for example, be done by consuming more of the household's own production instead of selling it, or through migration decisions. Furthermore, if the land received is used to produce marketable output, a land transfer can in itself affect income.

These are issues of reverse causality. Several approaches have been suggested in the literature on gift giving and altruism, which face similar issues. Kazianga (2006) estimate income using long-run rainfall data under the assumption of dependence between farm outcomes and rainfall. Mitrut and Nordblom (2010) do not explicitly correct for endogeneity of income in their study of Romanian gift transfers, but instead investigate the quality of the income data using household consumption. This paper uses a different approach and calculates a measure of the *pre-transfer income* for all households. This is possible due to the completeness of the transfer network data for all households in the sample.[12]

Using household-level data (2,028 households), linear predictions of realized income per capita as a function of land available to the household including transfers is estimated:

$$y_i = \alpha + \lambda_1 ln(A_i) + \beta z_i + \epsilon_i \tag{5}$$

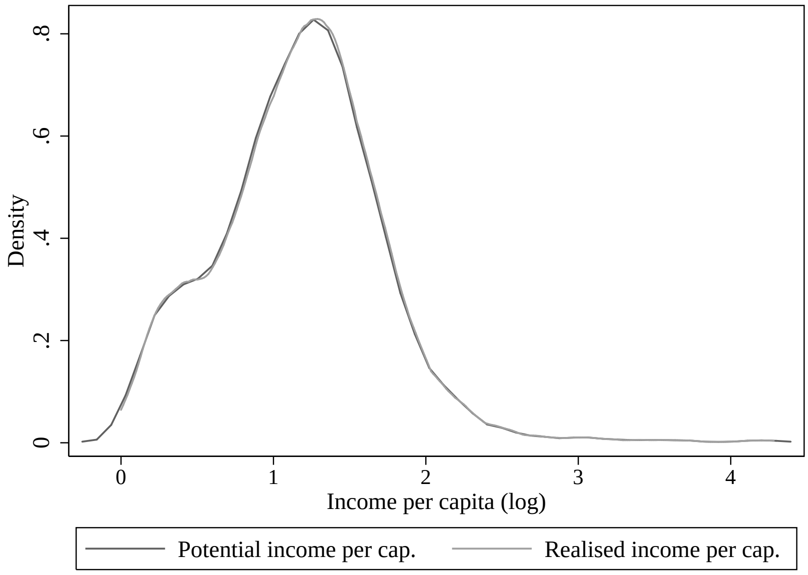
where $y_i$ is the log of income per capita for household $i$, and $ln(A_i)$ is the log of land available to household $i$. Also included is a set of control variables collected in vector $z_i$.

Denote the predicted realized log of income $\hat{y}_i$. Using the estimated parameters and the amount of land available to the household before transfers take place, $A_i^{exante}$, it is possible to compute the counterfactual income level *in the absence of any transfers*: $\hat{y}_i^{exante} = \hat{\alpha} + \hat{\lambda}_1 ln(A_i^{exante}) + \hat{\beta} z_i$. Therefore, $\Delta = (e^{\hat{y}_i} - e^{\hat{y}_i^{exante}})$ is a measure of how much income changes due to inter-household transfers. This measure can now be used to obtain an estimate of

[12] It is widely recognized that self-reported measures may contain measurement errors. Finding an association between land transactions and income in the presence of attenuation bias caused by measurement error in our dependent variable gives additional weight to the assumption of a true relationship.

potential income in the absence of land transfers: $\hat{y}_i^{potential} = log(e^{y_i} - \Delta)$.

Figure 3: Realized and potential income




As seen from Table A1, the available amount of land affects income positively and significantly, while household size affects income per household member negatively and significantly. However, the economic impact is limited: Figure 3 shows that there is little difference in the distribution of potential and realized income per capita for households participating in the land market.[13] This implies that realized monetary income is not highly endogenous to land transfers. Subsequent results therefore use actual income per capita instead of the constructed "potential" income. Results are robust to using the potential income variable.[14]

Second, even when including control variables, there may still be unobserved household-level characteristics affecting behavior. In the case of land transfers, a salient issue is that some households may be better farmers. Since unobserved household characteristics such as farming ability are likely to be

[13]The distribution is shown for the 1,140 households that actively participate in the land market and therefore experience a change in the amount of land cultivated due to land transactions.

[14]Two effects may occur from land transfers: First, if access to more land means that farmers substitute away from off-farm labor, the transfer may cause a decrease in non-agricultural income. Second, if access to land increases cash crop production, then agricultural income is likely to increase. By dividing total income into non-agricultural and agricultural income, we see that land does not affect non-agricultural income significantly, while land affects agricultural income positively and significantly. However, there is no difference in the distribution of potential and realized income for either non-agricultural or agricultural income.

both positively correlated with income and the size of land transactions, omitting farmers' ability is likely to bias our estimates upwards. This would result in a less negative coefficient estimate. Finding an effect in support of the norm-based access rule ($\gamma_1 < 0$) therefore limits the concern related to unobserved heterogeneity.

Nevertheless, we test the robustness of the result to unobserved heterogeneity by estimating equation 1 including either donor or recipient household fixed effects. Inclusion of household fixed effects wipes out all donor or recipient fixed effects, depending on the included set of fixed effects. This also means that when including donor fixed effects, only the effect of recipient income can be identified and vice versa. The model to be estimated can be written concisely as:

$$A_{ij} = \alpha + \lambda_1 y_s + \lambda_2 w_{ij} + \beta z_s + a_t + \epsilon_{ij} \tag{6}$$

for $s, t \in \{i, j\}$ and $s \neq t$, where the included variables are defined as outlined above. To illustrate, for recipient household $i$ and donor household $j$, equation 6 includes household-specific attributes $z_i$, the income of recipient $y_i$, and a donor household fixed effect $a_j$. In this example, a negative and statistically significant estimate of $\lambda_1$ is consistent with the norm-based access rule correcting for unobserved donor heterogeneity. This test is equivalent to the test normally performed in the principle-agent literature.

For the model of reciprocity, two-way donor and recipient fixed effects are included as a robustness check.

A final robustness check is carried out in order to ensure that the specific choice of the median is not driving results regarding the impact of population density and ethnic diversity. This is done by re-estimating the two models given by equation 3 and 4 using alternative quantiles to define highly population-dense and ethnically diverse villages.

# 4 Data and descriptive statistics

The data comes from a baseline survey conducted between February and May 2009 for the purpose of evaluating the national Community Driven Development Project (CDDP). The survey covers 60 randomly selected villages, representative of six out of eight Local Government Areas across different agro-ecological zones, and with populations between 200 and 1,000. The dataset contains three categories of information: (1) village level information, (2) a standard household survey, and (3) information on six networks: land transactions, labor transactions attached to the household head, transactions of agricultural input, credit, marriage, and detailed kinship information. The dataset contains information on transactions and their size between all households residing in the village for these six networks, as well as information on household endowments and characteristics.[15]

Five villages were dropped due to substantial amounts of missing household-level information in these villages. Second, as we are concerned with land transactions in rural areas, three semi-urban villages were dropped.[16] Table 6 includes village-level (52 observations) descriptive statistics. The average village has 521 inhabitants. This corresponds to a village population density of 234.62 inhabitants per square kilometer. The area denominator used in this calculation is the sum of the cultivated area of the village and the village itself. Therefore, the actual population density in the country will be lower than this number. The economic conditions resemble that of other rural communities in West Africa: Very few households have access to electricity, and the majority of households do not have access to improved water sources. The sample also represents the large ethnic diversity in The Gambia: The largest ethnic group (Mandinka) comprises only 55 percent of households in the sample and four other ethnic groups each comprise more than 5 percent of households in the sample. [17] There is a wide range of diversity across villages. Examining ethnic diversity using the Herfindahl fragmentation index, ethnic fragmentation inside villages ranges from 0 (completely homogeneous) to 0.84, with a mean of 0.28. The distribution of self-reported income measured in terms of the Gini coefficient is on average 0.31, with a maximum of 0.60 (1 denotes maximum inequality).

Rural villages in The Gambia are organized in compounds (*kundas*), which correspond to a group of people who work jointly on common fields, eat together, and organize daily activities (von Braun and Webb, 1989; Pamela, 2010). Depending on the size of the compound, independent cooking and consumption units (*dababas*) can co-exist within the compound. The household

[15]The data was collected using a structured group approach with a median household coverage rate in the villages of 94 percent. For detailed information on the sampling methodology and data description, see Jaimovich (2011).

[16]In terms of network activity, some 2 percent of the households in the semi-urban villages participate in the land market, while 10 percent participate in the labor market. The main reason for the absence of land sharing could possibly be the very small landholdings in these areas (0.243 hectares per household compared to 10.282 in the rural villages), as well as increased options for employment outside the village.

[17]Compared to the 2003 Census for The Gambia, the ethnic group Mandinka are slightly overrepresented and Serahules are underrepresented (Arcand and Jaimovich, 2012).

(i.e. an individual *dababa*) is used as the unit of analysis. If several households exist within one compound, the network between these will be present in the dataset. Around 16 percent of household heads in the sample are not the head of the compound in which they live.

Household-level descriptive statistics for all households, and separately for donors and recipients of land, are reported in Table 1. The data is consistent with the description of rural Gambia given in the last section: The largest households in the sample have more than 50 members (only 0.5 percent of the sample). These large households are partly explained by the polygamous nature of the rural Gambian society (49 percent of household heads have more than one wife). Households on average have five adult working members and the household head on average devotes four days annually to working on other farms. The households in our sample are predominately led by poorly educated men: Only 13 percent have any formal education. For the majority of households the main economic activity is related to agriculture (79 percent), through many households also engage in other income-generating activities. The average monetary income per capita is 2,931 Gambian Dalasis a year, PPP-equivalent to 301 USD a year, of which almost 15 percent stems from agricultural activities.[18] Interestingly, recipients and donors do not have significantly different income on average, though a larger share of the income of donors is derived from production of cash crops. Recipient households tend to be smaller in size and more likely to be female headed. While land recipients and non-receiving households have similar amounts of land with primary rights (around 8 hectares), land donors have substantially larger primary rights landholdings (around 17 hectares).

## Inter-household transfers

This subsection provides information on the network data. Information was collected on all transactions taking place within the last year for land and labor and agricultural input. We note two limitations of the data. First, the land network data do not include information about the contact related to the transaction of land. Qualitative field observations support the claim that most transactions are non-monetary and therefore do not come with a cost, while sometimes land transactions are reciprocated through a small symbolic payment in kola nuts, labor services, or cash (Eastman, 1990; Freudenberger,

[18]Using Penn World Tables PPP-adjusted exchange rate. Note that consumption and bartering of own production is not included in this figure.

Table 1: Descriptive statistics: Household level

| | All households | | By participation type (Mean) | | | T-test | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Std. Dev. | (a) Recipients | (b) Donors | (c) Non-participants | (a)-(b) | (a)-(c) | (b)-(c) |
| | N=2,029 | | N=818 | N=516 | N=860 | N=1,004 | N=1,678 | N=1,376 |
| Total Income per capita (1,000 GMD) | 2.931 | 3.330 | 2.751 | 2.915 | 2.996 | | * | |
| Agricultural income per capita (1,000 GMD) | 0.257 | 0.818 | 0.292 | 0.269 | 0.224 | | | |
| Other income per capita (1,000 GMD) | 2.674 | 3.268 | 2.460 | 2.646 | 2.772 | | * | |
| Agricultural share (share of income) | 0.146 | 0.255 | 0.159 | 0.173 | 0.127 | | * | * |
| Receive remittances | 0.458 | 0.498 | 0.455 | 0.529 | 0.437 | * | | * |
| Land owned with official rights (hec.) | 10.034 | 22.907 | 8.348 | 17.118 | 8.285 | * | | * |
| Household size | 13.358 | 8.952 | 14.49 | 15.434 | 11.750 | | * | * |
| Number of working adults | 4.893 | 3.894 | 5.279 | 5.535 | 4.390 | | * | * |
| Days the HH head worked on other farms | 3.631 | 11.018 | 3.592 | 4.083 | 3.295 | * | | * |
| Emigrated household member | 0.526 | 0.499 | 0.517 | 0.599 | 0.503 | * | * | * |
| Age of head | 52.679 | 16.13 | 53.243 | 55.122 | 51.274 | | | |
| Agricultural work | 0.785 | 0.411 | 0.793 | 0.797 | 0.766 | | | * |
| Non-agricultural | 0.178 | 0.383 | 0.167 | 0.157 | 0.203 | | | |
| HH has family links in the village | 0.946 | 0.226 | 0.932 | 0.977 | 0.944 | * | | * |
| HH head has family links in the village | 0.845 | 0.362 | 0.830 | 0.905 | 0.831 | * | | * |
| Wife of HH head has family links in the village | 0.519 | 0.500 | 0.515 | 0.589 | 0.506 | * | | * |
| HH have marriage links in the village | 0.46 | 0.499 | 0.433 | 0.556 | 0.449 | | | |
| Female headed household | 0.047 | 0.211 | 0.040 | 0.035 | 0.057 | | | |
| Illiterate | 0.444 | 0.497 | 0.458 | 0.409 | 0.442 | * | * | |
| Formal education | 0.125 | 0.331 | 0.104 | 0.132 | 0.136 | | | * |
| Monogamous | 0.458 | 0.498 | 0.438 | 0.428 | 0.483 | | * | * |
| Polygamous | 0.489 | 0.500 | 0.521 | 0.527 | 0.452 | | * | |
| Ethnicity: Mandinka | 0.548 | 0.498 | 0.553 | 0.541 | 0.564 | | | |
| Ethnicity: Fula | 0.171 | 0.377 | 0.181 | 0.140 | 0.177 | * | | * |
| Ethnicity: Wollof | 0.097 | 0.296 | 0.084 | 0.130 | 0.088 | * | | |

Note: 165 households participate on both sides of the market. Hence, the t-tests are performed for households either categorised as recipients, donors or non-participants. * denotes statistical differences at least at the 5 percent level.

2000; Jaimovich, 2011).[19] Moreover, findings in the anthropological literature suggest that outright sharecropping is not widespread in Gambia. Dey (1982) outlines how early sharecropping arrangements imposed by donor-implemented programs were unsuccessful and subsequently abandoned. If sharecropping is widespread in the present dataset, we would expect to find large labor transactions in the labor network. However, conditional on participation in the labor market, farmers work only 9.3 days on other farmers' land on average.

Second, while the dataset does not contain information on households located outside the village, it does include information about transactions between households in the village and households outside the village. It is therefore necessary to assume that the village is the natural domain for potential transactions. This is the case if land transactions are motivated by norm-based access rules where the village community is the unit of social security. This assumption is supported by the fact that external actors do not play a major role in the land network: Around 6 percent of the households in the sample either receive or send land to non-village members (no households both receive and send land in the external village market). The relatively low level of land transactions involving households residing outside the village is likely to be explained by the immobility of land. For the labor market, no households work on non-village members' farms, whereas 8 percent of the households receive external labor. The result on the sender side of the labor network is hardly surprising as the labor network is restricted to labor provided by the household head. It should be noted that the labor network does not include seasonal migrant workers known locally as "strange farmers" (Swindell, 1987). Finally, around 3 and 6 percent of the households in our sample either send or receive agricultural input from non-village members, respectively.

Table 2 shows descriptive transaction statistics for the land market. In each village, an average of one fifth of households sends land. The average donor sends land to more than 1 household, to a maximum number of 14 recipient households. The average amount of land transacted per land transaction is almost 2 hectares. In total, donors send an average of 33 percent of their initial land holdings (disregarding households that participate on both sides of the market). This implies that land is transacted from land-abundant households towards land-poorer households. This observation is confirmed in Figure 4. The figure shows changes in land distribution before and after transfers have taken

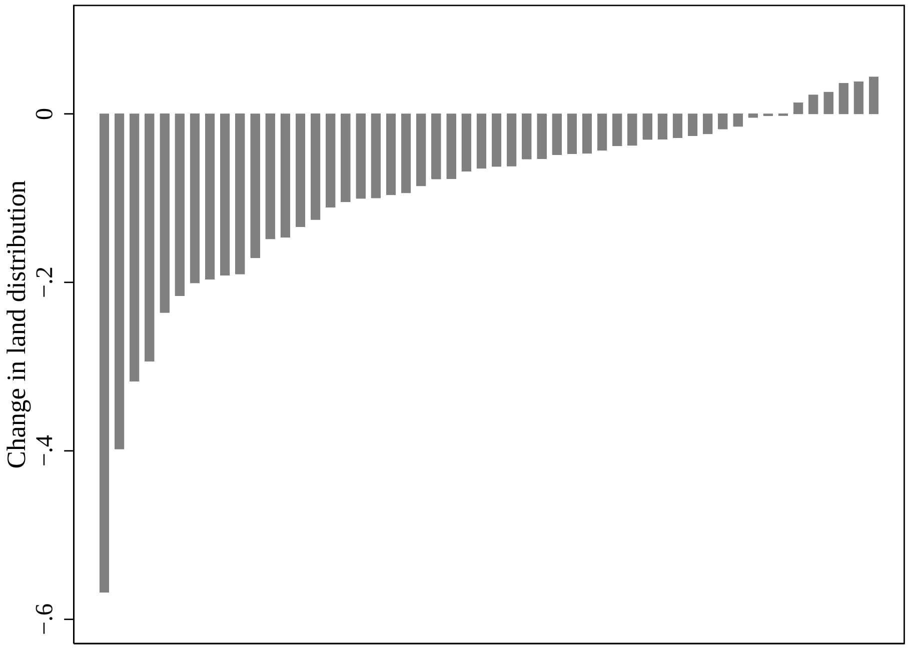
[19]Generally, the tenure system as described in the previous section is not comparable with a patron–client system as often observed South-East Asia. Rather, no landlord class exists in The Gambia (Pamela, 2010).

Table 2: Descriptive statistics: transfers

| | Mean | Std. | Min | Max | Obs. |
|---|---|---|---|---|---|
| Land market | | | | | |
| No. of households (across villages) | 38.02 | 14.45 | 11 | 70 | 52 |
| No. of donors (across villages) | 9.29 | 4.48 | 0 | 24 | 52 |
| No. of donors (% population, across villages) | 24.93% | 9.71% | 0 | 50.0% | 52 |
| No. of recipients per donor | 1.78 | 1.40 | 1 | 14 | 483 |
| No. of recipients (% population, across villages) | 44.42% | 22.22% | 0 | 118.6% | 52 |
| Size of the land transactions (hec.) | 1.83 | 1.50 | 0.25 | 20 | 859 |
| Size of the land transactions (% of donors' initial landholding)* | 32.72% | 26.14% | 0.08% | 100% | 466 |
| Labor market | | | | | |
| No. of labor-transacting hhs (across villages) | 15.21 | 8.43 | 2 | 37 | 52 |
| No. of labor-transacting hhs (% population, across villages) | 40.70% | 19.93% | 8.1% | 86.67% | 52 |
| No. of directly reciprocated land transactions | 10.48% | 30.64% | 0 | 100% | 859 |
| Size of labor transaction (days, conditional on direct reciprocity) | 9.07 | 14.60 | 1 | 90 | 90 |
| Agricultural input market | | | | | |
| No. of input-transacting hhs (across villages) | 17.31 | 9.25 | 4 | 43 | 52 |
| No. of input-transacting hhs (% population, across villages) | 47.08% | 19.75% | 12.12% | 97.67% | 52 |
| No. of directly reciprocated land transactions | 9.78% | 29.72% | 0 | 100% | 859 |

Note: Total number of land transactions is 859. Total number of recipients and donors is 657 and 483, respectively. * Excludes observations where the donor participates on both sides of the market.

Figure 4: Change in land equity by village




Note: Each bar denotes the change in the village-level Theil index of land usage rights due to land transfers. Negative values represent a fall in inequality

place. The land inequality figures are calculated using the Theil index. Each bar corresponds to a village, and negative values indicate a decrease in land inequality ex-post transactions, while positive values suggest that transactions increase land inequality. In the majority of villages, we observe a higher level of land equality in the wake of land transactions (i.e., negative values). A similar picture, though not reported, appears if we compare the distribution of land–labor ratios before and after transactions have taken place.

Table 2 also reports summary statistics for transactions of labor and agricultural input. Unfortunately, the type of agricultural input is unknown, and therefore the ways in which transacted inputs affect agricultural production are also unknown. As seen in Table 2, 41 and 47 percent of households across the 52 villages engage in labor or input transactions, respectively. Out of these, around 10 percent of the transactions are reciprocated by a directionally reverse transfer. Considering only households that reciprocate using the labor market, a recipient household on average reciprocates nine days of wage labor directly to the donor.[20] This corresponds to 0.79 working days per hectare of land received. The amount of reciprocated labor per hectare of land is not significantly

[20]A working day is defined as six hours of work.

different across population-dense or ethnically diverse villages at a 5 percent level.

# 5 Results

## Baseline results

Table 5 shows the baseline results of the effect of recipient and donor monetary income per capita on the land network.

The incomes of both the potential land donor and the potential land recipient are important for whether land is transacted (column (1)). As predicted by the norm-based access rule, recipient income negatively affects land transactions. The average partial effects on land transactions are reported in column (2). They show that on average, an increase in recipient income per capita by 10 percent decreases the amount of land by 0.03 hectares, or 300 square meters of land. The effect is significant at the 10 percent level. Donor income positively affects the probability of being a land sender which is also consistent with the presence of a norm-based access rule. This shows that recipient characteristics, which have been the focus of most of the inter-household gift giving literature due to data limitations, are not the only things that matter for land exchanges: Richer households are more likely to be senders of land. When splitting the income variable of the donor and recipient into agricultural and non-agricultural income (column (3)), it becomes clear that this effect is driven by variation in the non-agricultural recipient income, which is significant and negative at the 5 percent level. Recipient agricultural income, on the other hand, is positive and significant. This may caused by the fact that agricultural income picks up market-based transactions of cash-crop farmers: Households acquiring income from cash crops conduct more commercialized agriculture, and it makes sense that their access channels to land are more commercial as well. Donor non-agricultural income is still positive and significant.

Of particular interest is whether there is a stronger effect for the poorest households. The estimation results using a spline-based approach are shown as the third model. Transfer response to income is shown for the lowest to the highest income quartiles for both donor and recipient.[21] On the recipient side, only the lowest income quartile group is negative and statistically significant (only at the 10 percent level). On the donor side, only the highest quartile

[21] A substantial share of households have zero agricultural income. Therefore, only the third and fourth income quartiles can be constructed and included for agricultural income per capita in Table 5.

Table 3: Pro-poorness of land transactions

| | (1) Coefficients | (2) APE's | (3) Coefficients | (4) Coefficients |
|---|---|---|---|---|
| $i$'s total income | -0.351* | -0.003* | | |
| | (0.198) | (0.002) | | |
| $j$'s total income | 0.441* | 0.004* | | |
| | (0.257) | (0.003) | | |
| $i$'s non-agricultural income | | | -0.439** | |
| | | | (0.196) | |
| - *1st quartile* | | | | -0.748* |
| | | | | (0.387) |
| - *2nd quartile* | | | | 0.233 |
| | | | | (0.860) |
| - *3rd quartile* | | | | -0.772 |
| | | | | (1.021) |
| - *4th quartile* | | | | -0.442 |
| | | | | (0.488) |
| $j$'s non-agricultural income | | | 0.478* | |
| | | | (0.264) | |
| - *1st quartile* | | | | 0.662 |
| | | | | (0.851) |
| - *2nd quartile* | | | | 0.067 |
| | | | | (1.137) |
| - *3rd quartile* | | | | -0.675 |
| | | | | (1.536) |
| - *4th quartile* | | | | 1.024** |
| | | | | (0.455) |
| $i$'s agricultural income | | | 0.683** | |
| | | | (0.278) | |
| - *3rd quartile* | | | | 1.539 |
| | | | | (0.976) |
| - *4th quartile* | | | | 0.378 |
| | | | | (0.392) |
| $j$'s agricultural income | | | 0.267 | |
| | | | (0.338) | |
| - *3rd quartile* | | | | 0.052 |
| | | | | (1.114) |
| - *4th quartile* | | | | 0.280 |
| | | | | (0.491) |
| Observations | 87,788 | 87,788 | 87,788 | 87,788 |

Note: Tobit. Dependent variable: Amount of land $i$ receives from $j$. All regressions include control variables and village fixed effects. Marginal effects reported in column (2). Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

is positive and statistically significant. These results are consistent with the social security motive: Land is redistributed from the richest to the poorest households. In fact, the poorer the household, the larger the land transfer from the highest income quartile of donors. There are no significant effects for any quartile of agricultural income. The missing effect in the spline regression might be due to an increase in standard errors, rather than the lack of an effect.

An alternative explanation of a negative estimate on recipient income is that donors directly care about the well-being of the recipient. Such purely altruistic behavior has been highlighted in the inter-household gift literature (Cox et al., 1998; Kazianga, 2006; Mitrut and Nordblom, 2010). Another possible explanation of motives for land transactions is efficiency gains. If the household has too much land to cultivate using its own labor, it transacts land to other, landless households in order to increase allocative efficiency (Freudenberger, 2000). Using the same dataset, the effect of land transactions on allocative efficiency was explored by Beck and Bjerge (2014). They find that land transactions are efficiency enhancing by making land–labor ratios more equal across households. If households with a low land–labor ratio are also poorer, this could explain the present finding. This, however, is not a strong finding in the data.

### Controls

Control variable coefficients, found in Table A4 in the Appendix, are quite consistent across specifications. In line with Jaimovich (2011), kin-related households are more likely to transact land. Recipients with formal schooling are less likely to receive land. It is possible that educated household heads have more alternative income options and are therefore less likely to receive land. The share of households in the village with the same ethnicity as the recipient is negatively correlated with receiving land. This suggests that village minorities are allocated more secondary usage rights, perhaps because they are less likely to possess primary land rights. Moreover, donor households are typically older, larger in terms of the number of working adults, and better educated. This again shows that donor as well as recipient characteristics matter for inter-household land transactions.

Finally, a specific control variable of interest is whether the recipient household is landless prior to the land transaction. As seen in Table A4, landless households are more likely to receive land and less likely to donate land. This result carries important policy implications, as redistribution of land to landless households is likely to result in the crowding out of voluntary land transactions. To investigate whether this result is applicable to households with little land, we include the initial landholding of both the donor and recipient. The estimation results are shown in Table A5. The coefficient estimate on donors' initial landholding is positive and statistically significant, suggesting that households with more primary usage rights transact more land. However, the coefficient on the recipients' amount of initial land is insignificant. From this, what seems to be important is whether the household has any land rights at all. In summary, these results suggest that future land redistribution reforms in West Africa should take into account potential crowding-out effects of the social security mechanism: Land rights reform may not have as big an impact as one may expect if welfare effects are offset by less norm-based land redistribution.

## Reciprocity

The estimation results of equation 2 including both direct and indirect reciprocity indicators for the labor network are shown in Table 5. The dependent variable is equal to one if $i$ sends labor to $j$. There is strong evidence, like in Table 5, that land transactions are associated with labor transactions in the opposite direction. In terms of indirect reciprocity, the point estimates are all positive as expected but not statistically significant for either reverse or generalized reciprocity. Interestingly, the land endowment of donor households is positive and significant. If transactions were market based, we would expect this to be negative, given the discussion in Section 3. The land endowment of recipient ($j$) households is not significant at the standard 5 percent confidence level.

Turning to the agricultural input network, the estimation results of equation 2 including both direct and indirect reciprocity indicators are shown in Table 5. Like in the case of labor reciprocity, there is a high correlation between land transactions and reciprocating agricultural input transactions: Land market transactions are highly predictive of reciprocating input transactions. Also like in the case of the labor market, there are no consistent effects from the predictors of generalized reciprocity. Overall, the evidence for indirect reciprocity is weaker than one might expect, given the emphasis in much of the experimental, sociological, and human biology literature on the complexity of human nature. In particular, it is not possible to confirm the general findings in experimental studies that indirect reciprocity is an important phenomenon

Table 4: Reciprocity results – labor and inputs

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Direct reciprocity | 0.835*** | 0.826*** | 0.837*** | 0.828*** | 0.847*** | 0.857*** |
| | (0.091) | (0.093) | (0.095) | (0.092) | (0.096) | (0.095) |
| Reverse reciprocity | | | 0.063 | | | 0.030 |
| | | | (0.051) | | | (0.062) |
| Conditional reverse reciprocity | | | | 0.013 | | 0.035 |
| | | | | (0.047) | | (0.047) |
| Generalised reciprocity | | | | | 0.057 | 0.054 |
| | | | | | (0.037) | (0.044) |
| Land recipient $i$ ($\times 100$) | | 0.127** | 0.129** | 0.127** | 0.127** | 0.128** |
| | | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Land donor $j$ ($\times 100$) | | 0.128* | 0.123* | 0.128* | 0.115 | 0.115 |
| | | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Observations | 87,788 | 87,788 | 87,788 | 87,788 | 87,788 | 87,788 |

Note: Probit. Dependent variable: Dummy equal to one if $i$ sends labor to $j$. All regressions include control variables and village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

outside the laboratory (for a review, see Fehr and Schmidt, 2006).

Table 6 and Table A7 report the partial effects for direct reciprocity in the labor and agricultural input markets, respectively. Household $i$ is on average 4.8 percent more likely to directly reciprocate using the labor market if $i$ receives land from $j$. This is a relatively large effect given that only 10.5 percent of land recipient households directly reciprocate labor (reported in Table 2). We also find a relatively large effect looking at direct reciprocity in the agricultural input market: Land recipient households are on average 3.5 percent more likely than non-recipients to send labor to the donor. This compares to the previously reported summary statistics indicating that only 10 percent of the land recipients directly reciprocate using input.

## The strength of the social security norm

In order to investigate whether some villages exhibit stronger signs of the norm-based access rule equation, 3 and 4 are estimated. From the description in Section 2, we expect land transactions in villages with high density and villages with high ethnic diversity to be less controlled by norm-based access

Table 5: Reciprocity results – labor and inputs

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Direct reciprocity | 0.582*** | 0.592*** | 0.600*** | 0.595*** | 0.583*** | 0.589*** |
| | (0.067) | (0.067) | (0.068) | (0.066) | (0.068) | (0.065) |
| Reverse reciprocity | | | 0.049 | | | 0.101** |
| | | | (0.041) | | | (0.047) |
| Conditional reverse reciprocity | | | | 0.021 | | 0.017 |
| | | | | (0.038) | | (0.038) |
| Generalised reciprocity | | | | | -0.025 | -0.062* |
| | | | | | (0.029) | (0.032) |
| Land recipient $i$ ($\times 100$) | | 0.166** | 0.167** | 0.166** | 0.166** | 0.168** |
| | | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Land donor $j$ ($\times 100$) | | -0.089 | -0.096 | -0.088 | -0.082 | -0.082 |
| | | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Observations | 87,788 | 87,788 | 87,788 | 87,788 | 87,788 | 87,788 |

Note: Probit. Dependent variable: Dummy equal to one if $i$ sends inputs to $j$. All regressions include control variables and village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

rules and therefore to be less dependent on reciprocating transactions. One possibility is that ethnically diverse and densely populated villages share the common characteristic of being more welcoming towards newcomers. This does not appear to be the case: A cross-tabulation of the two dummies puts 13 to 15 villages in each of the four cells.

Table 5 reports the estimation results for equation 3. Allowing for varying effects between high- and low-population-dense villages, an interesting finding emerges (model 1a). The interaction terms of both donor and recipient income are insignificant, and point estimates are of opposite signs compared to the non-interacted income variables. The joint test of the sum of the interaction and the income variable yields an insignificant estimate. This indicates that transactions are not driven by norm-based access rules in these villages. The leftover main effect is now significant at the 5 percent level. When splitting villages by the level of ethnic diversity, the same result holds for recipient income. This is indicative of norm-based access rules being more important in these less diverse villages. However, the effect on donor income is now entirely captured by the interaction term and is not significant for low-diversity villages.

Table 6: Community differences and pro-poorness of land transfers

| Community characteristic | None | Density | Diversity | None | Density | Diversity |
|---|---|---|---|---|---|---|
| | (1) | (1a) | (1b) | (2) | (2a) | (2b) |
| $i$'s total income | -0.351* | -0.520*** | -0.466* | | | |
| | (0.198) | (0.197) | (0.248) | | | |
| $j$'s total income | 0.441* | 0.673** | -0.151 | | | |
| | (0.257) | (0.276) | (0.293) | | | |
| Characteristic is high * $i$'s income | | 0.465 | 0.226 | | | |
| | | (0.363) | (0.323) | | | |
| Characteristic is high* $j$'s income | | -0.829 | **1.030**** | | | |
| | | (0.648) | (0.442) | | | |
| $i$'s non-agricultural income | | | | -0.439** | -0.580*** | -0.486* |
| | | | | (0.196) | (0.205) | (0.250) |
| $j$'s non-agricultural income | | | | 0.478* | 0.656** | -0.086 |
| | | | | (0.264) | (0.281) | (0.330) |
| $i$'s agricultural income | | | | 0.683** | 0.244 | -0.101 |
| | | | | (0.278) | (0.347) | (0.524) |
| $j$'s agricultural income | | | | 0.267 | 0.521 | 0.033 |
| | | | | (0.338) | (0.392) | (0.578) |
| Characteristic is high* $i$'s non-agri income | | | | | 0.375 | 0.074 |
| | | | | | (0.340) | (0.317) |
| Characteristic is high* $j$'s non-agri income | | | | | -0.685 | **0.934**** |
| | | | | | (0.648) | (0.454) |
| Characteristic is high* $i$'s agri. income | | | | | **1.131**** | **1.131*** |
| | | | | | (0.525) | (0.598) |
| Characteristic is high* $j$'s agri. income | | | | | -0.819 | 0.338 |
| | | | | | (0.757) | (0.701) |
| Observations | 87,788 | 87,788 | 87,788 | 87,788 | 87,788 | 87,788 |

Note: Tobit. Dependent variable: Amount of land $i$ receives from $j$. All regressions include control variables and village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively. Numbers in bold estimates refer to cases where interaction + main effect is significant at the 5 percent level.

While this is not necessarily evidence against the functioning of norm-based access rules in low-diversity villages, it is somewhat puzzling. A closer look at the result reveals that it is driven by a few households with very high incomes that are also located in high-ethnicity villages. When we restrict the sample to links where the donor has an income equal to or less than the 99th percentile of income, the result disappears.

When we split the income variable into agricultural income and non-agricultural income (model 2a and 2b), another interesting finding emerges: The effect on non-agricultural income is again only significant and negative in the low-density villages (model 2a). In the baseline model (2a), there was a significant and positive effect from the recipient agricultural income, which we interpreted as being due to these households behaving more commercially. This effect is entirely picked up by the high-density villages, which is where we would expect commercialized agriculture to be present. Turning to ethnic diversity, the negative effect on non-agricultural recipient income is again only present in low-diversity villages, consistent with the theory that low-diversity villages adhere more to norm-based access rules. The effect is, however, only significant at the 10 percent level. The earlier effect of a positive and significant donor income is again found to be driven by high-diversity villages and only by differences in non-agricultural income. The positive effect on recipient agricultural income also found in the baseline specification (model 2) is likewise picked up by high-diversity villages.

Results of equation 4 are shown in Table 5. Interactions between the direct reciprocity predictor and population density and ethnic diversity are not statistically significant. However, the joint test of the sum of the interaction and the income variable is statistically significant in all cases. This implies that the observable patterns of reciprocal behavior do not differ between high- and low-density villages or between more and less ethnically heterogeneous villages. At least two different interpretations can explain this result: First, it is possible that the norm of reciprocal behavior is not affected in the same way as the norm regarding access to land. Second, it is possible that villages where market structures play a greater role also have more land-for-labor barter exchanges, which are observationally inseparable from reciprocal behavior in this dataset.

## Robustness checks

In this section, we conduct a series of robustness checks. First, we might worry that household unobservables in the baseline model drive inter-household

Table 7: Community differences and direct reciprocity

| Dep. variable: | Labor network | | | Agricultural input network | | |
|---|---|---|---|---|---|---|
| | (1) | (1a) | (1b) | (2) | (2a) | (2b) |
| Direct reciprocity | 0.835*** | 0.800*** | 0.848*** | 0.582*** | 0.554*** | 0.602*** |
| | (0.091) | (0.113) | (0.124) | (0.067) | (0.081) | (0.090) |
| Direct rec.* High density | | **0.092** | | | **0.066** | |
| | | (0.186) | | | (0.134) | |
| Direct rec.* High diversity | | | **-0.029** | | | **-0.046** |
| | | | (0.182) | | | (0.132) |
| Observations | 87,788 | 87,788 | 87,788 | 87,788 | 87,788 | 87,788 |

Note: Tobit. All regressions include control variables and village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively. Numbers in bold estimates refer to cases where interaction + main effect is significant at the 5 percent level.

transactions as well as households' tendency to reciprocate by other means. Second, the dummy variables for population density and ethnic diversity were based on the median cut-off between villages. This definition is somewhat arbitrary, and thus we investigate whether the results are sensitive to this choice.

### Pro-poorness of land transactions

Table 5 reports the estimation of the baseline model using donor or recipient fixed effects, estimated by OLS. The results are consistent in sign with the results reported in Table5. Moreover, the results are at least as significant as the baseline results in Table 5, except for total donor income per capita when including recipient fixed effects in column (1). Thus, it does not appear that household unobservables are driving the main results, although recipient fixed effects do seem to influence the result somewhat.

Table 8: Land transaction results including fixed effects

| | $h$ = Sender | | $h$ = Recipient | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Total income$_h$ | 0.005 | | -0.006*** | |
| | (0.003) | | (0.002) | |
| Non-agri. income$_h$ | | 0.006* | | -0.007*** |
| | | (0.004) | | (0.002) |
| Agri. income$_h$ | | -0.002 | | 0.005 |
| | | (0.004) | | (0.003) |
| Fixed effects | Receiver FE's | Receiver FE's | Sender FE's | Sender FE's |
| Controls | Link + sender char's | Link + sender char's | Link + recipient char's | Link + recipient char's |
| Observations | 87,788 | 87,788 | 87,788 | 87,788 |

Note: OLS. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively. Link-varying controls are included in all regressions.

**Reciprocity**

To check that the result on direct reciprocity is not driven by household unobservables, we re-estimate the model using a linear probability model including two-way fixed effects. In contrast to the analysis of unobserved heterogeneity in the land transaction equation above, we are able to correct for unobserved heterogeneity including donor and recipient fixed effects in equation 2. As previously mentioned, all household-specific variables are wiped out and only dyad-specific variables can be estimated. Because our direct reciprocity predictor is dyad-specific, in contrast to the income variables above, we are able to include both household fixed effects at the same time.[22]

The estimation results are reported in the Appendix, Table A8 and A9. Including donor and recipient fixed effects significantly changes the observed result for direct reciprocity in the agricultural input market: The positive and significant estimate on direct input reciprocity in Table 5 seems to be driven by household unobservables. Thus, only direct labor reciprocity is consistent across the different specifications.

[22]In particular, we estimate: $L_{ji} = \alpha + \delta_1 A_{ij} + \delta_2 w_{ij} + a_i + a_j + \mu_{ij}$. Notation is similar to equation 6 in Section 3.

**Alternative village-level cut-offs**

Figure A1 reports parameter estimates and a 5 percent confidence band using alternative cutoffs. The topmost row shows parameter estimates using population density to define cutoffs, while the lower row uses ethnic diversity. The two first columns shows the main effects for donors and recipients, while columns (3) and (4) report the joint test of interaction plus main effect, i.e., the parameter estimate for more population-dense and ethnically diverse villages. Moving to the right in each diagram corresponds to increasing the number of villages included in the main effect.

Low density effects of both donor and recipient income are significant and of the same sign as in the baseline model for most alternative values of the cutoff. Also consistent with the previous finding, there is no effect for donors and recipients across most cutoffs in more population-dense villages. Turning to the diversity results, the main effect result of a significant estimate on recipient income (subfigure 5) is not consistently negative or significant across cutoffs. It is, however, almost significant regardless of the cutoff. Low ethnic diversity does not appear to motivate donors of land further. The effect is significantly negative when few villages are covered by the main effect. However, this appears to be driven by a single outlying village, corresponding to the large negative spike in the figure. The somewhat puzzling finding of a highly significant and positive estimate of donor income in high diversity villages is robust – and the effect becomes increasingly larger when only the most ethnic diverse villages are included.

Figure A2 reports parameter estimates for direct reciprocity in the labor and agricultural input market, as well as a 5 percent confidence band using alternative cutoffs. Both the low and high density effects in the labor and input market are statistically significant and of the same sign as in Table 5 independent of alternative cutoff values. Turning to the ethnic diversity results, the main effect for more ethnically homogeneous villages and the joint effect for more ethnic heterogeneous villages are positive and significant throughout. Moreover, the magnitude of the estimated effects are quite stable across alternative cutoffs.

# 6 Conclusion

This paper investigated the strength of norms in a traditional West African land tenure system using a dataset which covers 52 rural villages in The Gambia. Specifically, we test whether land transactions are motivated by social security considerations driven by norm-based access rules and whether a land transaction comes with an obligation to directly or indirectly reciprocate using either the labor or agricultural input market. We proceed to test whether the strength of the norm-based access rule to land and reciprocity patterns differ depending on population density and ethnic diversity. In contrast to many previous studies on inter-household transactions, we are less affected by problems due to omitted variable bias, as information on both recipient and donors were included. Second, we check the extent of endogeneity of income to land transactions. We find that additional landholdings do increase monetary income, but the effect sizes are small. Finally, we are able to correct for unobserved household heterogeneity.

There are three main findings of the paper:

First, inter-household land transactions are found to be at least partly driven by social security considerations, as poor and landless households receive more land. We further find that the norm-based access rule only kicks in for the poorest households, those in the lowest income quartile. Donors are more likely to be rich households, i.e., those in the uppermost income quartile.

Second, we find evidence in support of direct reciprocity of labor, whereas direct reciprocity using the agricultural input market is no longer statistically significant when we correct for unobserved heterogeneity. To examine indirect reciprocity we develop different reciprocity predictors, which are then tested using the labor and input markets, respectively. We find no evidence that households are more inclined to give to those that have been generous towards others. This result is in contrast to findings of previous experimental studies. In conclusion, many land transfers are directly reciprocated, but the system of land transfers does not appear to be upheld by a complex system involving indirect reciprocity.

Third, we find that transaction behavior is different in villages characterized by higher population density and ethnic diversity, where land does not flow towards relatively poorer households. The former result is in line with land scarcity undermining traditional access rules to land (Platteau, 2002). Despite how long-time cooperation between ethnic groups who live side by side in the same villages and grow the same crops under similar environmental conditions has resulted in relatively homogeneous farming systems (Dey, 1982; Eastman, 1990), it seems that ethnic heterogeneity still has the potential to undermine community-level social security systems. This result may be explained by higher levels of trust in more ethnically homogeneous villages. In contrast, we did not find evidence that population pressure or high ethnic diversity influences patterns of direct reciprocity. In contrast to Jaimovich (2013) who found that increased market integration in the form of links with households outside the village reduced reciprocal behavior, we do not find that increased population pressure changes the pattern of direct reciprocity between villagers, even though this is also thought to increase market integration.

These results are comparable to previous empirical studies. For example, Devereux (1999, 2001) argues that traditional practices of redistribution from wealthier to poorer households are rapidly disappearing due to commercialization. This is exactly what we find in more population-dense villages, where commercialization of the land market supposedly is higher. However, the results of this paper also imply that community security systems in terms of access to land are still in effect in less densely populated villages in rural Gambia. This result suggests that the displacement of informal safety nets should be taken into account when land redistribution reforms are implemented. Redistribution of land towards landless households is likely to crowd out private inter-household transfers of land to the rural poor. As population density rises and land access becomes increasingly market based, the need for land redistribution reforms is likely to increase as the effect of norm-based access rules declines. There are concerns that formal land redistribution schemes intended to increase equality may have perverse outcomes (Bourlès and Bramoullé, 2013) in the form of increased inequality as the population density increases and where ethnic diversity is high.

What should one make of the results on indirect reciprocity? First, the evidence does not imply that households do not behave more nicely towards people that behave generously towards others, or that indirect reciprocity may not be an important balancing mechanism underlying community security systems. But indirect reciprocity should not be taken for granted in the explanation of inter-household transfers. Second, it is possible that we looked for indirect reciprocity in the wrong place. Perhaps land transactions are reciprocated in other markets or through other means, such as loyalty, respect, and devotion. Third, we acknowledge that our data have certain limitations. One potential limitation is that we do not know the sequence of the transactions and thus are unable to interpret the results of reciprocity causally. These caveats notwithstanding, we take the lack of evidence in support of indirect reciprocity as an indication that the network structures underlying village-based social security systems are less complicated than often believed.

# Appendix A

Table A1: Household-level income regression

| | (1) | |
|---|---|---|
| Cultivated land (log) | 0.028** | (0.011) |
| Household size (log) | -0.452*** | (0.019) |
| Observations | 2,028 | |
| R-squared | 0.612 | |

Dependent variable: Log of realized total income per capita. All regressions include household-level control variables similar to the subsequent dyad-regression as well as village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

Table A2: Descriptive statistics: village-level

| | Mean | Std. Dev. | Min. | Max. |
|---|---|---|---|---|
| Inhabitants | 521.23 | 184.46 | 130 | 1,077 |
| Area (hectares) | 402.16 | 353.32 | 20.77 | 1751.31 |
| Density (inhabitants per square km) | 234.62 | 182.58 | 25.75 | 980.14 |
| Ethnic diversity index (= 0 if homogenous) | 0.275 | 0.235 | 0 | 0.837 |
| Illiteracy share | 0.451 | 0.207 | 0 | 0.913 |
| No access to electricity | 0.970 | 0.042 | 0.837 | 1 |
| No private toilettes | 0.372 | 0.296 | 0 | 1 |
| Not improved water access | 0.912 | 0.131 | 0.431 | 1 |
| Gini (based on self-declared income per cap.) | 0.314 | 0.112 | 0.144 | 0.601 |
| Observations | 52 | | | |

Note: Density is calculated based on the sum of the cultivated and the inhabited village area.

Table A3: Descriptive statistics: Dyad-level

| | Mean | Std. Dev. | Min. | Max. |
|---|---|---|---|---|
| Receives land dummy | 0.010 | 0.098 | 0 | 1 |
| Amount of land received (hec.) | 1.828 | 1.496 | 0.25 | 20 |
| Send labor dummy | 0.014 | 0.116 | 0 | 1 |
| Send agricultural input dummy | 0.022 | 0.146 | 0 | 1 |
| Kinship tie | 0.129 | 0.335 | 0 | 1 |
| Direct labor reciprocity | 0.010 | 0.098 | 0 | 1 |
| Reverse indirect reciprocity | 0.087 | 0.282 | 0 | 1 |
| Generalised indirect reciprocity | 0.193 | 0.395 | 0 | 1 |
| Reverse indirect reciprocity (unconditional) | 0.221 | 0.415 | 0 | 1 |
| Observations | 87,788 | | | |

Table A4: Baseline results: Control variables

| | (1) | | (2) | | (3) | |
|---|---|---|---|---|---|---|
| Kinship tie | 1.550*** | (0.263) | 1.549*** | (0.261) | 1.547*** | (0.260) |
| *Recipient characteristics (i)* | | | | | | |
| Share of same ethnicity | -1.228*** | (0.295) | -1.242*** | (0.299) | -1.224*** | (0.293) |
| No land | 1.438*** | (0.312) | 1.496*** | (0.313) | 1.511*** | (0.314) |
| Age of head (log) | -0.004 | (0.005) | -0.004 | (0.005) | -0.003 | (0.005) |
| Adult labor (log) | 0.033 | (0.020) | 0.028 | (0.019) | 0.027 | (0.019) |
| Illiterate | -0.085 | (0.134) | -0.084 | (0.134) | -0.078 | (0.135) |
| Formal school dummy | -0.528** | (0.229) | -0.510** | (0.227) | -0.497** | (0.230) |
| Female head | -0.584 | (0.362) | -0.559 | (0.360) | -0.542 | (0.360) |
| Receive remittances | 0.206 | (0.189) | 0.216 | (0.184) | 0.213 | (0.183) |
| *Donor characteristics (j)* | | | | | | |
| Share of same ethnicity | 1.838** | (0.788) | 1.841** | (0.786) | 1.834** | (0.796) |
| No land | -4.219*** | (0.510) | -4.239*** | (0.506) | -4.217*** | (0.507) |
| Age of head (log) | 0.021*** | (0.006) | 0.021*** | (0.006) | 0.020*** | (0.006) |
| Adult labor (log) | 0.067** | (0.027) | 0.069** | (0.028) | 0.064** | (0.028) |
| Illiterate | -0.583** | (0.243) | -0.575** | (0.241) | -0.579** | (0.245) |
| Formal school dummy | 0.040 | (0.327) | 0.035 | (0.334) | 0.033 | (0.339) |
| Female head | -0.176 | (0.726) | -0.193 | (0.715) | -0.195 | (0.722) |
| Receive remittances | 0.107 | (0.231) | 0.105 | (0.231) | 0.110 | (0.232) |
| Observations | 87,788 | | 87,788 | | 87,788 | |

Dependent variable: Amount of land $i$ receives from $j$. continuation of Table 3. All regressions include village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

Table A5: Baseline results: initial land holdings

| | (1) | | (2) | | (3) | |
|---|---|---|---|---|---|---|
| $i$'s non-agricultural income | -0.439** | (0.196) | -0.301* | (0.177) | -0.435** | (0.197) |
| $j$'s non-agricultural income | 0.478* | (0.264) | 0.295 | (0.294) | 0.404 | (0.278) |
| $i$'s agricultural income | 0.683** | (0.278) | 0.569** | (0.281) | 0.661** | (0.279) |
| $j$'s agricultural income | 0.267 | (0.338) | 0.284 | (0.359) | 0.259 | (0.343) |
| No land (dummy): $i$ | 1.496*** | (0.313) | | | 1.410*** | (0.299) |
| No land (dummy): $j$ | -4.239*** | (0.506) | | | -3.764*** | (0.452) |
| Initial land holdings (hec.): $i$ | | | -0.014 | (0.012) | -0.008 | (0.009) |
| Initial land holdings (hec.): $j$ | | | 0.026*** | (0.006) | 0.022*** | (0.006) |
| Observations | 87,788 | | 87,788 | | 87,788 | |

Note: Tobit. Dependent variable: Amount of land $i$ receives from $j$. All regressions include control variables and village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table A6: Partial effects: Reciprocity results – labor

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Direct reciprocity | 0.048*** | 0.047*** | 0.048*** | 0.047*** | 0.049*** | 0.050*** |
| | (0.010) | (0.093) | (0.010) | (0.001) | (0.011) | (0.011) |
| Reverse reciprocity | | | 0.001 | | | 0.001 |
| | | | (0.001) | | | (0.001) |
| Conditional reverse reciprocity | | | | 0.000 | | 0.001 |
| | | | | (0.001) | | (0.001) |
| Generalised reciprocity | | | | | 0.001 | 0.001 |
| | | | | | (0.0001) | (0.001) |
| Land recipient $i$ (×100) | | 0.003** | 0.003** | 0.003** | 0.003** | 0.003** |
| | | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Land donor $j$ (×100) | | 0.003* | 0.003* | 0.003* | 0.003 | 0.003 |
| | | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Observations | 87,788 | 87,788 | 87,788 | 87,788 | 87,788 | 87,788 |

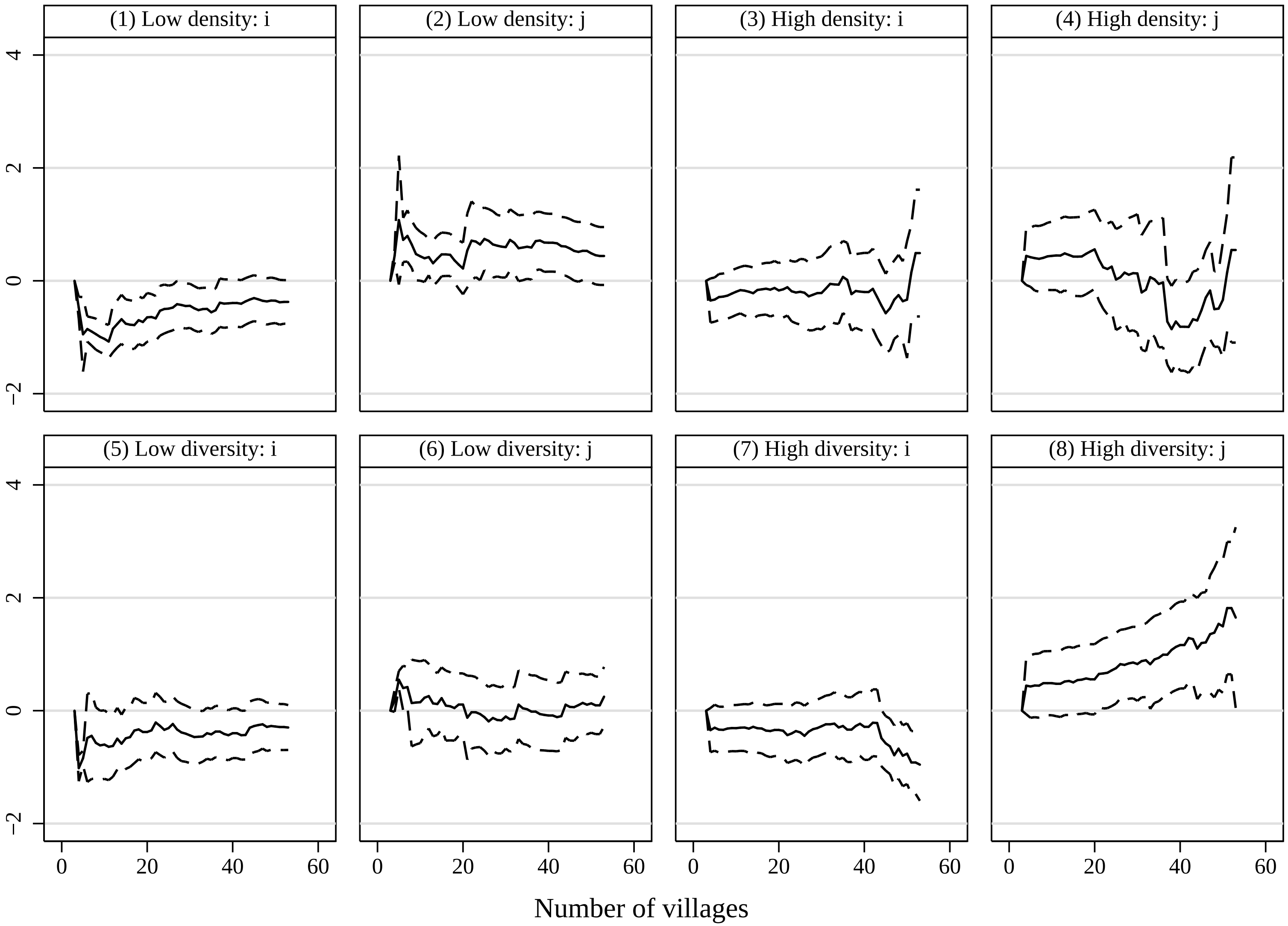
Note: Probit. Partiel effects. Dependent variable: Dummy equal to one if $i$ sends labor to $j$. All regressions include control variables and village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table A7: Partial effects: Reciprocity results – input

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Direct reciprocity | 0.035*** | 0.036*** | 0.037*** | 0.036*** | 0.035*** | 0.036*** |
| | (0.007) | (0.007) | (0.007) | (0.007) | (0.007) | (0.007) |
| Reverse reciprocity | | | 0.002 | | | 0.003* |
| | | | (0.001) | | | (0.002) |
| Conditional reverse reciprocity | | | | 0.001 | | 0.001 |
| | | | | (0.001) | | (0.001) |
| Generalised reciprocity | | | | | -0.001 | -0.002* |
| | | | | | (0.001) | (0.001) |
| Land recipient $i$ ($\times$100) | | 0.005** | 0.005** | 0.005** | 0.0025** | 0.0052** |
| | | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Land donor $j$ ($\times$100) | | -0.003 | -0.003 | -0.003 | -0.003 | -0.003 |
| | | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Observations | 87,788 | 87,788 | 87,788 | 87,788 | 87,788 | 87,788 |

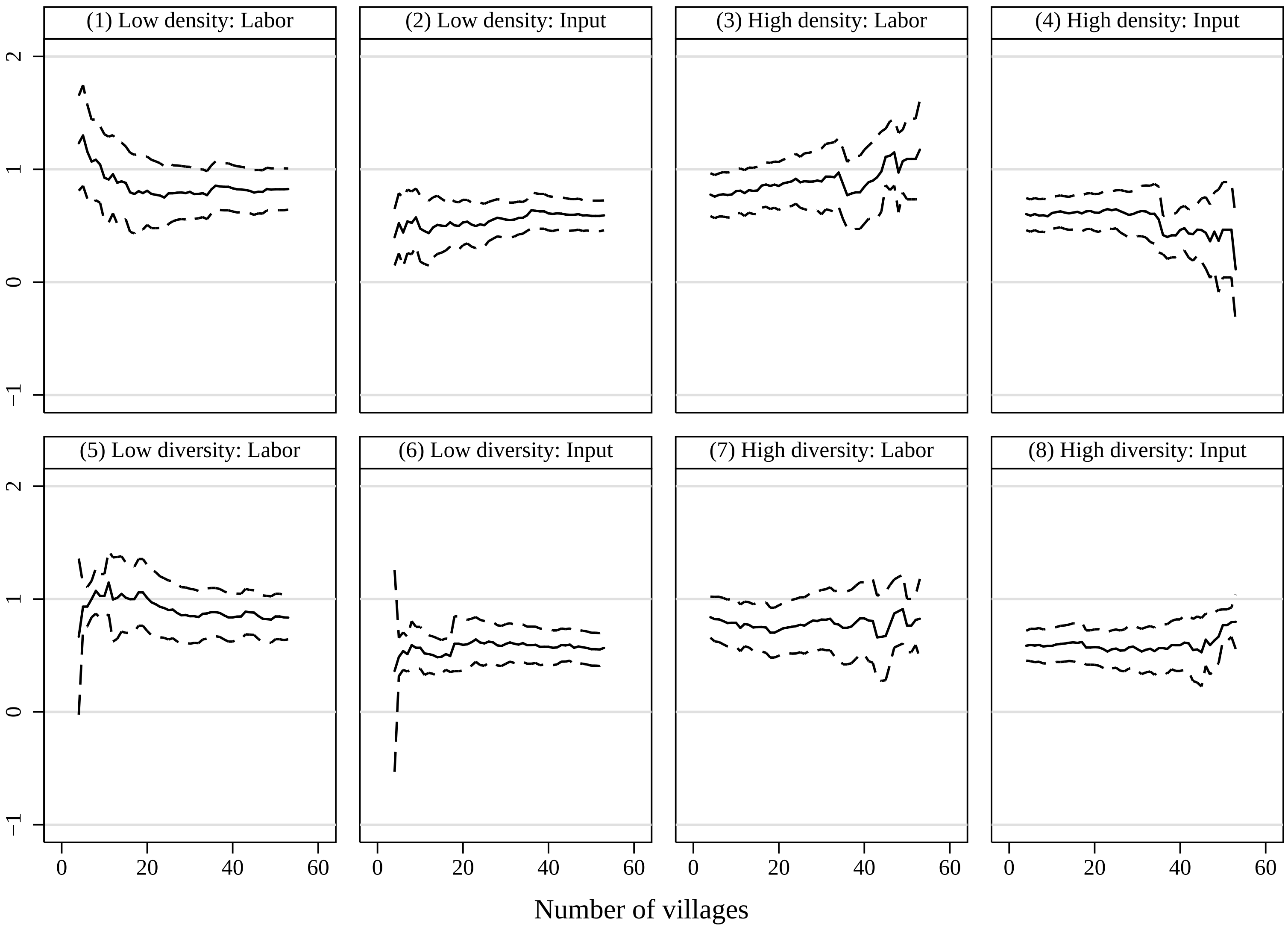
Note: Probit. Partiel effects. Dependent variable: Dummy equal to one if $i$ sends inputs to $j$. All regressions include control variables and village fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Figure A1: Alternative cutoffs on village characteristics: Land transactions




Note: $i$ and $j$ denote receivers and senders of land, respectively. The x-axis denotes the number of villages not included in the interaction, ranked by either density or ethnic diversity. In order to avoid spurious results, results are only shown for cut-offs where there are at least four villages in both main and interaction term.

Figure A2: Alternative cutoffs on village characteristics: Reciprocity




Note: The x-axis denotes the number of villages not included in the interaction, ranked by either density or ethnic diversity. In order to avoid spurious results, results are only shown for cut-offs where there are at least four villages in both main and interaction term.

Table A8: Reciprocity results including fixed effects - labor

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Direct reciprocity | 0.826*** | 0.081*** | 0.088*** | 0.083*** | 0.089*** | 0.085*** |
| | (0.093) | (0.014) | (0.015) | (0.015) | (0.015) | (0.015) |
| Reverse reciprocity | | | | | 0.000 | 0.002 |
| | | | | | (0.003) | (0.002) |
| Conditional reverse reciprocity | | | | | 0.000 | 0.001 |
| | | | | | (0.001) | (0.001) |
| Generalised reciprocity | | | | | 0.004** | 0.002 |
| | | | | | (0.002) | (0.002) |
| Estimation method | Probit | OLS | OLS | OLS | OLS | OLS |
| Controls | Yes | Yes | No | Yes | No | Yes |
| Initial land controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Fixed effects | Village | Village | Two-way | Two-way | Two-way | Two-way |
| Observations | 87,788 | 87,788 | 87,788 | 87,788 | 87,788 | 87,788 |
| R-squared | 0.131 | 0.028 | 0.011 | 0.024 | 0.011 | 0.024 |

Note: LPM. Dependent variable: Dummy equal to one if $i$ sends labor to $j$. Recipient and donor fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table A9: Reciprocity results including fixed effects - input

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Direct reciprocity | 0.592*** | 0.061*** | -0.000 | -0.000* | 0.000 | 0.000 |
| | (0.067) | (0.011) | (0.000) | (0.000) | (0.000) | (0.000) |
| Reverse reciprocity | | | | | -0.000** | -0.000** |
| | | | | | (0.000) | (0.000) |
| Conditional reverse reciprocity | | | | | 0.000 | 0.000 |
| | | | | | (0.000) | (0.000) |
| Generalised reciprocity | | | | | 0.000* | 0.000* |
| | | | | | (0.000) | (0.000) |
| Estimation method | Probit | OLS | OLS | OLS | OLS | OLS |
| Controls | Yes | Yes | No | Yes | No | Yes |
| Initial land controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Fixed effects | Village | Village | Two-way | Two-way | Two-way | Two-way |
| Observations | 87,788 | 87,788 | 87,788 | 87,788 | 87,788 | 87,788 |
| R-squared | 0.144 | 0.041 | 0.996 | 0.996 | 0.996 | 0.996 |

Note: LPM. Dependent variable: Dummy equal to one if $i$ sends inputs to $j$. Recipient and donor fixed effects. Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

# Appendix B

In the main text, equation 1 is estimated using a tobit model due to the large number of zero observations. This estimation approach is consistent with a corner-solution type of argument: Positive transfers are observed only when the latent transfer exceeds the cost of participating in the land market. Like a regular OLS, the tobit model assumes that the same mechanism is driving the decision to participate in the land market and the amount of land transferred conditional on participating. However, contrary to the OLS, the tobit model explicitly accounts for the non-linearity caused by the corner solution of zero land transfers. However, this comes at a cost of additional parametric assumptions. It is therefore of interest to investigate the robustness of the result in the main text alongside models with other or less restrictive assumptions.

Table B1, column (1) corresponds to the baseline specification in Table 5, column (2). In column (3), the same model is re-estimated using an OLS model. Results are shown in columns (1) and (3). To compare the results, the average partial effects for the tobit model are shown in column (2). Sign, magnitude and significance are consistent for the two estimators with the one exception that the coefficient on receivers' non-agricultural income is no longer statistically significant when estimated by OLS (column (3)).

One reason why the tobit model may fail is due to the assumption of a single parameter affecting both the decision to transfer and the magnitude of the transfer, conditional on transferring. In the absence of suitable instruments to separately identify the two effects, we take a simpler approach by estimating separate regressions for having a land transfer and the size of the transfer, conditional on participation. An important assumption underlying the two-part models is that the two residuals are independent (i.e. the unobservables which affect the decision to participate are independent of the unobservables that affect the decision of how much to transfer) and that errors are normally distributed with a constant variance and zero mean. We estimate both a truncated normal hurdle model proposed by Cragg (1971) and a log-normal hurdle model. The advantage of these models is that the effects of the explanatory variables are allowed to differ across the two dimensions of the transfer decision. The log-normal hurdle model is preferred as the normality assumption is more likely to be fulfilled.

Estimation results are reported in Table B1, columns (6) and (7). Apart from the coefficient estimate on senders' ($j$) agricultural income, the signs of the coefficient estimates are similar (this is also the case for the control variables). This result lends credibility to the assumption that the same mechanism is likely to drive both decisions. The positive estimate on senders' agricultural income suggests that households with a higher agricultural income are more likely to participate in the land market, but conditional on transferring less land. The variables of primary interest are the sender's and receiver's non-agricultural income. The result suggests that low income households are more likely to participate in the market, and also receive more land. In contrast, high-income senders are more likely to participate, whereas no significant difference is found between participating senders in terms of the amount of land they send. In summary, the baseline results are consistent with the same level of significance and sign across different estimation techniques.

Table B1: Baseline results – estimation technique robustness

| | (1) | (2) | (3) | (4) | (5) | (6) | | (7) | |
|---|---|---|---|---|---|---|---|---|---|
| | Tobit | Tobit | OLS | Tobit | OLS | Normal hurdle | | Log-normal hurdle | |
| | | Partial effects | | | | Probit | OLS | Probit | OLS (log) |
| $i$'s non-agri. income | -0.439** | -0.004** | -0.007*** | -0.439** | -0.007*** | -0.076** | -0.837* | -0.076** | -0.121** |
| | (0.196) | (0.002) | (0.002) | (0.196) | (0.002) | (0.038) | (0.432) | (0.038) | (0.059) |
| $j$'s non-agri. income | 0.478* | 0.005 | 0.006* | 0.478* | 0.006* | 0.098* | -0.028 | 0.098* | 0.055 |
| | (0.264) | (0.003) | (0.004) | (0.264) | (0.004) | (0.053) | (0.210) | (0.053) | (0.056) |
| $i$'s agri. income | 0.683** | 0.007* | 0.005 | 0.683** | 0.005* | 0.142** | 0.196 | 0.142** | 0.027 |
| | (0.278) | (0.003) | (0.003) | (0.279) | (0.003) | (0.062) | (0.385) | (0.062) | (0.079) |
| $j$'s agri. income | 0.267 | 0.002 | -0.002 | 0.267 | -0.001 | 0.063 | -0.638 | 0.063 | -0.193* |
| | (0.338) | (0.003) | (0.004) | (0.338) | (0.004) | (0.069) | (0.455) | (0.069) | (0.114) |
| Observations | 87,788 | 87,788 | 87,788 | 87,138 | 87,138 | 87,788 | | 87,138 | 859 |

Note: Dependent variable is the amount of land $i$ receives from $j$, except for truncated normal hurdle first tier estimation and probit estimation where the dependent variable is equal to one if receives land from , and zero otherwise. All regressions include control variables and village fixed effects. Marginal effects reported in column (2). Standard errors are clustered on the village level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

## References


Alesina, A., R. Baqir and W. Easterly (1999): "Public goods and ethnic divisions," *The Quarterly Journal of Economics*, 114, 1243–1284.

Arcand, J.-L. and D. Jaimovich (2012): "Does ethnic diversity decrease economic interactions? Evidence from exchange networks in rural Gambia," .

Bank, W. (1993): "The Gambia - An assessment of Poverty," Report No. 11941-GM.

Beck, U. R. and B. Bjerge (2014): "Efficiency of land markets: Network level evidence of the importance of social ties," *Working paper*.

Besley, T. and R. Burgess (2000): "Land reform, poverty reduction, and growth: Evidence from India," *The Quarterly Journal of Economics*, 115(2), 389–430.

Bolton, G. E., E. Katok and A. Ockenfels (2005): "Cooperation among strangers with limited information about reputation," *Journal of Public Economics*, 89, 1457–1468.

Boserup, E. (1965): *The Conditions of Agricultural Growth: The Economics of Agrarian Change under Population Pressure*. London: George Allen & Unwin LTD.

Bourlès, R. and Y. Bramoullé (2013): "Altruism in Networks," Working Paper Nr 56.

Cashdan, E. A. (1985): "Coping with Risk: Reciprocity Among the Basarwa of Northern Botswana," *Man (N.S.)*, 20, 454–4745.

Chavas, J.-P., R. Petrie and M. Roth (2005): "Farm Household Production Efficiency: Evidence from The Gambia," *American Journal of Agricultural Economics*, 87, 160–179.

Cox, D. (1987): "Motives for Private Income Transfers," *Journal of Political Economy*, 95, 508–546.

Cox, D. and G. Jakubson (1995): "The connection between public transfers and private interfamily transfers," *Journal of Public Economics*, 57, 129–167.

Coxa, D., B. E. Hansenb and E. Jimenez (2004): "How responsive are private transfers to income? Evidence from a laissez-faire economy," *Journal of Public Economics*, 88, 2193–2219.

Dercon, S. and J. D. Weerdt (2006): "Risk-sharing networks and insurance against illness," *Journal of Development Economics*, 81.

Devereux, S. (1999): "'Making less last longer': Informal safety nets in Malawi," ICS Discussion Paper 373.

——— (2001): "Livelihood Insecurity and Social Protection: A Re-emerging Issue in Rural Development," *Development Policy Review*, 19(4), 507–519.

Dey, J. (1982): "Development Planning in The Gambia: The Gap between Planners' and Farmers' Perceptions, Expectations and Objectives," *World Development*, 10(5), 377–396.

Eastman, C. (1990): "Gambian usufruct tenure. Help or hindrance to agricultural production?," *Land Use Policy*, pp. 27–40.

Fafchamps, M. and S. Lund (2003): "Risk-sharing networks in rural Philippines," *Journal of Development Economics*, 71, 261–287.

Fafchamps, M. and M. Söderbom (2014): "Network Proximity and Business Practices in African Manufacturing," *World Bank Economic Review*, 28, 99–129.

Fehr, E. and A. Falk (2002): "Psychological foundations of incentives," *European Economic Review*, 45, 687–724.

Fehr, E. and S. Gašchter (2000): "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspective*, 14, 159–181.

Fehr, E. and K. M. Schmidt (2006): *The Economics of fairness, reciprocity and altruism - experimental evidence and new theories*chap. 8.

Fenske, J. (2011): "Land tenure and investment incentives: Evidence from West Africa," *Journal of Development Economics*, 95, 137–156.

Fong, C., S. Bowles and H. Gintis (2006): *Strong reciprocity and the welfare state*. In: Kolm, S.-C., Mercier Ythier, J. (Eds.), Handbook of Economics of Giving, Altruism and Reciprocity, vol. 1. Elsevier.

Freudenberger, M. S. (1993): "Institutions and natural resource management in The Gambia: A case study of the Foni Jarrol district," LTC Research Paper 114.

——— (2000): "Tenure and Natural Resources in The Gambia: Summary of Research Findings and Policy Options," Land Tenure Center. University of Wisconsin-Madison. Working paper no. 40.

Gajigo, O. and A. Saine (2011): "The effects of government policies on cereal consumption pattern change in the Gambia," *Review of African Political Economy*, 38(130), 517–536.

Gisselquist, R. M. (2013): "Ethnic divisions and public goods provision, revisited," *Ethnic and Racial Studies*, pp. 1–23.

Jaimovich, D. (2011): "Macrostructure and microstructure: Evidence from overlapping village networks in The Gambia," *Working Paper*.

——— (2013): “Missing links, missing markets: Internal exchanges, reciprocity and external connections in the economic networks of Gambian villages,” MPRA Paper No. 44080.

Kazianga, H. (2006): “Motives for household proivate transfers in Burkina Faso,” *Journal of Development Economics*, 79, 73–117.

Kolm, S.-C. (2006): *Introduction to the Economics of Giving, Altruism and Reciprocity*. In: Kolm, S.-C., Mercier Ythier, J. (Eds.), Handbook of Economics of Giving, Altruism and Reciprocity, vol. 1. Elsevier.

Mazzocco, M. and S. Saini (2012): “Testing Efficiency Risk Sharing witrh Heterogenous Risk Preferences,” *American Economic Review*, 1, 428–468.

Mitrut, A. and K. Nordblom (2010): “Social norms and gift behavior: Theory and evidence from Romania,” *European Economic Review*, 54, 998–1015.

Morduch, J. (1995): “Income Smoothing and Consumption Smoothing,” *Journal of Economic Perspectives*, 9(3), 103–114.

Nowak, M. and K. Sigmund (1998): “Evolution of indirect reciprocity by image scoring,” *Journal of Theoretical Biology*, 194, 561–574.

Pamela, K. (2010): *Land, labour and entrustment : West African female farmers and the politics of difference*. Leiden, Boston.

Platteau, J.-P. (1991): *Traditional Systems of Social Security and Hunger Insurance: Past Achievements and Modern Challenges*pp. 112–70. Clarendon Press, Oxford,.

——— (2002): “The gradual erosion of the social security function of customary land tenure arrangements in lineage-based societies,” Discussion Paper No. 2002/26.

——— (2006): *Solidarity norms and institutions in village societies: static and dynamic considerations*. In: Kolm, S.-C., Mercier Ythier, J. (Eds.), Handbook of Economics of Giving, Altruism and Reciprocity, vol. 1. Elsevier.

Platteau, J.-P. and P. G. Sekeris (2010): “On the feasibility of power and status ranking in traditional setups,” *Journal of Comparative Economics*, 38, 267–282.

Sacco, P. L., P. Vanin and S. Zamagni (2006): *The economics of human relationships*. In: Kolm, S.-C., Mercier Ythier, J. (Eds.), Handbook of Economics of Giving, Altruism and Reciprocity, vol. 1. Elsevier.

Seinen, I. and A. Schram (2006): “Social status and group norms: Indirect reciprocity in a repeated helping experiment,” *European Economic Review*, 50, 581–602.

STAMM, V. (2004): "The World Bank on Land Policies: A West African Look at the World Bank Policy Research Report," *Journal of the International African Institute*, 74, 670–678.

SWINDELL, K. (1987): "Family Farms and Migrant Labour: The Strange Farmers of the Gambia," *Canadian Journal of African Studies*, 12(1), 3–17.

VON BRAUN, J. AND P. J. R. WEBB (1989): "The Impact of New Crop Technology on the Agricultural Division of Labor in a West African Setting," *Economic Development and Cultural Change*, 37, 513–534.

WOODHOUSE, P. (2003): "African Enclosures: A Default Mode of Development," *World Development*, 31, 1705–1720.

ZAK, P. J. AND S. KNACK (2001): "Trust and Growth," *The Economic Journal*, 111, 295–321.

# Chapter 4

# The importance of structural mechanisms in network formation: The case of rural Gambia

Benedikte Alkjærsig Bjerge*

**Abstract**

This paper investigates the explanatory value of network architecture. I allow for endogenous networking mechanisms to overcome the assumption of dyad independence using the recent development of exponential random graph (ERG) models. Structural mechanisms in the form of reciprocity and transitivity are found to be important determinants of network formation. Inability to account for structural mechanisms leads to upward biased parameter estimates of household attributes and dyad-specific characteristics, including kinship. This suggests that households forming exchange networks take into account the structure of the community network and the structure resulting from additional partnerships.

**Keywords:** exponential random graph models, network architecture, structural mechanisms, inter-household transactions

**JEL classifications:** A14, D85, O12

*Benedikte Alkjærsig Bjerge (benedikte.bjerge@econ.ku.dk) is employed at the Department of Economics, University of Copenhagen. The author thanks John Rand, Marianne Hvidt, and participants at the development economic seminar at the University of Copenhagen for their helpful comments. I also thank Dany E. Jaimovich and the research team, including principal investigator Jean-Louis Arcand, investigators Yiu Por Chen and Yong He, and finally, field coordinators Cheikh Ibrahima Fall Diop, Eric Djimeu Wouabe, Malek Garbouj, and Slavica Zec for making the household survey available to us.

# 1 Introduction

From the research of Jackson and Wolinsky (1996) and Bala and Goyal (2000), economists have come to a general understanding that the entire graph is relevant in network formation. While this burgeoning theoretical literature has modeled formation of social networks endogenously (for a review see Jackson, 2008), the empirical literature on link formation within development economics increasingly uses dyadic regressions.[1] The main advantage of dyadic regressions is that it takes the pair of actors (dyad) as the unit of analysis rather than the individual. However, the approach relies on the assumption of dyad independence. This assumption is particularly unrealistic when two dyads involve the same actor: Tie formation between actors $i$ and $j$ is likely to depend on the link between actors $i$ and $k$, and vice versa. Moreover, dyadic regressions exclusively consider attributes of the nodes and characteristics of the links in the analysis of network formation. This leaves out the potential for endogenous networking mechanisms that are only indirectly related to nodes and link-specific characteristics. For example, it excludes the well-established notion that human beings value symmetry in relations: the tendency for friendship to be returned (reciprocity) and for friends of friends to befriend one another (triadic closure). Ignoring reciprocity and triadic closure, one is likely to overestimate any tendency towards homophily as reciprocated ties and closed triangles among members of the same category are alone attributed to homophily (i.e., Mayer and Puller, 2008; Goodreau, Kitts and Morris, 2009).

The aim of this paper is to examine the consequences of dyad independence between pairs by considering the importance of endogenous networking mechanisms. I consider exponential family models in order to examine the architecture of the network and refer to them as exponential random graph (ERG) models. The advantage of the ERG modeling framework lies in its capacity to represent the effects of complex local network structures under the assumption of dyad dependence across links. This class of models has a long history in the network literature, while it is less studied by economists. Part of the explanation for this is that ERG models are only able to answer questions related to how a network forms, while economists traditionally have been more interested in the question of why the network forms (Jackson, 2008).

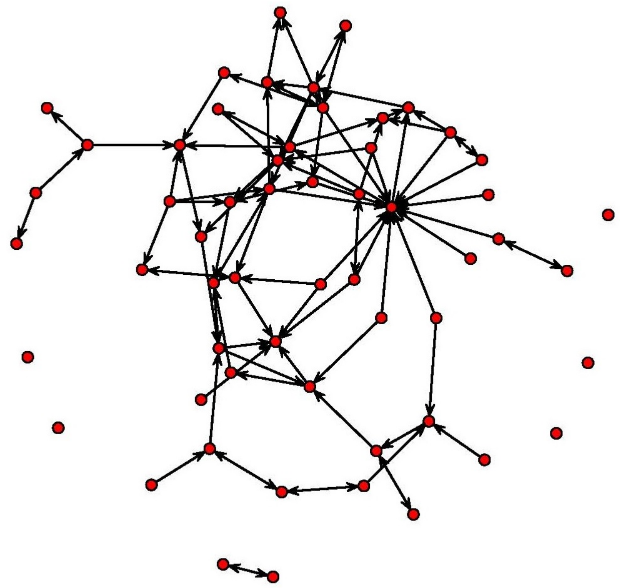
[1] An estimating equation is said to be dyadic if each observation corresponds to a pair of nodes. Examples of studies estimating dyadic regressions to study network formation in rural communities in developing countries include Fafchamps and Gubert (2007); Arcand and Fafchamps (2012); Dercon and Weerdt (2006); Conley and Udry (2010).

The primary contribution of this paper is to further understand the formation mechanisms that influence the economic composition of networks by distinguishing traditional characteristics (such as age and ethnicity) from endogenous structural mechanisms. I focus on two global structural mechanisms shown to be important across different sociodemographic settings (Snijders, Pattison, Robins and Handcook, 2006; Robins, Snijders, Wang, Handcock and Pattison, 2007; Goodreau, 2007; Goodreau, Kitts and Morris, 2009; Wimmer and Lewis, 2010): reciprocity and transitivity. Direct reciprocity occurs when the friendship nomination by the sender is reciprocated by the nominee. In the economic literature, reciprocity is argued to be a relevant enforcement mechanism underlying risk-sharing arrangements in village societies (Platteau, 2006). Beck and Bjerge (2014) also consider the importance of direct reciprocity related to the norm-based rule for access to land using the same dataset as the present study. They find that land transactions come with an obligation to reciprocate in the labor market. Transitivity refers to the tendency for friends of friends to become friends, leading to closed networks (i.e., triads).[2] These provide agents with dyadic constraints and facilitate monitoring and enforcement through balancing mechanisms. I investigate whether actors choose their economic network partners solely on the basis of personal attributes, or whether the potential network structure resulting from their additional partnerships is also an underlying consideration.

This paper focuses on network data collected in a large number of rural villages in The Gambia. The nodes represent households, and the edges indicate a transaction in the form of labor, land, or agricultural input, and in what direction the transaction was made. This setting is ideal, as rural societies in developing countries are highly dependent on their personal network of contacts. Particularly in the absence of economic and financial institutions, informal social arrangements are likely to drive market and non-market transactions within the villages, and personal relations therefore carry a certain economic value to the individual household as well as to the society as a whole. Figure 1 depicts one such network of transactions between rural households in The

[2]This paper use the term "reciprocity" to describe direct reciprocity. Indirect reciprocity (transfer returned to the sender by a non-recipient third party) is a subset of the concept of triadic closure. For instance, consider reverse reciprocity: If $i$ transfers to $j$ and $k$ transfers to $i$, then $j$ might decide to transfer to $k$, which would result in a closed triangle. However, triangles may also exist in absence of indirect reciprocity: If $i$ transfers to both $j$ and $k$, either or both $j$ and $k$ may decide to transfer to each other. In the remaining part of this paper, I focus exclusively on triangles of which indirect reciprocity examined in more detailed by Beck and Bjerge (2014) is only a subset. The lack of evidence in support of indirect resiprocity in Beck and Bjerge (2014) may be due to the inability to account for dyad dependence.

Figure 1: Transaction network



Gambia; the arrows on the edges represent the direction of the transaction.

This paper makes two contributions. First, it applies an alternative method to model network formation in rural villages, showing the importance of dyad dependence. The large dataset covering many rural villages provides an unusual opportunity for empirical replication across 37 villages to function as robustness for the claims made. Recent studies that examine network formation in a development economic setting only provide evidence for a single village or a limited group and in many cases suffer from incomplete network data (Hiwatari, 2010; Potter and Handcock, 2010).[3] While the literature on friendship networks in schools does better in terms of using more than one case for empirical replication (Goodreau, Kitts and Morris, 2009), these studies generally suffer from the limit on the number of friends a student can nominate to be his or her friend.[4]

[3]Hiwatari (2010) investigate whether actors in rural Uzbekistan choose their ROSCA (Rotating Savings and Credit Associations) partners using a network comprising 45 nodes, while Potter and Handcock (2010) offer insights on within-family resource exchange using incomplete household networks in a single Malawian village. A different set of empirical studies use affiliation networks to create social network data for the study of network formation of animal traders. For instance, Ortiz-Pelaez, Ashenafi, Roger and Waret-Szkuta (2012) create a trading network linking a selected set of ten farmers from 75 different villages in Ethiopia to ten markets. A limitation of this study is that is does not include higher-order structural terms.

[4]Robins, Snijders, Wang, Handcock and Pattison (2007) examine 20 highly different but well-known datasets available in UCINET. The number of nodes in these networks range from 10 to 39.

Second, this paper highlights the importance of both the number of links and the structural mechanisms in determining economic link formation. The only previous empirical studies that explicitly recognize the role of network architecture in rural villages are Krishnan and Sciubba (2009) and Comola (2010). Krishnan and Sciubba (2009) modify the model by Jackson and Wolinsky (1996) in respect to network heterogeneity and empirically investigate the predicted equilibrium outcome using data on labor-sharing arrangements in Ethiopia. They show how differences in quality and endowments of participants are key determinants of the structure of the network.[5] Also departing from Jackson and Wolinsky's connections model, Comola (2010) proposes a structural model of endogenous network formation allowing for indirect contacts to generate externalities. Testing a single rural village in Tanzania, she finds that agents also consider the wealth and the position of indirect contacts when evaluating the net advantage of forming a link. Taken together, these studies show that the network structure has an explanatory value that is ignored in reduced form estimations. While both of these papers take a game-theoretical approach and test specific equilibrium predictions derived from endogenous network formation theory, this paper takes as its point of departure the random graph literature inspired by physicists and identifies the determinants of network formation by modeling a probability distribution of the entire network. While the ERG approach allows the researcher to account for the number of links of indirect contacts, in contrast to the previous studies, it fails to consider externalities generated by indirect contacts.[6]

The empirical findings suggest that the assumption of dyadic independence and previous studies' inability to account for structural mechanisms lead to upward biased parameter estimates of household attributes. Moreover, I find statistical evidence that the economic exchange network in rural Gambia cannot be described by individual and dyadic attributes alone, but that higher-order structural mechanisms also determine network formation. This implies that a farmer's propensity to send and receive depends on the giving behavior of others and not only on household attributes. In particular, and in line with previous research examining friendship networks, I demonstrate that reci-

[5]In particular, Krishnan and Sciubba (2009) find that symmetric networks where all (or almost all) links are established between farmers who tend to have the same number of partners are more likely to emerge among households that do not differ in quality. Heterogeneity in quality is more likely to be observed among participants in asymmetric network structures characterized by few central households.

[6]In a recent working paper by Chandrasekhar and Jackson (2013) additional caveats of the ERG models are pointed out.

procity and transitivity are of overwhelming importance for the formation of economic exchange networks, and that they specifically reduce the effects of kinship ties, established in the literature as one of the main drivers of mutual support arrangements (i.e. Weerdt, 2004; Comola, 2010; Jaimovich, 2011). The importance of direct reciprocity is in line with previous studies using the same data (Jaimovich, 2013), but this paper further shows the sensitivity of the importance of direct reciprocity to the exclusion of higher-order structural mechanisms such as transitivity.

From a policy point of view, important insights about the structure of African economies can be gained from recognizing that people are embedded within social networks. Failure to incorporate this knowledge into decision making will lead to inappropriate economic policies, as interactions in one network spill over to other areas of economic activity (Dasgupta, 2002). By this reasoning, it is important to understand the driving forces of tie formation in order to design appropriate social policies at the micro level. The major lesson to be taken from this paper is that structural mechanisms, as compared to household observables, are more important in understanding network formation between farmers in rural villages.

The remainder of this paper is structured as follows: In the next section, I discuss possible channels of network formation related to the rural communities. In Section 3, I describe the survey data used for the empirical analysis. Section 4 presents the ERG models and discusses the challenges that arise in estimation. Section 5 presents the empirical results of the analysis, followed by a conclusion in Section 6.

# 2 Theoretical framework

Considerable uncertainty exists as to which tie-formation mechanisms influence the economic composition of networks. This uncertainty is further aggravated when the endogeneity of tie formation is taken into account. The function of this section is therefore to develop a theoretical framework that allows me to consider how various tie-generating mechanisms influence the overall network composition and how these mechanisms in turn are related to the sociodemographic structures of a rural population.[7] The framework starts from the basic mechanisms explaining friendship formation among high school students

[7]The generated network patterns then feed back into the sociodemographic structures. However, the cross-sectional dataset considered in this paper does not allow me to investigate network dynamics.

put forward by Wimmer and Lewis (2010) and focuses on four types of tie-generating mechanisms.

## Principles of tie formation in rural villages

First, the probability that two households establish a tie depends on the set of potential network partners and on the distribution of these households over the social categories and groups within this set. Most important for network composition is the size of potential network partners: The smaller the relative size of the village, the more likely it is that the villagers will form external ties. Similarly, the smaller the size of within-village groups, the greater the probability that households will establish out-of-group ties. This availability hypothesis can also be interpreted in terms of homophily: If households prefer to group with similar others, then the chance of out-of-group ties decreases with the relative size of these groups. The distribution over social categories is also important as people belonging (for example) to different social layers may be reluctant to establish a relationship: If household $i$ has a lower social status compared to household $j$, then $i$ may hesitate to form a tie (transact) with $j$, as the likelihood of reciprocity decreases with the level of social status.

Second, two households are likely to develop a tie if they regularly engage in joint activities (Feld, 1981). In other words, independent of group size, households can be brought into exchange relationships through the proximity mechanism. For example, if two individuals are participating in the same community activities over a long period of time, then they are likely to develop some kind of relationship with each other. These proximity effects can emerge through a wide range of factors, including shared institutional environments such as schools attended, common workplaces, kinship, spatial proximity, or voluntary organizations/groups (Kossinets and Watts, 2009). This proximity effect depends on the distribution of households over different groups and physical space as these often create boundaries for interactions and thus affect the probability of link formation. These groups or social categories in turn are indirectly structured by processes of selection and sorting. These may be historically determined. For instance, in the case of rural Gambia, the descendants of the first settlers have primary usage rights due to opposition to the sale of land to non-family members. This has created groups of landless and land-abundant households.

The third mechanism is homophily, which refer to an individual's preference to befriend similar others, where similar refers to those who share an attribute in a relevant category (McPherson, Smith-Lovin and Cook, 2001). Homophily may be preferred in uncertain environments for solidarity and social insurance, as homophily increases ease of communication, improves predictability of behavior, and fosters relationships of trust and reciprocity. In the setting of rural villages, homophily may arise from mutual ethnic categorization or same gender, educational level, and community position. The converse mechanism, namely heterophily, is also likely to be present in economic network formation. Individuals may voluntarily choose to establish economic relations with dissimilar individuals who possess skills, resources, and know-how that are complementary to their own and relevant for overcoming particular economic constraints.

Finally, these mechanisms of opportunity and homophily/heterophily should be distinguished from endogenous networking mechanisms that are only indirectly related to different sociodemographic structures. Several structural mechanisms can be identified. First, two individuals might become friends because they both like to socialize and are therefore able to develop a large number of ties with others. This suggests that tie formation also depends on an individual's degree of sociality, which can be thought of as the expansiveness (size) of the personal network (Goodreau, Kitts and Morris, 2009). Second, balance theory predicts that unreciprocated ties and aversion between one's friends produce social and psychological tensions (Cartwright and Harary, 1956), which means that networks are often characterized by high degree of *reciprocity* and *triadic closure*. Reciprocity refers to the tendency of $i$ to befriend $j$, if $j$ is already friends with $i$. Hence, in terms of rural economic networks, if household $i$ borrows inputs from household $j$, then $i$ is likely to return the favor whenever $j$ is in need, ceteris paribus. Similarly, transitivity refers to the tendency for friends of friends to become friends , implying network closure. Common to these structural mechanisms is that they are all balanced, which means that they all rely on the observation that humans value symmetry in relations.

While structural mechanisms create pressure for extended ties to be reciprocated and "open" triangles to be closed, independent of characteristics, there is still a possibility for economic network formation to result in asymmetric relations. For instance, some households may be more dependent on economic transfers than others. In addition, households endowed with more wealth and ability to generate income are expected to be in a better position to supply more input to their network partners. Moreover, better-endowed households are likely to be more appealing partners; however, they might also have a lower incentive to establish a relationship. If transactions of agricultural input are altruistic, transactions are likely to flow from more- towards less-endowed households. Dependent on the asymmetry in household endowment, this may create "hubs" representing households sending to many others and "spokes" representing households that receive from many others.

Until recently, most techniques used to model social networks assumed a large degree of independence among ties. Consequently, structural mechanisms were masked, and their potential contribution to observed homogeneity remained unclear. To avoid the upward bias to homophily in the presence of a tendency towards reciprocity, structural mechanisms need to be taken into account. This can also be illustrated by an example: In the presence of a tendency towards reciprocity and same-status homophily, individual $i$ may choose $j$ due to same-status effects, while $j$ may choose $i$ due to symmetry considerations. Similar, in terms of transitivity closure, $i$ may choose to befriend $j$ because they share a common partner, $k$, and not because of same-status preferences (Moody, 2001). This implies that structural mechanisms influence the observed degree of network homogeneity through an enhancing effect due to an individual's tendency to value symmetry in relations. In summary, an accurate estimate of either process requires controlling for the other using information about the attribute composition of all ties and the count of all triangles.

# 3 Data and setting

The network data studied in this article come from a baseline survey conducted between February and May 2009 for the purpose of evaluating a national Community Driven Development Project (CDDP). The data is representative of six out of eight Local Government Areas across different agro-ecological zones, and with populations between 200 and 1,000. The dataset contains three categories of information: (1) village-level information, (2) a standard household survey, and (3) detailed information on six networks: land, labor, input, credit, marriage, and kinship. The network dataset is unique in the sense that it includes information on transactions between all households residing in each village, as well as on household endowments and composition.[8]

To collect information on transactions between households, each household was asked to nominate the entire set of households with whom they transacted

[8]For detailed information on the sampling methodology and data description, see Jaimovich (2011).

land, labor, credit, and agricultural input within the last year. Information was collected on the direction of the transaction as well as the amount transacted.[9] The edges in the dataset, as illustrated in Figure 1, are therefore directed. Information on the level of the transactions is not exploited in the empirical analysis. I consider a network of economic transactions to be defined as those dyads in which households have transacted land, labor, or agricultural input.[10] Hence, if a transaction is observed from household $j$ to household $i$, then it can be any of the three mediums (i.e., labor, agricultural input, or land).

The driving factor of the labor network is the scarcity of household labor, in particular before and after the rainy season. Accordingly, fewer than half of the household members on average are regarded as working members. Each working member has the official right to cultivate the equivalent of 0.4 hectares of land. Additional allocation of land is determined by the village founders, the village cheif also called the *Alkalo* and his direct relatives. This means that households not entitled with a sufficient amount of land must borrow plots on a seasonal or annual basis. Most often, transactions in land networks are moneyless and are directly assigned by the village privileged, but they sometimes include rent contracts (Arcand and Jaimovich, 2010). Finally, the input network is defined as exchanges of production units such as tools, cattle, fertilizer, and seeds.

The level of analysis is the household. Rural villages in The Gambia are organized into compounds (*kundas*), which correspond to a group of people who work jointly on common fields, eat together, and organize daily activities (von Braun and Webb, 1989; Pamela, 2010). Depending on the size of the compound, independent cooking and consumption units (*dababas*) can co-exist within the compound. The household (i.e., the individual *dababas*) is used as the unit of analysis. If several households exist within one compound, the network between these will be present in the dataset. Some 16.6 percent of the household heads in the sample are not the head of the compound in which they live.

While the dataset includes information about the connections between households in the village and households outside the village, it does not contain household-level information on external partners residing outside the village.

[9]Information on the input network only includes whether a transaction took place, and thus does not include the level of the transaction. Moreover, the labor network was only collected for the household head.

[10]In comparison, Jaimovich (2011) uses a broader definition of the economic network, including labor, land, input, and credit. I choose not to include credit because credit networks are somewhat different from the other networks directly related to agricultural production.

It is therefore necessary to assume that the village is the natural domain for economic transactions related to agricultural input. This assumption is supported by the fact that external actors are not very important in any of the three networks. 7 percent of the households in the sample send land to external villagers and 9 percent receive land from non-village members (no households both receive and send land in the external village market). The relatively low level of land transactions involving households residing outside the village is likely to be explained by the immobility of land. For the labor market, no households work on non-village members farms, whereas 7 percent of the households receive external labor. The result on the sender side of the labor network is hardly surprising as the labor network is restricted to labor provided by the household head. It should be noted that the labor network does not include seasonal migrant workers, known locally as "strange farmers" (Swindell, 1987). Finally, around 3 and 9 percent of the households in our sample send and receive agricultural input, respectively, from non-village members.

Each village network may be represented by an asymmetric $n \times n$ matrix $\mathbf{Y}$ and $n \times q$ matrix $\mathbf{X}$ of household attributes, where $n$ is the number of households. The entries of the $\mathbf{Y}$ matrix, called the adjacency matrix, all take the value 0 or 1, with $Y_{ij} = 1$ indicating the presence of a transaction from $j$ to $i$, and vice versa. Since there is no limit on the number of allowed transactions reported, this means that the data is complete, and the directness of the data means that no assumption of over- or underreporting needs to be imposed.[11] The total sample used in the empirical analysis includes 37 villages. It was not possible to fit a general ERG model due to poor convergence in the remaining 15 villages in the dataset.[12]

The household attributes included in $\mathbf{X}$ are chosen as in the shorter run they are considered exogenous. It is important that these variables be exogenous in order to guarantee the property of dyadic independence to be explained in Section 4. The analysis is focused on five household attributes. Three of these are related to the household as a whole and include ethnicity, household

[11]Most studies examining school friends allow students to nominate five friends and impose the assumption of symmetry ($Y_{ij} = Y_{ji}$). For instance, in the case where $A$ nominates $B$ as his friend, but $B$ does not nominate $A$, the researcher imposes the assumption of underreporting and establishes an undirected friendship link between $A$ and $B$. An exception to this rule is Hunter et al. (2008), where if either student fails to nominate the other, there is no mutual friendship.

[12]The original sample included 60 villages. Out of these, eight villages were initially excluded from the analysis: (i) Five villages were initially excluded due to a large number of missing household attributes, (ii) and three villages were excluded due to their semi-urban nature not being comparable with the rural setting examined.

size, and whether the household is related to the *Alkalo*. The additional two attributes are defined by the head of the household: age and illiteracy status. At first, relatedness to the *Alkalo* may seem endogenous. However, in the vast majority of the villages considered, the position of village chief is inherited passed on from the first settlers in the village. Descendants of the first settlers are households with the primary usage rights to the land, and thus family ties to the *Alkalo* proxy for households' social and political status in the village.

Summary statistics across all households in each of the 37 villages are reported in Table 1. The 37 villages correspond to a sample of 1,570 households. The largest households have more than 50 members, and the average household has 13 members. The large households are partly explained by the polygamous nature of rural Gambian society (48 percent of the household heads have more than one wife). Households on average have five adult working members, and almost half the household heads are illiterate. For the majority of households, the main economic activity is related to agriculture (80 percent), though many households also engage in other income-generating activities. The kinship network is dense, and thus almost all households have at least one kin within the village. Ethnic heterogeneity is vast, with no ethnic group accounting for more than 53 percent of the sample. Table 1 also provide t-tests of the difference between the sample of villages included in the empirical analysis and the sample of villages excluded. Heterogeneity across the samples is generally caused by ethnic heterogeneity as well as differences in the density of kinship ties disaggregated into ties of the household head, kin of the wife (wives) of the household head, and marriage ties.

## Village 10

While average statistics of all the models estimated within each of the 37 villages are reported in the empirical analysis, I choose to focus on a single illustrative village, Village 10, which comprises 57 households corresponding to 757 inhabitants.[13] The results from Village 10 may not necessarily be generalized to the whole population of villages. In particular, the parameter estimates for Village 10 may be numerically quite different for other villages because the parameters depend on the number of nodes (i.e., households) in a complicated way, as discussed in Section 2. However, when I consider all 37 villages, I find

[13]Village 10 was not selection based on any pre-specified rule. It was however the median village in the initial sample of 20 villages considered. Estimation results for three additional villages are shown in the Appendix, Table A2-A4.

Table 1: Household descriptives

| | (1) Villages in ERG sample | | (2) Village 10 | | (3) Excluded villages | | T-test | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. | (1)-(2) | (1)-(3) | (2)-(3) |
| Total income per capita (1,000 GMD) | 3.079 | 3.590 | 3.150 | 1.461 | 2.764 | 2.417 | | * | |
| Agricultural share (share of income) | 0.139 | 0.245 | 0.057 | 0.141 | 0.158 | 0.272 | * | | * |
| Receives remittances | 0.461 | 0.499 | 0.351 | 0.481 | 0.413 | 0.493 | | * | |
| Land owned with official rights (ha.) | 10.098 | 24.962 | 3.000 | 3.868 | 8.656 | 14.186 | * | | * |
| Household size | 13.153 | 8.724 | 13.281 | 10.404 | 13.538 | 9.746 | | | |
| Number of working adults | 4.836 | 3.890 | 5.246 | 7.510 | 4.921 | 3.783 | | | |
| Age of head | 52.301 | 16.053 | 54.684 | 16.716 | 53.743 | 16.404 | | | |
| Agricultural work | 0.803 | 0.398 | 0.737 | 0.444 | 0.755 | 0.431 | | * | |
| Non-agricultural work | 0.164 | 0.371 | 0.211 | 0.411 | 0.207 | 0.405 | | * | |
| HH has family links in the village | 0.955 | 0.208 | 0.965 | 0.186 | 0.970 | 0.170 | | | |
| HH head has family links in the village | 0.864 | 0.343 | 0.895 | 0.310 | 0.825 | 0.381 | | * | |
| Wife of HH head has family links in the village | 0.499 | 0.500 | 0.386 | 0.491 | 0.572 | 0.495 | | * | * |
| HH has marriage links in the village | 0.618 | 0.486 | 0.351 | 0.481 | 0.672 | 0.470 | * | * | * |
| Female-headed household | 0.047 | 0.212 | 0.123 | 0.331 | 0.046 | 0.209 | * | | * |
| Illiterate | 0.446 | 0.497 | 0.632 | 0.487 | 0.452 | 0.498 | * | | * |
| Monogamous | 0.466 | 0.499 | 0.439 | 0.501 | 0.448 | 0.498 | | | |
| Polygamous | 0.483 | 0.500 | 0.421 | 0.498 | 0.497 | 0.500 | | | |
| Ethnicity: Mandinka | 0.532 | 0.499 | 0.018 | 0.132 | 0.557 | 0.497 | * | | * |
| Ethnicity: Fula | 0.161 | 0.368 | 0.105 | 0.31 | 0.242 | 0.428 | | * | * |
| Ethnicity: Wollof | 0.108 | 0.311 | 0.000 | 0.000 | 0.065 | 0.246 | * | * | * |
| Ethnicity: Jola | 0.076 | 0.265 | 0.000 | 0.000 | 0.084 | 0.278 | * | | * |
| Ethnicity: Serehuleh | 0.020 | 0.141 | 0.035 | 0.186 | 0.030 | 0.170 | | | |
| Ethnicity: Sererr | 0.073 | 0.260 | 0.544 | 0.503 | 0.021 | 0.144 | * | * | * |
| Ethnicity: Manjago | 0.011 | 0.106 | 0.088 | 0.285 | 0.000 | 0.000 | * | * | * |
| Ethnicity: Non-Gambian | 0.011 | 0.106 | 0.140 | 0.350 | 0.002 | 0.042 | * | * | * |
| Household observations | 1,570 | | 57 | | 571 | | | | |
| Villages | 37 | | 1 | | 16 | | | | |

Note: The t-test across columns (1) and (2) does not include Village 10 in the sample shown in column (1). * denotes statistical differences at least at the 5 percent level.

similar qualitative results. According to column (2) in Table 1, the economic condition in Village 10 resembles the larger set of Gambian villages surveyed with the exception of the amount of land owned with official rights and the ethnic composition of the village. While the ethnic Mandinka group is the largest ethnic group in The Gambia, they are among the ethnic minorities in Village 10.

The number of transactions in Village 10 corresponds to the degree. Since we have a balanced adjacency matrix, the number of receiving links is equal to the number of sending links considering the overall network. However, the in- and out-degree for an individual household $i$ can be unbalanced, where the in-degree is the sum of $i$'s receiver links and the out-degree is the sum of $i$'s sending links. In order words, $i$ can have more (or less) receiving links than sending links. The total number of transaction (i.e., edges) in Village 10 is 96, and the total number of triangles is 19. Some 5 households do not participate on either side of the market. There is a tendency for sending households to send to more than one household, while fewer households receive from many households. However, there is one interesting exception to this pattern: One household receives input from 13 households (this can also be seen visually in Figure 1 in the introduction). In comparison, the maximum number of out-degree ties is 6. The total number of possible ties in the village is $n \times (n-1)$, where $n$ is equal to the number of households. Out of the 3,192 possible ties, the density of the kinship network amounts to 8 percent. A visualization of the economic network as well as the kinship network is shown in Figure 1 and A1.

# 4 Estimation method: ERG modeling

ERG models allow the researcher to consider in-group preferences on all levels at the same time. This method also enables one to take the effects of relative group size, proximity, sociality, and balancing mechanisms into consideration simultaneously. Hence, the aim when generating an ERG model is to find the set of parameters that maximizes the probability that any random graph generated is identical to the observed network. To ensure that the reader is familiar with the technique applied, a general introduction follows below.

## Statistical framework

ERG models have been developed over the last 20 years as a method of directly modeling the underlying forces which create social networks. Essentially, this class of models works as a pattern recognition device, looking for consistencies in the way social ties are structured, as well as the associations between ties and individual attributes. In ERG modeling, the possible ties among actors in a network are regarded as random variables, and the general form of the model is determined by assumptions about the dependencies among these variables. This paper considers social selection ERG models, which assume that the pattern of ties is explained by the relative prevalence of a range of overlapping "sub-graphs," usually called "configurations". A configuration can be a simple tie between two nodes (i.e., an edge) or a more complex three-tie configuration (a "triangle").

The ERG models can be expressed mathematically in the following form:

$$Prob\left(\mathbf{Y}=\mathbf{y}|\mathbf{X}\right)=\left(\frac{1}{\kappa}\right)exp\left[\sum_{A}\eta_{A}g_{A}\left(\mathbf{y},\mathbf{X}\right)\right] \tag{1}$$

where $\mathbf{Y}$ is the matrix of $Y_{ij}$ corresponding to a network tie between two members $i$ and $j$ of a set of $N$ actors. $Y_{ij}=1$ if there is a link between the actors, and 0 otherwise. $\mathbf{y}$ is the matrix of observed ties, where $y_{ij}$ correspond to the realized value of $Y_{ij}$, and $\mathbf{X}$ is a matrix of attributes. $\kappa$ is a normalizing constant which ensures that the equation has a proper probability distribution, while summation in the model is taken over all $A$ configurations. Finally, $g_A\left(\mathbf{y}\right)$ is the network statistic corresponding to configuration $A$, and each statistic is associated with a parameter, $\eta_A$, indicating the importance of the configuration. A high (and positive) parameter value implies that the configuration is expressed more often than expected by chance (or at least more often than if the ties were formed at random). In contrast, configurations with a negative parameter value have a less-than-chance probability of being present, controlling for all other configurations in the model.[14]

The simplest model of interest is a single-parameter model that posits an equal probability for all edges in the network. This model is known as

[14]Equation 1 can be re-expressed as the conditional log-odds (logit) of individual ties: $log(P(Y_{ij}=1|N,Y_{ij,-ij}))=\sum_{a=1}^{A}\eta_A\delta g_A(\mathbf{y})$, where $Y_{ij,-ij}$ denotes all dyads other than $Y_{ij}$ and represents the dependence between ties. The interpretation of $\eta_A$ is then as follows: If forming a tie increases the configuration $g_A$ by one, then ceteris paribus, the log-odds of that tie forming increases by $\eta_A$. For example, if $A$ refers to a triangle, then $\eta_A$ represents the increase in log-odds of a tie forming that would close exactly one triangle, assuming nothing else changes and all other model effects have been accounted for (Hunter, 2007).

the Bernoulli model and assumes *dyadic independence*. In the ERG modeling framework, this corresponds to a model with a $g_A(\mathbf{y},\mathbf{X})$ vector of statistics that contains only a single element, namely the number of edges in the network. Under dyadic independence the model can be further extended to include a matrix of observable attributes, $\mathbf{X}$. In such models the probability of any tie does not depend on the value of other ties, but only on the attributes of the two actors involved. This model is equivalent to the dyadic regression model.[15] By contrast, a model including endogenous tie formation exhibits *dyadic dependence*. Examples include the Markov dependence model introduced by Frank and Strauss (1986) and the triad closure model described in this paper.

To estimate ERG models under dyadic dependence, assumptions on the nature of the configurations ($\eta_A$) are needed. The Markov dependence assumption introduced by Frank and Strauss (1986) states that two possible edges in a graph are only conditionally dependent when a common actor is involved in both. Markov random graph models, however, often lead to difficulties as the algorithms for parameter estimation may not converge when triangulation is high or there are high-degree nodes in the degree distribution (i.e., "hubs" or "spokes"). Recent developments have therefore implemented the Markov chain Monte Carlo (MCMC) maximum likelihood estimation procedure (Snijders, 2002; Handcock, 2003). The Monte Carlo estimation simulates a distribution of random graphs based on a starting set of parameter values generated by pseudo-likelihood. These parameters are repeatedly refined by comparing the simulated distribution of graphs against the observed data. Simulations from the convergent estimates will then produce a distribution of graphs in which the observed graph is typical for all effects in the model. Despite these recent developments, estimates gained through this procedure are often empirically implausible, such that the parameter space of the model contains either almost complete graphs or almost empty graphs.[16] This problem is known as *degeneracy* and occurs when a model is poorly specified. In practice, degeneracy often implies that parameter estimates never converge (Handcock, 2003). I assess the model fit in Section 5 and find that the preferred (full) model is well specified.

[15] It should be noted that ERG models that exhibit dyadic independence may also include other terms such as reciprocity, homophily, and differences in actor attributes.

[16] The parameter space refers to the range of possible parameter values for all of the configurations in the model. A complete graph is a graph where all possible ties exist, while an empty graph is a graph with no ties.

### New specifications

In a promising attempt to overcome problems of degeneracy, Snijders, Pattison, Robins and Handcook (2006) proposed three new specifications: higher-order star, triangular, and two-path effects. Following the partial conditional dependence assumption by Pattison and Robins (2002),[17] Snijders et al. (2006) assume that the presence of some ties affects the propensity for closed structures to emerge in the network. Combined with Markov dependence, this assumption enables the examination of properties for multiple triangulation and connectivity. These new specifications are at the core of this paper, and I therefore briefly present them in turn.

First, the new statistics for triangular configurations accommodate the tendency that two nodes share more than one partner, allowing for densely clustered areas in the network. In terms of a friendship network, the underlying effect of higher-order triangles might be that $i$ tends to choose actor $j$, who is a friend of their mutual friend $k$. The strength of this choice increases the number of friends shared by $j$ and $i$.[18] This complex triangular configuration is termed a geometrically weighted edge-wise shared partner (GWESP) statistic. The geometrical weight expresses the expectation that higher-order triangles, where two nodes share many partners, are less likely than lower-order triangles, where two nodes share fewer partners.

Second, basic stars are replaced by more complex star configurations (geometrically weighted degree), which separate the probability of observing stars of all possible orders into two discrete terms: an in-degree (GWID) effect for in-stars of all possible orders and an out-degree (GWOD) effect for out-stars of all possible orders. The corresponding in-degree parameter relates to *popularity effects* and the out-degree parameter to *activity effects*.[19] A positive in-degree parameter suggests a preference for connections with higher-degree nodes, leading to a core–periphery structure, but with a core of limited size.

Finally, Snijders et al. (2006) introduced configurations related to the dis-

[17] Partial conditional dependence means that whether two pairs (edges) $Y_{ij}$ and $Y_{kl}$ are conditionally dependent, given the rest of the graph, depends not only on whether they share nodes but also on the pattern of ties in the rest of the graph. In practice, the main effect of the partial dependence assumption is that it allows additional configurations to effect the formation of ties.

[18] This increase is not linear, and beyond a certain number, finding an additional friend $k$ does not add greatly to the chances of the multiple two-path becoming closed (Hunter, 2007).

[19] The terms were defined for friendship networks. For the economic network considered in this paper, *popularity effects* referring to households that receive inputs from many households can seem a bit misleading when resources in these villages are generally scarce.

tribution of shared partners of actors who are not tied themselves: the geometrically weighted dyad-wise shared partner (GWDSP). This new parameter is equivalent to a triangle without a base. This parameter is comparable to what Burt (1992) identified as structural holes. Thus, a positive parameter may be interpreted as a structural imbalance, representing a situation where actor $i$ is not connected with $j$ despite having one or more partners in common (Wimmer and Lewis, 2010).

## ERG model specification

The model to be estimated includes terms of sociality, homophily, and structural mechanisms in the form of reciprocity, triad closure, and star effects. The vast data covering many rural Gambian villages provide an unusual opportunity for empirical replication. The models are fit to each village, which in turn allows us to compare 37 sets of coefficients.

I account for *sociability* based on the counts of ties observed. Three statistics are included: First, the total number of ties in a given village acts as an intercept. Since the magnitude of the intercept is directly affected by network density (fraction of possible network links realized), I expect the density of economic transactions between households to decrease with village population. Second, the total number of ties for all households with attributes $i$ represents the tie probability relative to the reference category. The reference category is household not kin-related to the *Alkalo* and not illiterate. By including sociability for these groups, we allow for heterogeneity across these attribute classes and homogeneity within the individual attribute class. The third is the number of ties for sending and receiving households with attributes $k$. These attributes are continuous in nature and include the household size and age. This count measure determines whether the number of sending and receiving ties is correlated with the size and age of the receiving and sending household.

I account for *homophily* in educational and social status, as well as ethnicity. In order for the model to be appropriate across a large range of villages, I consider homogeneous homophily across the attribute categories (i.e., *uniform homophily*). To illustrate the possible presence of *differential homophily* across specific individual ethnic groups, the model is re-estimated for Village 10 including homophily terms corresponding to the ethnic groups in the village.

Kinship is often found to be a key driver of network formation between households (De Weerdt, 2002; Fafchamps and Lund, 2003; Udry and Conley, 2004; Dercon and Weerdt, 2006). Typically, a household has a long-lasting relationship with family members, and the family as a group is likely to punish uncooperative behavior, thus inducing norms and trust. To accommodate the importance of kinship ties, a dyad-specific variable is included that takes the value 1 if $i$ and $j$ are kin-related to the household head, the wife (wives) of the household head, or by a marriage tie, and 0 otherwise.

Finally, for reasons discussed above, I consider the importance of different structural mechanisms. First, I account for reciprocity by including a dyad-specific variable measuring the probability that a transaction is reversed through a transaction in the opposite direction. Second, I investigate triad closure using GWESP to capture the phenomenon wherein two nodes tend to share at least one partner, producing densely clustered areas. Third, I investigate the presence of popularity effects where some households receive more transactions compared to other households. For this I include the term GWID and thereby allow for all possible k-star configurations. Finally, I explore the possibility that some nodes receive transfers from the same set of senders without being connected (GWDSP). Higher-order out-degree stars were generally found to be unimportant, and thus GWOD is not included in the estimations.

I estimate five models for each village. The first model (Model 1), also referred to as the baseline model, only includes household attribute coefficients in terms of sociability and homophily. Sociability is included together with homophily as homophily terms only account for the different group size while ignoring differences in average sociability across groups. As discussed in Section 2 different sociability across groups may influence the extent of homophily. In practice, inclusion of sociability has been shown to modify homophily effects (Wimmer and Lewis, 2010; Goodreau, Kitts and Morris, 2009). Model 2 extends Model 1 by including the dyad-specific indicator for whether two households are kin related. Many studies examining local network formation in rural villages are unable to account for kin ties due to data limitations (for instance Fafchamps and Gubert, 2007; Krishnan and Sciubba, 2009). Comparison of Models 1 and 2 illustrates the importance of kinship networks in rural communities and how exclusion may result in misleading conclusions regarding the network-formation process. Model 3 exclusively models the structural variables together with an edge parameter to detect the importance of structural mechanisms. To determine whether the observed tendency to transact with similar households is modified by the balancing mechanism of reciprocation, Model 4 includes a reciprocity term. To further understand the importance of structural mechanisms, Model 5 extends Model 4 by including triadic closure (GWESP), two-path (GWDSP), and in-star effects (GWID). I refer to Model 5 in the rest of the paper as the full model.[20]

# 5 Results

Subsection 5 focuses on changes in the importance of determinants when structural mechanisms are included under the assumption of dyad independence. First, the estimation results across the five models are discussed for Village 10. Second, I compare the results across the 37 villages. Having established some general findings across villages, Subsection 5 builds the best possible ERG model of the observed network structure in Village 10, followed by an assessment of model fit.

## Village 10

Parameter estimates of the five ERG models for Village 10 are reported in Table 2. Model 1, reported in column (1), shows how network ties are formed with regard to household attributes. Surprisingly, only a few household attributes are found to be statistically significant under the assumption of dyad independence. The negative and significant intercept indicates that the degree distribution is lower than what would be expected if ties were formed at random. The log-odds of the formation of a tie that is completely heterogeneous (the number of members differs from each other on all attributes) is equal to the intercept -3.02 (coefficient of the edges).

In the next step, the dyadic indicator of kinship is included (Model 2 shown in column (2)). The effect is positive and statistically significant, suggesting that a household is more likely to transact with kin. The large impact from kinship ties is consistent with previous studies that examine network formation in rural villages (i.e. Comola, 2010; Jaimovich, 2011). Comparing Models 1 and 2, the coefficient of the sender effect of being related to the *Alkalo* and the coefficient of the receiver effect of household size become insignificant. It

[20] All ERG models and goodness-of-fit plots in this article were generated using ergm package, which is part of the statnet packages of statistical network analysis in R (Handcock et al., 2003). Models 1, 2, and 4 assume dyadic independence and thus can be calculated using pseudo-likelihood estimation. Models 3 and 5, however, require MCMC estimation due to the incorporation of higher-order terms. Variability in the MCMC estimations was reduced by implementing long Markov chains, selecting a burn-in of 1 million toggles, an MCMC sample size of 10,000, and an interval between successive samples of 10,000 toggles. Finally, the step length was set to 0.25 for further stability.

is further evident that the sign of the receiver coefficient of being related to the *Alkalo* changes: Households related to the *Alkalo* are less likely to receive economic input. Inclusion of kinship ties amplifies the coefficient estimate of ethnic homophily, and it becomes significant. This may be driven by (i) a high correlation between kin and same ethnicity, or (ii) the fact that we have not controlled for ethnic sociability. Table A1 controls for ethnic sociability and separates out the homophily effect to allow for heterogeneous ethnic homophily (i.e., differential homophily). Correcting for sociability does not change the result for ethnic homophily; however, the result disappears once differential ethnic homophily is included together with ethnic sociability.[21]

The estimation of Model 3, reported in column (3), shows the importance of structural mechanisms in Village 10. The positive and significant effect of reciprocity means that economic assistance is likely to be reciprocated. Taken together with the negative and significant edge parameter, this indicates that households have few other economic network partners apart from the ones that are reciprocated. Closing one or more triangles is also a relatively important structural mechanism (GWESP), though the magnitude is considerably lower than that for reciprocity. Controlling for reciprocity and the tendency to close triangles, there are fewer unclosed triangles than would be expected to form by chance (negative coefficient of GWDSP). Given the discussion in Section 2, this result is less surprising since unclosed paths are structurally unbalanced. The positive effect of the higher-order k-trinagles compelled with the negative two-parts effect suggests that the exchange network tends to be cliguelike, with possibly several different denser clusters of farmers (Snijders, Pattison, Robins and Handcook, 2006). Hence, the evidence does not support the presence of structural holes. Finally, the negative and significant coefficient of in-stars (GWID) indicates that high in-degrees are less likely in this network and thus farmers tend to be high-degree nodes within clusters of farmers they transact with rather than between clusters.

Model 4 introduces reciprocity into Model 2 in order to determine the impact of the balancing mechanism on household attributes. As in Model 3, the coefficient of reciprocity is statistically significant and positive. Separating out the balancing effect of reciprocity from kinship and ethnic homophily mecha-

[21] Table A1 also shows in column (3) that excluding ethnic sociability but allowing for ethnic heterogeneity indicates that uniform ethnic homophily is solely driven by the ethnic Fula group. However, the effect disappears when we control for sociability in column (3). It should be noted that the Akaike Information Criterion (AIC) and likelihood-ratio test suggest that the model estimated in column (3) is marginally better than the model estimated in column (4).

nisms, the coefficients decrease by at least 28 percent. In fact, comparing the size of the coefficients, it is seen that a transaction that symmetrizes a dyad is statistically more likely to occur than a transaction between two households that share the same ethnicity. More specifically, the edge coefficient of -3.67 in Model 4 refers to the log-odds of a tie forming between two households assigned to different household attributes and that do not reciprocate. This log-odd increases by 2.55 if the tie establishes a mutual transaction relationship, whereas only by 0.68 and 1.34 if the tie is between same-ethnicity or kin-related households, respectively.

The estimation results of the full model (Model 5) are reported in column (5). Comparison across models shows that significance and signs are similar to Model 3 and Model 4. Indeed, the most important principle of networking overall is the tendency to reciprocate transactions of input for agricultural production. This is in contrast to Weerdt (2002) and Comola (2010), who both find that the variable with the highest impact is kinship ties. The fall in the coefficient estimate of kinship ties is due to the fact that reciprocated exchanges are within the extended family. Comparing coefficient estimates of household attributes between Model 2 and Model 5, it is evident that the inclusion of structural mechanisms reduces the effect of household attributes, with the exception of the receiver effect for households related to the *Alkalo*.

### Comparison of the full model across all villages

I first consider which of the five models best fits the data by using the likelihood-based measures of Akaike Information Criterion (AIC) and likelihood-ratio tests. For all 37 villages, these measures indicate that the full model fits best but with a varying margin. The increase in log-likelihood between Model 1 and the full model (Model 5) ranges between 20 and 300, with six new terms in the full model relative to Model 1. The observed log-likelihoods for the additional models are all between the baseline model and the full model, with exception of 7 villages. Since the full model fits best for the majority of villages, I consider these coefficient estimates to be the best estimates of the true magnitude of sociability, homophily, and structural mechanisms.

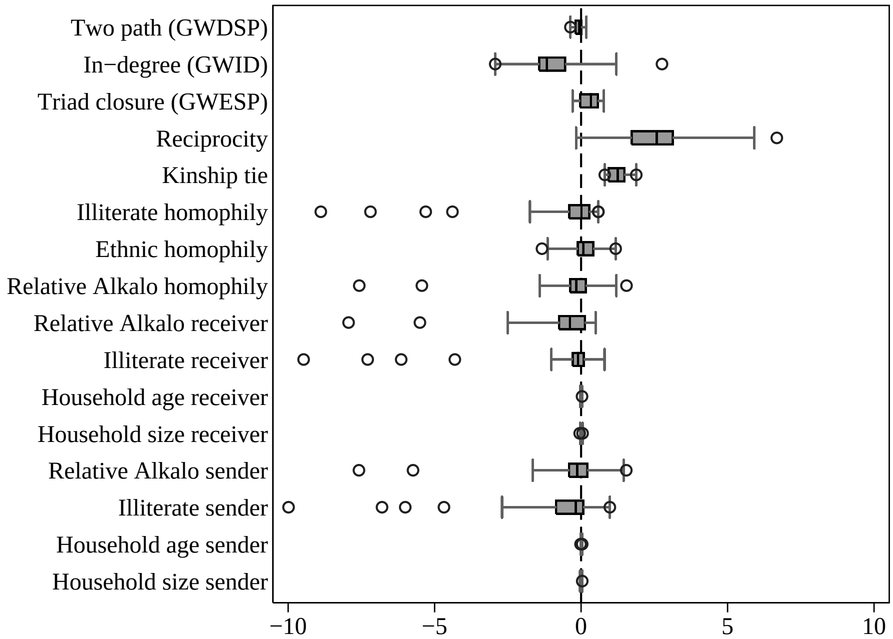
Figure 2 charts all the coefficient estimates from the full model using boxplots in order to show the range across the 37 villages. The household attributes should be interpreted with caution as significance varies across villages. The chart shows a general tendency for negative coefficient estimates of household attributes. Sociability effects are generally largest for households related to

Table 2: Results for Village 10

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| Edges | -3.673*** | -3.020*** | -3.106*** | -3.367*** | -2.587*** |
| | (0.628) | (0.659) | (0.269) | (0.616) | (0.650) |
| Household size sender | -0.019 | -0.005 | | -0.004 | -0.005 |
| | (0.015) | (0.012) | | (0.013) | (0.011) |
| Household age sender | 0.003 | -0.004 | | -0.004 | -0.003 |
| | (0.006) | (0.006) | | (0.007) | (0.007) |
| Illiterate sender | -0.116 | 0.289 | | 0.297 | 0.265 |
| | (0.228) | (0.251) | | (0.249) | (0.254) |
| Relative Alkalo sender | 0.534** | -0.189 | | -0.014 | -0.139 |
| | (0.229) | (0.247) | | (0.256) | (0.256) |
| Household size receiver | -0.034** | -0.001 | | 0.000 | 0.000 |
| | (0.016) | (0.010) | | (0.011) | (0.009) |
| Household age receiver | 0.006 | -0.002 | | -0.002 | -0.002 |
| | (0.006) | (0.006) | | (0.007) | (0.005) |
| Illiterate receiver | -0.265 | 0.081 | | 0.003 | 0.087 |
| | (0.229) | (0.251) | | (0.249) | (0.214) |
| Relative Alkalo receiver | 0.462** | -0.674*** | | -0.674*** | -0.493** |
| | (0.230) | (0.249) | | (0.256) | (0.215) |
| Relative Alkalo homophily | 0.076 | -0.488* | | -0.374* | -0.386* |
| | (0.221) | (0.252) | | (0.225) | (0.233) |
| Ethnic homophily | 0.092 | 0.868*** | | 0.677*** | 0.683*** |
| | (0.222) | (0.293) | | (0.253) | (0.253) |
| Illiterate homophily | 0.009 | -0.477* | | -0.382* | -0.397* |
| | (0.221) | (0.246) | | (0.223) | (0.238) |
| Kin ties | | 1.718*** | | 1.335*** | 1.295*** |
| | | (0.234) | | (0.206) | (0.216) |
| Reciprocity | | | 2.932*** | 2.548*** | 2.655*** |
| | | | (0.380) | (0.396) | (0.407) |
| Triad closure (GWESP) | | | 0.632*** | | 0.522*** |
| | | | (0.201) | | (0.199) |
| In degree (GWID) | | | -1.082*** | | -0.986** |
| | | | (0.418) | | (0.426) |
| Two paths (GWDSP) | | | -0.178** | | -0.210*** |
| | | | (0.078) | | (0.080) |
| AIC | 962.736 | 823.519 | 808.032 | 792.947 | 783.495 |
| BIC | 1,037.592 | 902.408 | 838.374 | 877.904 | 886.658 |
| Log Likelihood | -469.368 | -398.760 | -399.016 | -382.473 | -374.747 |

Note: Akaike Information Criterion (AIC), and Bayesian information criterion (BIC). ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Figure 2: Coefficients from the full model, plotted across all 37 villages

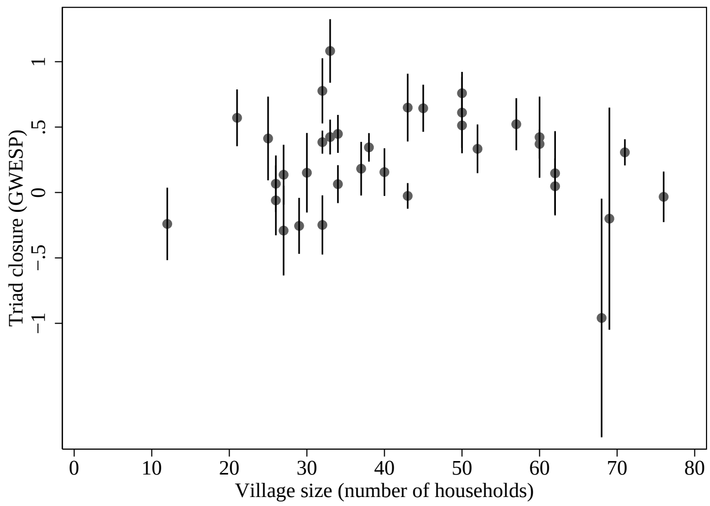



Note: Boxplots follow the Turkey method. Boxes represent quartiles, whiskers extend to the most extreme data point within 1.5 times the interquartile range from the edge of the box, and points represent outliers.

the *Alkalo*: These households exhibit slightly less solidarity than non-related households (negative mean on sender and receiver). The coefficient of kinship ties is positive and significant in all estimations across the 37 villages. This confirms the intuition that households are more likely to engage in agricultural input transactions with kin, possibly due to lower transaction costs and larger "costs" associated with refusing kin access to production input in times of need.

While the household attributes are largely insignificant and close to zero when the structural mechanisms are included in the full model, the coefficients associated with the balancing mechanisms are generally positive and statistically significant. In line with the findings for Village 10, the positive and significant estimate on reciprocity in all estimations (except in six villages) confirms that reciprocity is the most important principle of networking overall (Wimmer and Lewis, 2010): Households are more likely to engage in reciprocal relationships than those that would be formed at random. Moreover, the triad closure (GWESP) coefficient for the full model is positive in all estimates but only statistically significant in 20 out of 37 villages. To illustrate the more complicated interpretation of the GWESP coefficient, consider the average coefficient for the larger villages in terms of the total number of households (approximately 0.5 according to Figure 3). If a tie will close one triangle and no actor pairs in that triangle have a shared partner ex-ante to the transaction, then the log-odds will increase by 1.5 relative to an otherwise similar tie that would not close a triangle (Goodreau, Kitts and Morris, 2009). If the tie will close two triangles, the additional increase is approximately halved due to the positive geometrical weight (set to 0.3), which determines how quickly the influence of triangles levels off. Thus, completing a 2-triangle when a triangle already exists only result in an additional increase of $0.5 \times (1 - exp(-0.3)) = 0.26$ (for more details see Hunter, 2007). The relationship between network size and the GWESP coefficient, shown in Figure 3, results from the non-linear nature of triadic effects. Households in smaller villages average more exchange partners in common than those in larger villages. However, this is more likely to happen by chance in smaller villages: The probability of closing triangles by chance declines with the square of network size. The relationship between magnitude of effect and village size reflects the fact that higher-order relational processes operate differently in communities of different sizes. If households prefer some level of social closure, they must exert more effort in larger populations to create it.

Figure 3: Triad closure (GWESP) coefficient: full model



### Comparing results across models

As hypothesized, estimation results for Village 10 suggested that the coefficient estimates attenuate as structural mechanisms are included. This subsection compares coefficient estimates for all villages across the different models to establish whether this is a more general finding. Figure 4 shows plots for four selected statistics over all villages. Each dot in the plots corresponds to the coefficient estimate in one village, and the dotted line is the 45 degree line.

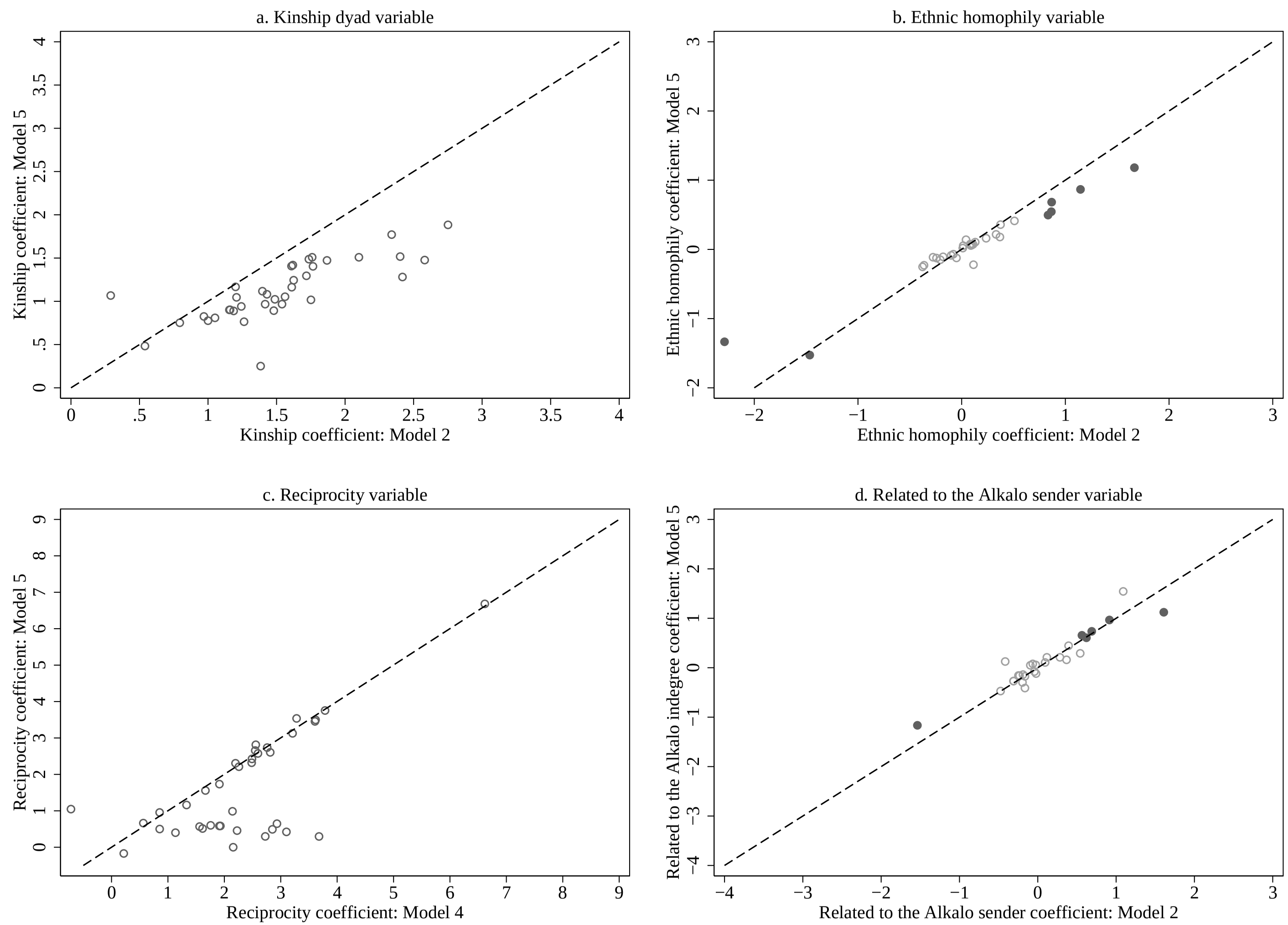
Analysis of the nested models on all villages shows that the kinship coefficient from the full model is less than the corresponding coefficient in Model 2 (Figure 4, plot a). The estimated coefficient for kinship on average declines by 32 percent, suggesting that assortive mixing by kinship in rural Gambian villages to a large extent is a function of structural mechanisms. Figure 4, plot b shows the apparent change in the ethnic homophily coefficient between Model 2 and the full model. At first it seems as if the coefficient estimates are similar across the two models. However, the coefficients (colored dots in plot b) that are statistically significant are mainly located to the left of the 45 degree line.

For reciprocity, shown in plot c, we find that additional structural mechanisms attenuate the effect associated with reciprocation in approximately half the villages. One reason that reciprocity is affected by the inclusion of triangle closure is that reciprocity is likely to be found in triangles. Hence, once the triad closure term is included, the magnitude of the reciprocity coefficient attenuates. Finally, plot d show no visual changes in the sender variables for being related to the *Alkalo*.

## An ERG model of network structure: Village 10

Having established some general findings across villages, I now derive the best possible ERG model of the observed network structure in Village 10. Due to the large heterogeneity of villages, it was necessary to consider more general models (Models 1-5) allowing for non-significant household attribute terms. Given that there is no generally accepted strategy for developing ERG models, I take an inductive strategy to find the model that best fits the general characteristics of Village 10. Unlike regression analysis, where an inductive approach is highly discouraged, the construction of realistic network models normally involves an extended trial-and-error process of simulation and refinement (Goodreau, 2007). However, for the specification of the final model, I developed a transparent and replicable procedure. I first ran separate models

Figure 4: Comparison of coefficient estimates across models



for each tie-formation mechanism and then combined all the significant sociability and homophily terms into a single model. These terms were then added to those structural terms that were found to be significant.[22] Alternatively, one could have used a backward selection procedure, where nonsignificant effects are stepwised deleted from the model (Snijders, Pattison, Robins and Handcook, 2006).

The estimation of Model 6 presented in Table 3 represents the best approximation of how the network structure itself was generated. Triadic closure, while important, does not represent the dominant principle of tie formation among rural households. However, reciprocity continues to be the most important mechanism of tie formation. Another important characteristic of the link between households that is often excluded in dyadic regressions due to data limitations is kinship ties between households. Kinship relations are found to be the second most important tie-formation mechanism. No additional household attributes were found to be important in Village 10.

## Assessing model fit

Having obtained coefficient estimates, we need to consider how well the models fit the observed data. I take three approaches to assess model fit for Village 10. I separately consider Models 5 and 6, and compare the model fit of Model 5 with that of Models 1 and 4.

First, I simulate a new network based on coefficient estimates obtained at random and compare these to the observed data in order to investigate how different they are. For the full model, 58 percent of the one million proposals were accepted.

Second, I examine the diagnostics for the MCMC model-fitting process. Figures A2-A4 show what happens to the model statistics during the last iteration of the MCMC estimation procedure.[23] The left-hand plots represent

[22]As this model includes dyad-dependence terms, the MCMC estimation procedure was applied. In line with the previous models, variability was reduced by implementing long Markov chains, selecting a burn-in of 1 million toggles, an MCMC sample size of 10,000, and an interval between successive samples of 10,000 toggles. Finally, the step length was set to 0.25 for further stability. The final parameters were obtained using 50 iterations of this process, each time using the finishing values of the previous cycle as a starting point for the next. This process was repeated twice with the same outcome. The final model (Model 6) is shown in Table 3. Using this approach, the remaining model has the most important determinants of tie formation in Village 10.

[23]A burn-in of 1,000 was applied. The burn-in is the number of steps in the simulation chain before the simulated network is drawn from the default of 1,000. When simulating a network the burn-in gives the chain a chance to move away from the starting network so that the output is approximately independent of initial conditions.

Table 3: Model of tie formation: Village 10

| | Model 6 |
|---|---|
| Edges | -3.721*** |
| | (0.195) |
| Relative *Alkalo* receiver | -0.208 |
| | (0.186) |
| Kinship | 1.245*** |
| | (0.204) |
| Reciprocity | 2.489*** |
| | (0.391) |
| Triad closure (GWESP) | 0.362* |
| | (0.186) |
| In degree (GWID) | -0.743* |
| | (0.415) |
| AIC | 779.943 |
| BIC | 816.353 |
| Log Likelihood | -383.971 |

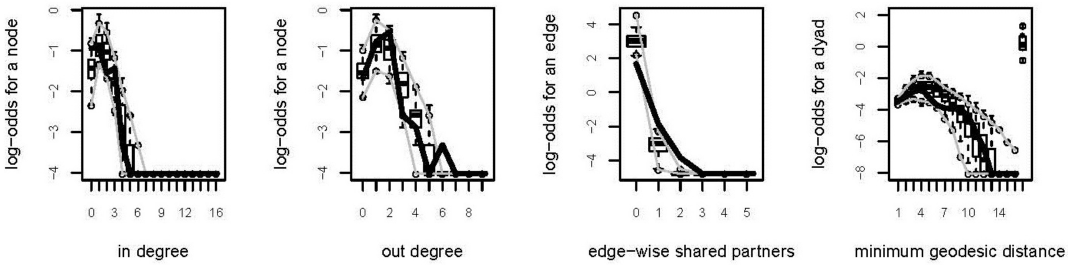
Note: ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

the chain as one time series for each model's statistics, while the right-hand side summarizes the chain in a histogram. Both are normalized to the observed data represented by 0. Simulation from the parameter estimates suggests a stable, well-behaved model with a single region of graphs. This can be seen from the figures, as there is no long tail to the distribution of, for example, edges, and across the simulation there is no "leakage" away from the bulk of the simulated graphs. This indicates that the model presented is not near degeneracy and the obtained estimates are a good representation of the data.

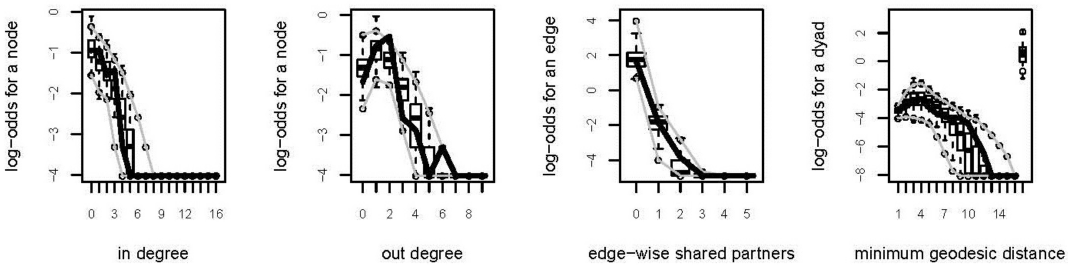
Third, it is desirable to offer a visual representation of how the processes included in the model are capable of accurately reproducing key features of the network's global structure. A set of 100 randomly generated networks was therefore simulated using the parameters from the three models. They were then compared to the actually observed network along four diagnostic parameters. Figure 5 charts the comparisons. The three rows correspond to Models 1, 4, and 5 presented in Table 2. In each case the dark solid line represents a given statistic from Village 10, while the boxplot represent the same statistics for the 100 simulated networks. The first plot in each row show the in-degree distribution, tabulated across all households in the network, while the second plot in each row shows the out-degree distribution. The model does not capture

Figure 5: Goodness-of-Fit: Village 10

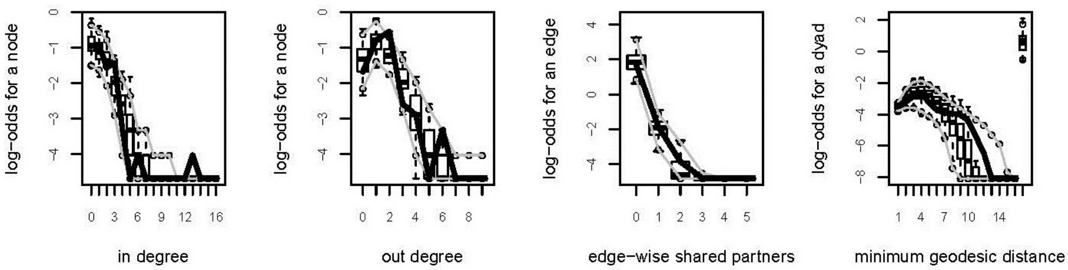
a. Model 1




b. Model 4




c. Model 5




well the large in-degree node observed in the data; while it does a better job on the out degree, it also face challenges in capturing the higher out-degree nodes. The third plot in each row represents the distribution of shared partners (i.e., number of exchange partners in common) tabulated across all pairs of households that transact. It thus provides a sense of the level and scale of clustering. Model 4 largely underestimates the number of shared partners, while Model 5 captures well the number of shared partners. On the other hand, none of the models comes close to matching the higher-order statistics. In fact, Model 5 sharply underestimates the observed higher-order statistics.

Turning to Model 6, some 64 percent of the one million proposals were accepted, corresponding to an increase of 5 percentage points compared to Model 5. Figure A5 compares the simulations to the actually observed network. The two plots in the first row show the in-degree and out-degree distribution.

Apart from a small jump in out-degree in the observed data not captured by the model, the statistics are remarkably improved compared to the statistics for Model 5. Moreover, shared partners and geographic distance are well captured in the model.

# 6 Conclusion

This paper investigated the consequences of dyad independence between pairs of households by allowing for endogenous networking mechanisms among households in 37 rural villages in The Gambia. I applied exponential random graph (ERG) modeling to investigate whether households choose their economic network partners solely on the basis of household attributes, or whether the potential network structure resulting from their additional partnerships is also an underlying consideration. Specifically, I test whether mechanisms in the form of reciprocity (tendency for households to return the transfer) and triadic closure (tendency for partners with whom a household transacts to start transacting among themselves) are important determinants of network formation. I proceeded by comparing coefficient estimates for all villages across different models to establish whether the exclusion of structural mechanisms attenuates the impact of standard household attributes. There are three main findings of this paper:

First, in line with previous studies, kinship ties are found to be an important determinant of network formation among rural households in The Gambia. However, the effects of kinship ties are significantly amplified when households' tendency to reciprocate transactions is not taken into account. Hence, part of the effect normally attributed to kinship ties is in fact driven by households' valuation of symmetry in relations.

Second, I find statistical evidence that the economic exchange network in rural Gambia cannot be described by household and dyadic attributes alone, but that higher-order structural mechanisms also determine network formation. In particular, and consistent with previous research examining friendship networks, I demonstrate that reciprocity and transitivity are important in the formation of exchange networks.

Third, I find that the inability to account for structural mechanisms leads to upward biased parameter estimates of household attributes and dyad-specific characteristics including kinship and reciprocity. This suggests that the households forming exchange networks take into account the structure of the community network and the structure resulting from additional partnerships. What these results show is that network architecture has an explanatory value which is disregarded when we focus exclusively on reduced form estimations allowing only for observable characteristics. The lesson from this paper is that household attributes, and kinship in particular, help explain network formation but are likely to mask underlying structural mechanisms, meaning that the actual impact from observables is less than what is often expected.

# Appendix

Figure A1: Kinship network: Village 10

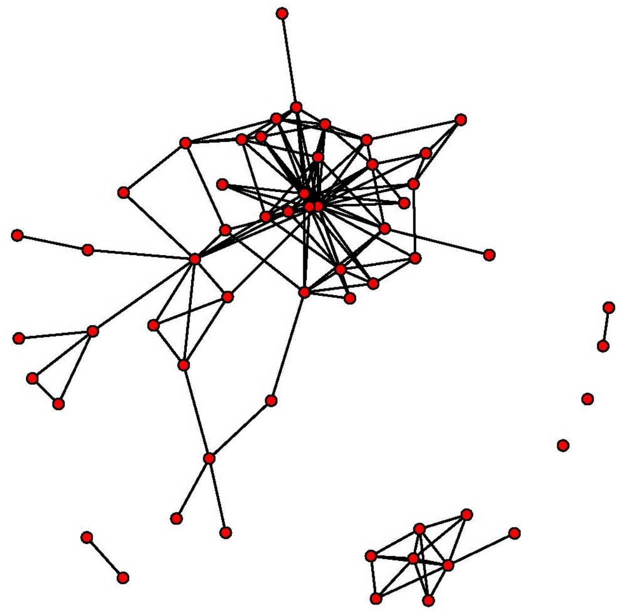


Table A1: Ethnic homophily

| | | (1) | (2) | (3) | (4) |
|---|---|---|---|---|---|
| | Edges | -3.020*** | -3.751** | -2.958*** | -3.827*** |
| | | (0.659) | (1.465) | (0.680) | (1.482) |
| *Iniform Ethnic homophily:* | | 0.868*** | 0.813** | | |
| | | (0.293) | (0.351) | | |
| *Differential Ethnic homophily:* | Fula | | | 1.027*** | 0.590 |
| | | | | (0.370) | (0.585) |
| | Sererr | | | 1.308 | 1.560 |
| | | | | (0.809) | (0.986) |
| | Manjago | | | -0.487 | -0.064 |
| | | | | (1.047) | (1.173) |
| | Non Gambian | | | 1.466 | 1.320 |
| | | | | (1.086) | (1.223) |
| *Soliability ethnic receiver:* | Fula | | -0.549 | | -0.715 |
| | | | (0.834) | | (0.866) |
| | Sarehuleh | | 0.841 | | 0.886 |
| | | | (0.903) | | (0.902) |
| | Sererr | | 0.972 | | 1.106 |
| | | | (0.827) | | (0.870) |
| | Manjago | | 0.074 | | -0.015 |
| | | | (0.821) | | (0.833) |
| | Non Gambian | | -1.121 | | -0.923 |
| | | | (0.862) | | (0.876) |
| *Solidarity ethnic sender:* | Fula | | 0.664 | | 0.549 |
| | | | (1.095) | | (1.111) |
| | Sarehuleh | | 0.668 | | 0.736 |
| | | | (1.229) | | (1.229) |
| | Sererr | | -0.382 | | -0.236 |
| | | | (1.209) | | (1.256) |
| | Manjago | | 0.621 | | 0.485 |
| | | | (1.128) | | (1.155) |
| | Non Gambian | | 1.106 | | 1.194 |
| | | | (1.078) | | (1.083) |
| AIC | | 823.519 | 818.500 | 831.915 | 827.875 |
| BIC | | 902.408 | 970.210 | 947.215 | 1015.995 |
| Log Likelihood | | -398.760 | -384.250 | -396.958 | -382.937 |

Note: Additional control variables in terms of household attributes are included in all estimations. For differential ethnic homophily the excluded category is Sarehuleh. The ethnic category "Other" is not shown in order to save space. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Figure A2: MCMC diagnostics: Model 5

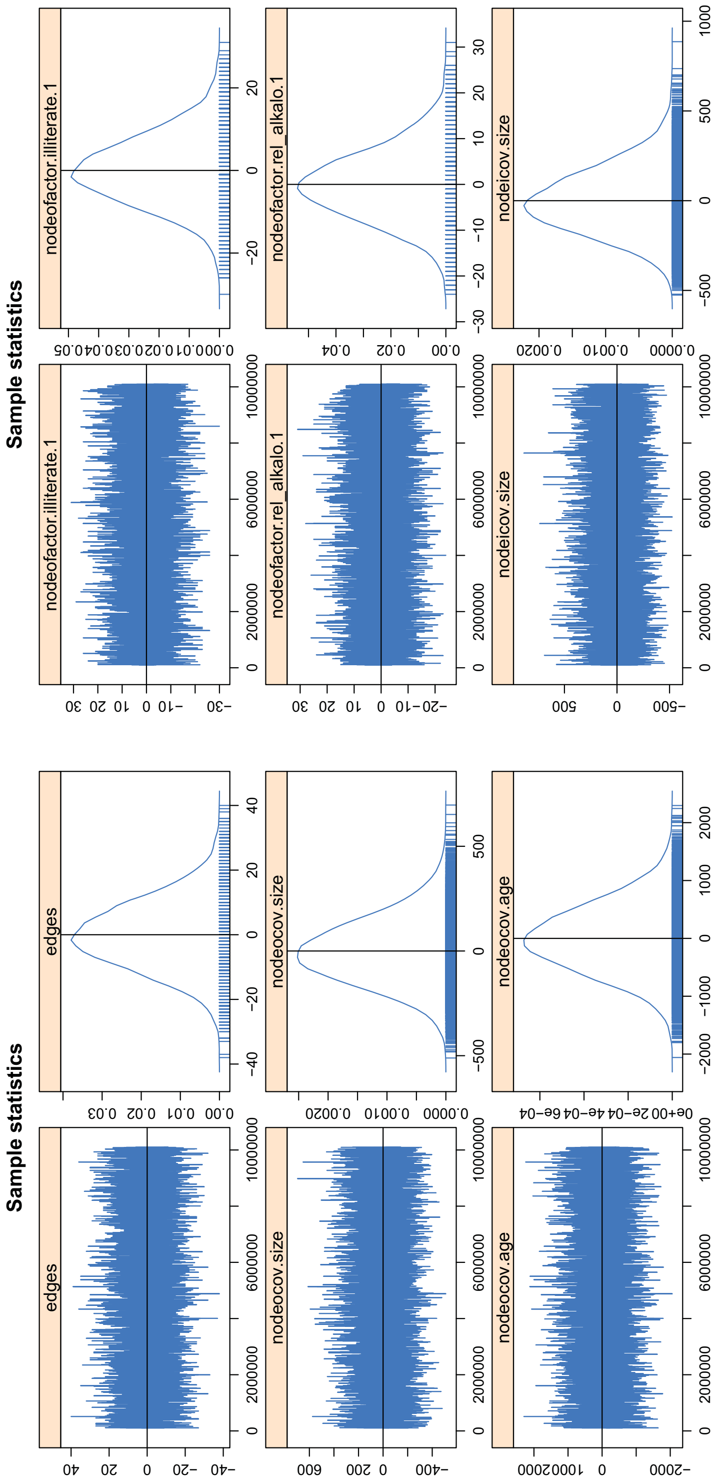


Figure A3: MCMC diagnostics: Model 5 (continued)

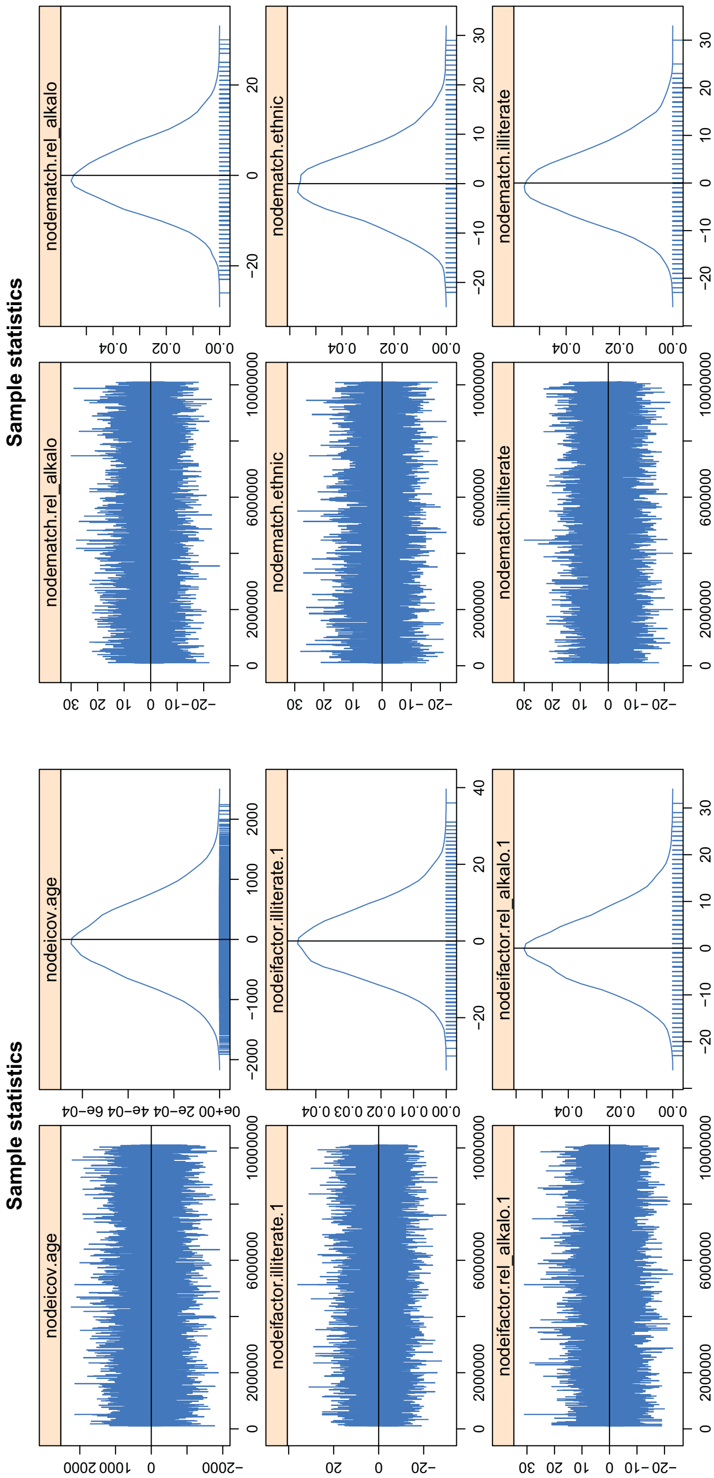


Figure A4: MCMC diagnostics: Model 5 (continued)

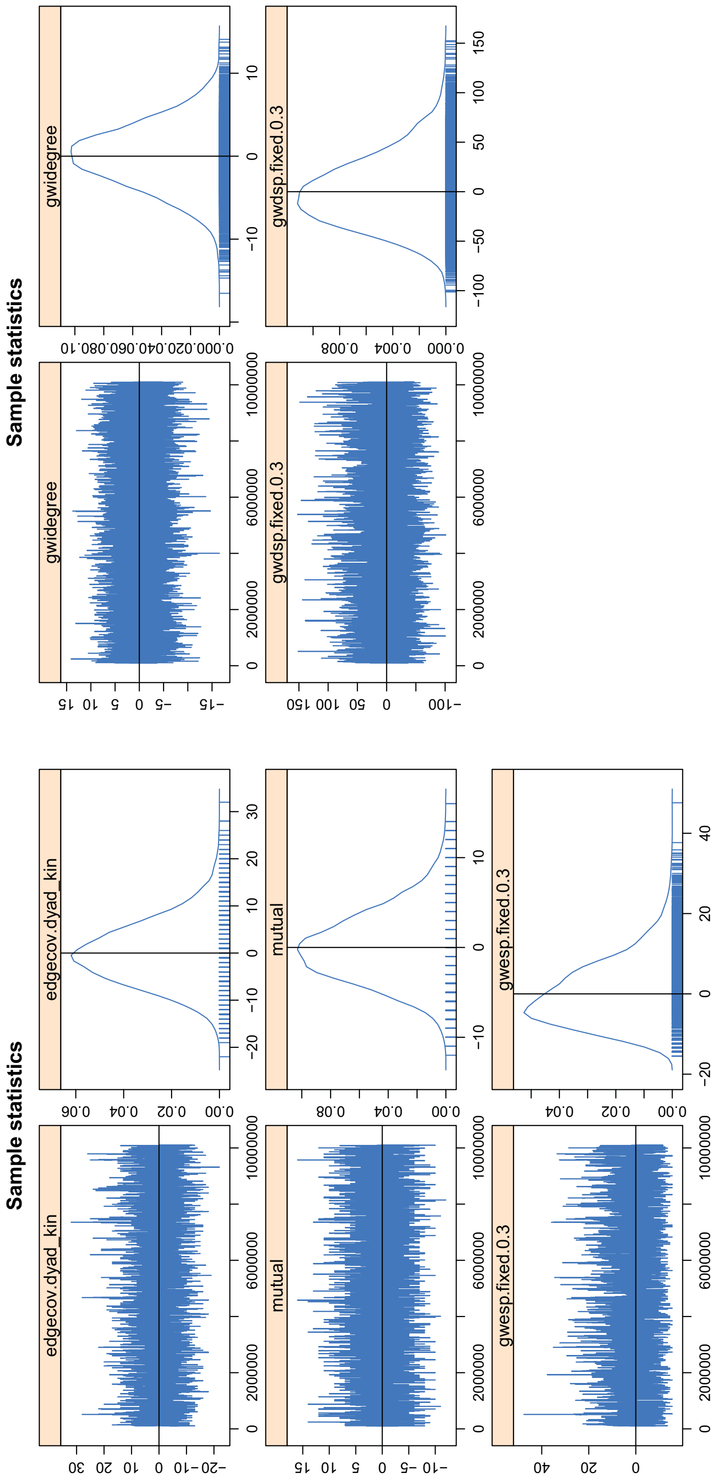


Figure A5: Goodness-of-Fit: Model 6 – Village 10

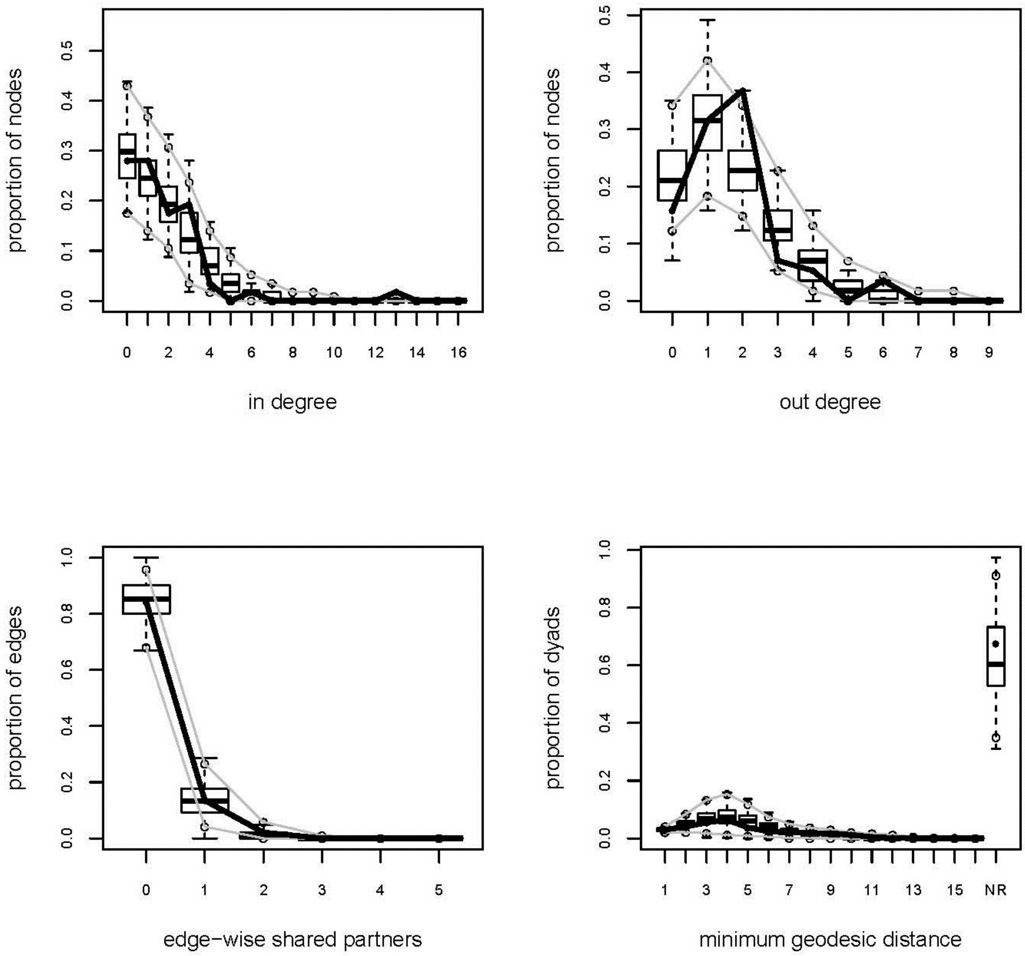


Table A2: Village 4

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| Edges | 1.234 | 0.347 | -2.812*** | -0.358 | -0.547 |
| | (0.901) | (0.949) | (0.507) | (0.927) | (1.304) |
| Household size sender | -0.007 | -0.005 | | -0.011 | -0.007 |
| | (0.014) | (0.014) | | (0.015) | (0.013) |
| Household age sender | -0.028*** | -0.026*** | | -0.022*** | -0.017** |
| | (0.008) | (0.008) | | (0.009) | (0.008) |
| Illiterate sender | -0.045 | -0.128 | | -0.124 | -0.143 |
| | (0.387) | (0.398) | | (0.396) | (0.355) |
| Relative *Alkalo* sender | -1.690*** | -1.537** | | -1.430** | -1.165** |
| | (0.590) | (0.597) | | (0.596) | (0.588) |
| Household size receiver | 0.018 | 0.023 | | 0.025* | 0.022 |
| | (0.014) | (0.014) | | (0.014) | (0.014) |
| Household age receiver | -0.023*** | -0.021** | | -0.016* | -0.017* |
| | (0.008) | (0.009) | | (0.009) | (0.010) |
| Illiterate receiver | -0.010 | -0.094 | | -0.066 | -0.109 |
| | (0.393) | (0.403) | | (0.406) | (0.404) |
| Relative *Alkalo* receiver | -1.018* | -0.797 | | -0.517 | -0.701 |
| | (0.589) | (0.596) | | (0.621) | (0.684) |
| Relative *Alkalo* homophily | -0.117 | -0.060 | | -0.070 | -0.225 |
| | (0.580) | (0.585) | | (0.581) | (0.559) |
| Ethnic homophily | 0.383 | 0.375 | | 0.308 | 0.358 |
| | (0.321) | (0.337) | | (0.312) | (0.291) |
| Illiterate homophily | -0.154 | -0.203 | | -0.155 | -0.147 |
| | (0.371) | (0.381) | | (0.346) | (0.363) |
| Kin ties | | 1.187*** | | 0.984*** | 0.889*** |
| | | (0.307) | | (0.309) | (0.270) |
| Reciprocity | | | 1.281*** | 1.330*** | 1.160** |
| | | | (0.470) | (0.504) | (0.504) |
| Triad closure (GWESP) | | | 0.960*** | | 0.571*** |
| | | | (0.233) | | (0.217) |
| In degree (GWID) | | | 0.919 | | 0.858 |
| | | | (0.927) | | (0.975) |
| Two paths (GWDSP) | | | -0.119 | | -0.187* |
| | | | (0.100) | | (0.100) |
| AIC | 340.468 | 326.713 | 328.522 | 321.881 | 317.906 |
| BIC | 388.951 | 379.236 | 348.723 | 378.444 | 386.591 |
| Log Likelihood | -158.234 | -150.356 | -159.261 | -146.940 | -141.953 |

Note: Estimation results for Village 52 using ERG models. Corresponds to Table 2 for Village 10. Akaike Information Criterion (AIC), and Bayesian information criterion (BIC). ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Table A3: Village 31

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| Edges | -1.450** | -1.739** | -2.858*** | -1.987*** | -1.998*** |
| | (0.725) | (0.752) | (0.291) | (0.754) | (0.743) |
| Household size sender | -0.004 | -0.007 | | -0.006 | -0.004 |
| | (0.016) | (0.017) | | (0.016) | (0.016) |
| Household age sender | -0.014 | -0.015 | | -0.015 | -0.014 |
| | (0.009) | (0.009) | | (0.009) | (0.009) |
| Illiterate sender | -0.244 | -0.238 | | -0.252 | -0.262 |
| | (0.260) | (0.271) | | (0.280) | (0.283) |
| Relative *Alkalo* sender | -0.822*** | -0.616** | | -0.572* | -0.500* |
| | (0.272) | (0.281) | | (0.295) | (0.284) |
| Household size receiver | -0.004 | -0.007 | | -0.006 | -0.006 |
| | (0.016) | (0.017) | | (0.017) | (0.014) |
| Household age receiver | -0.005 | -0.007 | | -0.004 | -0.002 |
| | (0.009) | (0.009) | | (0.009) | (0.007) |
| Illiterate receiver | 0.220 | 0.244 | | 0.272 | 0.190 |
| | (0.256) | (0.267) | | (0.278) | (0.226) |
| Relative *Alkalo* receiver | -0.562** | -0.366 | | -0.278 | -0.095 |
| | (0.272) | (0.281) | | (0.281) | (0.240) |
| Relative *Alkalo* homophily | 0.198 | 0.256 | | 0.220 | 0.227 |
| | (0.226) | (0.233) | | (0.227) | (0.222) |
| Ethnic homophily | 0.633* | 0.369 | | 0.324 | 0.179 |
| | (0.335) | (0.348) | | (0.330) | (0.283) |
| Illiterate homophily | 0.553*** | 0.457** | | 0.405* | 0.418** |
| | (0.214) | (0.219) | | (0.208) | (0.212) |
| Kin ties | | 1.397*** | | 1.238*** | 1.117*** |
| | | (0.214) | | (0.208) | (0.212) |
| Reciprocity | | | 1.404*** | 1.134*** | 1.072*** |
| | | | (0.386) | (0.384) | (0.400) |
| Triad closure (GWESP) | | | 0.343** | | 0.156 |
| | | | (0.175) | | (0.182) |
| In degree (GWID) | | | -1.342*** | | -1.409*** |
| | | | (0.502) | | (0.485) |
| Two paths (GWDSP) | | | 0.017 | | 0.005 |
| | | | (0.060) | | (0.061) |
| AIC | 780.845 | 742.099 | 757.455 | 736.623 | 730.152 |
| BIC | 845.074 | 811.681 | 784.217 | 811.557 | 821.143 |
| Log Likelihood | -378.422 | -358.050 | -373.727 | -354.312 | -348.076 |

Note: Estimation results for Village 31 using ERG models. Corresponds to Table 2 for Village 10. Akaike Information Criterion (AIC), and Bayesian information criterion (BIC). ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Table A4: Village 52

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| Edges | -2.035*** | -2.915*** | -3.813*** | -3.602*** | -3.500*** |
| | (0.754) | (0.861) | (0.365) | (0.696) | (0.793) |
| Household size sender | 0.009 | -0.010 | | -0.011 | -0.011 |
| | (0.012) | (0.013) | | (0.017) | (0.016) |
| Household age sender | 0.000 | 0.006 | | 0.005 | 0.005 |
| | (0.009) | (0.010) | | (0.013) | (0.012) |
| Illiterate sender | -0.084 | -0.049 | | 0.123 | 0.080 |
| | (0.247) | (0.269) | | (0.309) | (0.296) |
| Relative *Alkalo* sender | -0.228 | -0.233 | | -0.170 | -0.160 |
| | (0.271) | (0.291) | | (0.334) | (0.332) |
| Household size receiver | 0.013 | -0.006 | | 0.000 | 0.000 |
| | (0.012) | (0.013) | | (0.016) | (0.016) |
| Household age receiver | -0.001 | 0.005 | | 0.002 | 0.001 |
| | (0.009) | (0.010) | | (0.013) | (0.012) |
| Illiterate receiver | -0.234 | -0.217 | | -0.290 | -0.283 |
| | (0.247) | (0.268) | | (0.302) | (0.294) |
| Relative *Alkalo* receiver | -0.220 | -0.200 | | -0.106 | -0.109 |
| | (0.269) | (0.290) | | (0.337) | (0.338) |
| Relative *Alkalo* homophily | -0.565** | -0.247 | | -0.152 | -0.227 |
| | (0.238) | (0.254) | | (0.204) | (0.226) |
| Ethnic homophily | 0.813*** | 0.865*** | | 0.552*** | 0.545*** |
| | (0.203) | (0.224) | | (0.184) | (0.168) |
| Illiterate homophily | -0.290 | -0.316 | | -0.207 | -0.210 |
| | (0.222) | (0.238) | | (0.198) | (0.204) |
| Kin ties | | 2.581*** | | 1.605*** | 1.477*** |
| | | (0.212) | | (0.180) | (0.177) |
| Reciprocity | | | 3.888*** | 3.607*** | 3.455*** |
| | | | (0.360) | (0.386) | (0.373) |
| Triad closure (GWESP) | | | 0.489*** | | 0.345*** |
| | | | (0.123) | | (0.109) |
| In degree (GWID) | | | 0.535 | | 0.174 |
| | | | (0.798) | | (0.726) |
| Two paths (GWDSP) | | | -0.005 | | -0.061 |
| | | | (0.055) | | (0.052) |
| AIC | 870.478 | 725.431 | 694.159 | 626.610 | 620.743 |
| BIC | 933.460 | 793.662 | 720.402 | 700.089 | 709.967 |
| Log Likelihood | -423.239 | -349.716 | -342.080 | -299.305 | -293.371 |

Note: Estimation results for Village 52 using ERG models. Corresponds to Table 2 for Village 10. Akaike Information Criterion (AIC), and Bayesian information criterion (BIC). ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

# References


Arcand, J.-L. and M. Fafchamps (2012): "Matching in Community-Based Organizations," *Journal of Development Economics*, 98, 203–219.

Arcand, J.-L. and D. Jaimovich (2010): "Social and Economic Networks in Rural Gambia," *Working paper.*

Bala, V. and S. Goyal (2000): "A non-cooperative codel of network formation," *Econometrica*, 68, 1181–1231.

Beck, U. R. and B. A. Bjerge (2014): "Are inter-household land transactions pro-poor?," Working paper.

Burt, R. S. (1992): *Structural Holes.* Cambridge, MA: Harvard University Press.

Cartwright, D. and F. Harary (1956): "Structural Balance: A Generalization of Heider's Theory," *The Psychological Review*, 63, 277–293.

Chandrasekhar, A. G. and M. O. Jackson (2013): "Tractable and consistent random graph models," Working paper. Available at SSRN: http://ssrn.com/abstract=2150428.

Comola, M. (2010): "The Network Structure of Mutual Support Links: Evidence from Rural Tanzania," *Working paper.*

Conley, T. and C. Udry (2010): "Learning about a new technology: Prinapple in Ghana," *American Economic Review*, 100, 35–69.

Dasgupta, P. (2002): "Social Capital and Economic Performance: Analytics*," Working Paper.

Dercon, S. and J. D. Weerdt (2006): "Risk-sharing networks and insurance against illness," *Journal of Development Economics*, 81.

Fafchamps, M. and F. Gubert (2007): "The formation of risk sharing networks," *Journal of Development Economics*, 83, 326–350.

Fafchamps, M. and S. Lund (2003): "Risk-sharing networks in rural Philippines," *Journal of Development Economics*, 71, 261–287.

Feld, S. (1981): "The focused organization of social ties," *American Journal of Sociology*, 86, 1015–1035.

Frank, O. and D. Strauss (1986): "Markov Graphs," *Journal of American Statistical Association*, 81, 823–842.

Goodreau, S. M. (2007): "Advances in exponential random graph (p*) models applied to a large social network," *Social Networks*, 29, 231–248.

Goodreau, S. M., J. A. Kitts and M. Morris (2009): "Birds of a feather, or friend of a friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks," *Demography*, 46(1), 103–125.

Handcock, M. (2003): "Assessing degeneracy in statistical models of social networks," Working Paper no. 39. Center for Statistics and the Social Sciences, University of Washington.

Hiwatari, M. (2010): "A Network Structure of ROSCAs (Rotating Savings and Credit Associations) : ERGMs (Exponential Random Graph Models) Applied to a Leaders' Network in Rural Uzbekistan," Discussion Paper, Series A, 221: 1-24.

Hunter, D. R. (2007): "Curved exponential family models for social networks," *Social Networks*, 29, 216–230.

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

Jackson, M. O. and A. Wolinsky (1996): "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71(0108), 44–74.

Jaimovich, D. (2011): "Macrostructure and microstructure: Evidence from overlapping village networks in The Gambia," *Working Paper*.

——— (2013): "Missing links, missing markets: Internal exchanges, reciprocity and external connections in the economic networks of Gambian villages," MPRA Paper No. 44080.

Kossinets, G. and D. J. Watts (2009): "Origins of Homophily in an Evolving Social Network," *American Journal of Sociology*, 115(2), 405–450.

Krishnan, P. and E. Sciubba (2009): "Links and Architecture in Village Networks," *The Economic Journal*, 119, 917–949.

Mayer, A. and S. L. Puller (2008): "The Old Boy (and Girl) Network: Social Network Formation on University Compus," *Journal of Public Economics*, 92, 329–347.

McPherson, M., L. Smith-Lovin and J. M. Cook (2001): "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 27, 415–444.

Moody, J. (2001): "Race, school integration, and friendship segregation in America," *American Journal of Sociology*, 107, 679–716.

Ortiz-Pelaez, A., G. Ashenafi, F. Roger and A. Waret-Szkuta (2012): "Can Geographical Factors Determine the Choices of Farmers in the Ethiopian Highlands to Trade in Livestock Markets?," *PLoS ONE*, 7, e30710. doi:10.1371/journal.pone.0030710.

Pamela, K. (2010): *Land, labour and entrustment : West African female farmers and the politics of difference*. Leiden, Boston.

Pattison, P. E. and G. L. Robins (2002): "Neighbourhood Based Models for Social Networks," *Sociological Methodology*, 22, 301–337.

Platteau, J.-P. (2006): *Solidarity norms and institutions in village societies: static and dynamic considerations*. In: Kolm, S.-C., Mercier Ythier, J. (Eds.), Handbook of Economics of Giving, Altruism and Reciprocity, vol. 1. Elsevier.

Potter, G. E. and M. S. Handcock (2010): "A description of within-family resource exchange networks in a Malawian village," *Demographic Research*, 23, 117–152.

Robins, G., T. Snijders, P. Wang, M. Handcock and P. E. Pattison (2007): "Recent developments in exponential random graph (p*) models for social networks," *Social Networks*, 29, 192–215.

Snijders, T. A. B. (2002): "Markov Chain Monte Carlo Estimation of Exponential Random Graph Models," *Journal of Social Structure*, 3.

Snijders, T. A. B., P. E. Pattison, G. L. Robins and M. S. Handcock (2006): "New Specifications for Exponential Random Graph Models," *Sociological Methodology*, 36(1), 99–153.

Swindell, K. (1987): "Family Farms and Migrant Labour: The Strange Farmers of the Gambia," *Canadian Journal of African Studies*, 12(1), 3–17.

Udry, C. and T. Conley (2004): "Social Networks in Ghana," *Working paper*.

von Braun, J. and P. J. R. Webb (1989): "The Impact of New Crop Technology on the Agricultural Division of Labor in a West African Setting," *Economic Development and Cultural Change*, 37, 513–534.

Weerdt, J. (2002): "Risk-sharing and endogenous network formation," World Institute for Development Economics (UNU-WIDER), Discussion Paper No. 2002/57.

Weerdt, J. D. (2004): *Risk Sharing and Endogenous Network Formation*. Oxford University Press, Chapter 10.

Wimmer, A. and K. Lewis (2010): "Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook," *American Journal of Sociology*, 116(2), 583–642.

# Chapter 5

# Network benefits from co-membership in Mozambican business associations

Benedikte Alkjærsig Bjerge*

**Abstract**

Using network data to indicate whether any two firms are members of the same business association, this paper looks for evidence of assortative matching into associations among manufacturing firms in Mozambique. The results show that co-membership is restricted to specific members, whereas general membership in any business association is not determined by location and sector. Next, the paper examines the idea that business associations facilitate exchange of information about new technologies and business practices. Controlling for self-selection, firms' business practices are not found to be strongly correlated across co-membership status. This suggests that diffusion effects of technology and business practices are limited between members of the same business association in Mozambique.

**Keywords:** business associations, manufacturing firms, knowledge diffusion

**JEL classifications:** O1, D2, L2, L6

---

*Benedikte Alkjærsig Bjerge (benedikte.bjerge@econ.ku.dk) is employed at the Department of Economics, University of Copenhagen. The author thanks John Rand, Ulrik Richardt Beck, Finn Tarp, Neda Trifkovic, and participants at seminars at the University of Copenhagen for their helpful comments. I also thank my fieldwork colleagues, Søren Schou and Cardoso Jose. Without their support this study would not have been possible. I acknowledge financial support from the Development Economic Research Group (DERG) and am grateful for productive professional collaboration with colleagues at the National Directorate of Studies and Policy Analysis (DNEAP) in the Ministry of Planning and Development (MPD). Special thanks are also due to Steven Glover, Vincenzo Salvucci, and the Confederation of Business Associations (CTA) in Mozambique, who provided assistance in obtaining specific location information for business association headquarters.

# 1 Introduction

Recent years have brought about a more positive attitude towards the potential of business associations.[1] This interest is predicated on the premise that business associations may constitute an important feature in the public–private dialogue, and can help promote firm productivity through the provision of non-financial services as well as help build social capital by creating a forum for firms to identify business partners and new practices (Doner and Schneider, 2000; Goldsmith, 2002).[2] Existing evidence is largely qualitative in nature and focuses on individual business associations and their impact on state policy (Moore and Hamalai, 1993; Lucas, 1994; Nadvi, 1999; Heilbrunn, 1997; Heilman and Lucas, 1997; Bräutigam, Rakner and Taylor, 2002; Goldsmith, 2002). For instance, a World Bank study suggested that the Ugandan Manufacturing Association was closely involved in the broad set of reforms undertaken in the country (Devarajan, Dollar and Holmgren, 2001). Despite the policy interest in business associations, little research has been devoted to understanding the composition of business associations and the benefits of joining such associations in developing countries.

This paper provides elements of an answer taking a firm-level approach, using a micro, small, and medium-sized enterprise (MSMEs) survey from Mozambique conducted in 2012, and a smaller ethnographic study undertaken in the Maputo area and Beira. For the empirical analysis, a dyadic dataset of unique firm pairs is constructed. I consider firms to be linked if they are both members of the *same* business association.

The empirical analysis is divided into two steps. First, I investigate the composition of business associations by conducting a multivariate analysis on assortative matching applying dyadic regressions.[3] This approach allows me to overcome the problem that assortative criteria are often correlated. Using information about all possible firm pairs, I investigate whether two firms are more likely to belong to the same business association if they resemble each

[1] In earlier years negative attitudes towards business associations were partly due to the negative presumption inherent in New Institutional Economics (NIE) against special interest groups. This was first introduced by Mancur Olson in his book *The Logic of Collective Action: Public Goods and the Theory of Groups* from 1965. Olson emphasized that interest groups, like business associations, always pursue distributive objectives, seeking unproductive rents rather than the common or public interest.

[2] The term “business association” includes all formal membership organizations of businesspeople or firms concerned with business issues (Moore and Hamalai, 1993).

[3] An estimating equation is said to be dyadic if each observation corresponds to a pair of firms. Dyadic regressions are increasingly used to study network formation by economists (Fafchamps and Gubert, 2007).

other along different dimensions (for example, educational level of the owner and firm size) and are more proximate in terms of geographical distance or sectoral affiliation.

I then examine whether linked firms are more similar in terms of business practices. This analysis is motivated by the ethnographic study and the reported survey answers. These showed that the most important benefit received by Mozambican MSMEs is that business associations provide a forum for their members to interact (IIM, 2012). This suggests that diffusion of new knowledge and business practices is likely to take place between firms that are close in terms of co-membership in business associations. To test for evidence of diffusion along social networks based on co-membership, I estimate dyadic regressions comparing business practices across linked and unlinked firms. This test follows the approach suggested by Fafchamps and Söderbom (2014), who tested for presence of diffusion between trading partners in two African countries. If productivity-enhancing knowledge diffuses between co-members, then entrepreneurs who are members of the same business association are expected to have more similar business practices compared to non-members. Finding that co-members are more similar in business practices would suggest that adoption decisions are strategic complements: The incentive to adopt a certain practice increases if co-members have already adopted the practice. On the other hand, if members are more similar to non-members, then this suggests that firms' adoption decisions are strategic substitutes: The incentive to adopt a certain practice is reduced if co-members have already adopted the practice.

Presumably, enterprises that face similar challenges will tend to group together, which in turn increases the probability of adopting similar business practices. This, however, introduces self-selection into associations. Under the assumption that residential proximity to association headquarters reduces the cost of two firms joining the *same* business association, I correct for selection on unobservables by creating a predictor for collocation based on firms' distance to association headquarters. After controlling for the geographical distance between firm pairs ($ij$) and other firm characteristics, I argue that distances to a set of relevant association headquarters affect firms' decision to collocate, but does not independently affect the individual decision to adopt specific business practices.

This paper makes three contributions. First, I build quantitative evidence regarding the composition of business associations. Whether enterprise development can be achieved through business associations largely depends on their composition. If business associations are composed primarily of large firms or only specific sectors are represented, then interventions channeled through them are likely to reflect preferences and interests of these groups, rather than the broader business community. Knowledge of the composition of business associations is thus of interest to policy makers if professional business associations are expected to improve the business environment and enhance economic development.

Second, the paper contributes to the literature that focuses on the potential channels underlying knowledge diffusion between firms. The literature on agglomeration effects attributes a key role in knowledge diffusion and information sharing about business practices to geographical proximity (Audretsch, 1998). Fafchamps and Söderbom (2014) also find that geographical distance between firms matters for a firm's decision to adopt business practices. They investigate diffusion among manufacturing firms in Ethiopia and Sudan and find little evidence in support of diffusion between trading partners. Randomized field experiments have also been used to directly measure the causal effects of a firm's business peers on subsequent firm performance (Fafchamps and Quinn, 2012, 2013). Fafchamps and Quinn (2013) find evidence of diffusion of business practices among treated entrepreneurs who were in the same group. However, the authors also find evidence that performance among treated firms on average is more different ex-post compared to firms not in the same group.[4] They attribute these seemingly contradictory findings to firm heterogeneity: expected benefits from forming new links differ across firms. The mixed evidence on the importance of social networks for firms' adoption of new business practices and firm performance questions the narrative that diffusion of ideas and practices is essential for aggregate economic performance.

The final contribution lies in the focus on Mozambique, a country where a significant share of donor funds are channeled towards business associations each year (USAID, 2008).[5] The main objective underlying capacity building of business associations in Mozambique is that these are to work in conjunction with the formal legal system and help improve the business environment, which has been identified as the driving factor behind the low firm-level productivity (ICA, 2003, 2009; Sakar, 2004).[6]

[4] Other suggested diffusion channels are competition between domestic and foreign firms operating in the same market (Kraay, Soloaga and Tybout, 2002), and between spin-off firms and parent firms (Franco and Filson, 2006; Muendler, Rauch and Tocoiand, 2012).

[5] See also DI (2013); DANIDA (2012).

[6] Previous findings suggest that firms that show greater efforts at technology upgrading

Findings of this paper suggest that same-sector firms located close together are more likely to join the same business association. However, membership in business associations more generally is found not to be determined by sector and geographical location. Controlling for selection, I find limited evidence of knowledge diffusion of business practices between co-members, and almost no evidence of strategic complementarity in innovation practices or introduction of new technologies. Rather, if firm $i$ invests in R&D, then co-members of the same business association are less likely to adopt similar practices (i.e., strategic substitution). Moreover, there is strong evidence to suggest that firms located near each other differ more with respect to innovation and technology upgrading. These results complement the existing evidence on diffusion between manufacturing firms in sub-Saharan Africa and confirm previous findings that diffusion effects are limited between small and medium sized enterprises (Fafchamps and Söderbom, 2014).

The findings provide some insights into the nature of diffusion between firms and the underlying motivation for joining business associations. For example, emphasis on associations' ability to facilitate a forum for firms to interact indicates that associations play a more general part in development by building social capital (the ability to trust and work cooperatively with others) among members. The absence of assortative matching into business association more broadly contrasts the concerns raised in the public choice literature that business associations are likely to be captured by elites, but is consistent with a wide diversity of associations in many developing countries. Finally, the pattern of limited diffusion is consistent with the observed slow convergence of productivity between firms both across and within sectors in sub-Saharan Africa (Gelb, Meyer and Ramachandran, 2014).

The paper is structured as follows: I describe the survey data used in the empirical analysis in the next section. In Section 3, I present the empirical methodology. Section 4 outlines the testing strategy and discusses issues related to identification using distances to associations' headquarters. Section 5 presents the results of the analysis, followed by a concluding discussion in Section 6.

and worker training also show significantly higher productivity as measured by both sales and value added per worker (ICA, 2009). These are exactly the strategies business associations are likely to help provide.

# 2 Data

To shed light on the driving forces behind association membership and the implication for firms' business practices, I rely on two types of data: (i) an ethnographic study of business associations and business owners/managers, and (ii) a larger quantitative survey, including a special section on business associations.

## 2.1 Ethnographic fieldwork

The purpose of the fieldwork was to get a better understanding of the driving forces of association membership and its interplay with a firm's business networks. Specially, the objectives of the assessment were as follows: (1) To better understand the qualitative aspects of firm dynamics with respect to association membership, as well as the perceived benefits (and limitations) associated with business networks. (2) To investigate the formal (and informal) role of business associations and how they can help MSMEs overcome constraints in their business environment through the provision of services and information sharing. The fieldwork was carried out through semi-structured interviews with the owners of 25 enterprises, and spokespersons from 3 formal business associations. All interviews were conducted in August 2012. The interview was structured by a list of questions prepared in close connection to the survey answers. To achieve a better understanding of the driving forces behind a firm's decision not to join, both members and non-members were interviewed. As membership probability increases with size, micro enterprises are underrepresented.

**Why do firms collocate in business associations?**

The ethnographic study revealed some interesting insights into why firms collocate in business associations.

First, as outlined in the introduction, the main benefit of being a member of a business association is that associations provide a forum for firms to meet and interact. Not surprisingly, it was found that the opportunity to create new business links depends on the individual business association. This implies that firms may collocate in business associations to speed the flow of knowledge and new ideas. The underlying premise of this statement is that business associations provide a network of connections beyond those that are narrowly economic, facilitating opportunities for enterprises to network externally, to seek matches in the market, and to exchange specialist knowledge within a privileged group. This is consistent with McMillan and Woodruff's (2000) argument that networks created through business associations lower the cost of information gathering, resulting in better-informed manufacturers.

Second, the interviews revealed that many firms join business associations to get access to government subsidies such as VAT reductions on domestic purchases. These subsidies are often only accessible through association membership, even though membership is not free.[7] Hence, the subsidy tasks delegated to business associations give incentives for specific groups of firms to join sector-specific business associations, creating a grossly higher membership rates among some firms. To confirm that the methodology applied in the present paper is appropriate, questions related to tax exemptions are used to examine whether association members are more similar compared to non-members.

A third reason for collocation, often put forward in the developed country literature, is to reduce costs of access to non-financial services. Non-financial services may include training courses, technical consultancy, organization of trade fairs, seminars, conferences, promotion of business contacts, and the collection and dissemination of knowledge on issues affecting members (Doner and Schneider, 2000). Interviews revealed that entrepreneurs generally felt that the direct benefits normally associated with membership are weak. However, interviews confirmed that business associations in Mozambique arrange meetings and seminars supporting informal interaction and knowledge diffusion, though they are not perceived to be frequent enough. With few exceptions, the lack of non-financial services is largely explained by the fact that associations are still in their infant stage of building capabilities to meet firms' demands.

Finally, individual MSMEs are unlikely to be capable of taking part effectively in the interchange with government due to lack of expertise and time. In this case, membership in a group of like-minded businesses may empower firms and allow them to speak with a single voice through the business association. While Doner and Schneider (2000) discuss different cases where the business associations have been successful in impacting private sector development in developing countries, interviews with firms suggests that Mozambican business associations are still too weak to take an active part in the public–private

[7] For instance, in response to massive protests over increases in bread prices, the government introduced a subsidy on wheat flour to compensate bakeries in September 2010. The subsidy was only payable through the bakery association (Mozambican Association of Bakers, AMOPAO), and was only available to registered members.

dialogue.[8]

## 2.2 Firm-level survey

The paper uses a recent firm-level survey carried out in 2012. The survey covers manufacturing MSMEs in seven provinces (Maputo City, Maputo Province, Gaza, Sofala, Manica, Tete, and Nampula) throughout Mozambique. Firms to be interviewed in 2012 were identified based on the previous 2006 DNEAP and 2009 World Bank ICA survey (DNEAP, 2006; ICA, 2009). Additional new firms for the 2012 survey were identified using a snowball approach. Hence, firms were not drawn independently of each other; rather, the survey design relied on the local knowledge of the firms interviewed to identify other nearby manufacturing entrepreneurs. In provinces with surviving firms, these served as starting points for the snowball sampling approach. As a supplement, the updated 2002 census of enterprises (the CEMPRE) conducted by the National Institute of Statistics was used to identify broad "areas of industrial activity" and pick a random firm to start the snowball. Some of the firms were identified by the enumerators using other local information. In provinces where there were no previously interviewed firms, these areas were used as starting points (for more details see IIM, 2012). The sample was stratified by district location with the aim of including at least 50 percent of the manufacturing firms in each province. In addition, the sample only includes privately owned manufacturing enterprises that started operating prior to 2009.[9] The 2012 questionnaire builds on the one used in the 2006 DNEAP survey, but adds a new module on business associations. The new module on business associations requests the name of at most three business associations of which the firm is a member, as well as characteristics of its membership and satisfaction with the business association (regarding the most important association if more than one membership was reported).

The empirical analysis focuses on registered firms, as non-registered firms are not members of business associations.[10] Applying this selection criteria,

[8]Interviews with two business associations revealed two main factors preventing associations from representing the interests of the private sector. The first is the absence of a formalized cooperation channel between the government and the private sector. Second, the leaders of the association experienced a lack of government willingness to cooperate in the development of draft laws and new legislation.

[9]A privately owned firm is defined as a firm with a state share of ownership not higher than 50 percent. Manufacturing firms are defined as firms with no less than 50 percent of their sales in the manufacturing sectors of the International Standard Industrial Classification (ISIC) rev. 3 (ISIC categories 15-37).

[10]Firms without a unique tax payer identification number (NUIT) are categorized in the

the cross-section used for the empirical analysis includes 531 firms.

### Descriptive statistics

Table 1 provides information for all firms, and separately for those participating in business associations. Some 14 percent of the firms surveyed are members of a business association. The surveyed firms on average have 20 full-time employees, while the largest firm employs 251 workers. Less than 4 percent undertake export activities, and the majority of the firms are registered as sole proprietorships. This is partly explained by the large share of micro enterprises in the Mozambican economy.[11] The firms are predominately lead by educated men: Some 69 percent of the entrepreneurs have completed at least secondary education. Almost 40 percent of the firms in the sample are located in Maputo City, and more than 80 percent of the firms are concentrated in only five 2-digit sectors (food, apparel, wood, fabricated metal products, and furniture, etc.).

Furthermore, comparing members and non-members, it is evident that membership firms are significantly older, as well as larger in terms of the number of full-time employees.[12] These are also the firms more likely to be foreign owned and to export. Educational differences are driven by the top and bottom educational levels. Hence, entrepreneurs with only a primary education are less likely to be member of a business association, whereas entrepreneurs with a university degree are more likely to join a business association. Membership probability varies only slightly with firm location: Firms located in Tete province are less likely to join a business association. Finally, firms producing food and beverages and firms in chemicals, rubber, and plastic are significantly more likely to be a member of a business association.

Table 2 shows descriptive statistics regarding membership by firm size. The majority of the named business associations (81 percent) are registered as business associations under the Confederation of Business Associations (CTA). The CTA is an umbrella organization currently representing 70 business asso-

survey as non-registered (informal). Substantial shares of the firms that report having a NUIT are not registered, even though this in principle is needed.

[11]More than three quarters of the manufacturing firms in Mozambique are micro, and less than 1 percent are large (IIM, 2012).

[12]The membership shares for medium and large firms are lower compared to the sample across eight African countries used in Goldsmith (2002). He finds that 67 percent of small and medium-sized firms (fewer than 100 employees) are members of a business association, with an even larger membership share among large firms (100 or more employees). Part of the explanation is overrepresentation of larger firms. Moreover, it is not obvious whether the sample used by Goldsmith was selectively chosen to target larger firms and, if so, for what purpose.

Table 1: Descriptive statistics: All firms and by membership status

| | All | | Members | | Non-members | |
|---|---|---|---|---|---|---|
| | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. |
| Firm size | 19.878 | 34.207 | 54.757 | 52.578 | 14.230*** | 26.231 |
| Firm age | 14.94 | 12.186 | 22.878 | 16.110 | 13.654*** | 10.916 |
| State involvement | 0.009 | 0.097 | 0.054 | 0.228 | 0.002*** | 0.047 |
| Export | 0.034 | 0.181 | 0.081 | 0.275 | 0.026** | 0.160 |
| Sole proprietorship | 0.744 | 0.437 | 0.365 | 0.485 | 0.805*** | 0.396 |
| Partnership | 0.228 | 0.42 | 0.554 | 0.500 | 0.175*** | 0.380 |
| Limited liability company | 0.017 | 0.129 | 0.068 | 0.253 | 0.009*** | 0.093 |
| Other | 0.011 | 0.106 | 0.014 | 0.116 | 0.011 | 0.104 |
| Female owner | 0.094 | 0.292 | 0.108 | 0.313 | 0.092 | 0.289 |
| Foreign owner | 0.13 | 0.337 | 0.243 | 0.432 | 0.112** | 0.315 |
| Only primary school | 0.311 | 0.463 | 0.149 | 0.358 | 0.337*** | 0.473 |
| Secondary school | 0.196 | 0.397 | 0.122 | 0.329 | 0.208* | 0.406 |
| High school | 0.303 | 0.46 | 0.338 | 0.476 | 0.298 | 0.458 |
| University degree | 0.19 | 0.393 | 0.392 | 0.492 | 0.158*** | 0.365 |
| Ethnicity: African | 0.845 | 0.362 | 0.722 | 0.451 | 0.865*** | 0.342 |
| Ethnicity: European | 0.082 | 0.275 | 0.167 | 0.375 | 0.069*** | 0.253 |
| Ethnicity: Asian | 0.073 | 0.26 | 0.111 | 0.316 | 0.066 | 0.249 |
| Member of Frelimo | 0.322 | 0.468 | 0.270 | 0.447 | 0.330 | 0.471 |
| Maputo C | 0.36 | 0.48 | 0.405 | 0.494 | 0.352 | 0.478 |
| Maputo P | 0.094 | 0.292 | 0.095 | 0.295 | 0.094 | 0.292 |
| Sofala | 0.16 | 0.367 | 0.189 | 0.394 | 0.155 | 0.363 |
| Nampula | 0.098 | 0.297 | 0.122 | 0.329 | 0.094 | 0.292 |
| Manica | 0.13 | 0.337 | 0.122 | 0.329 | 0.131 | 0.338 |
| Tete | 0.089 | 0.284 | 0.014 | 0.116 | 0.101** | 0.301 |
| Gaza | 0.07 | 0.255 | 0.054 | 0.228 | 0.072 | 0.259 |
| Food, bev, tobacco | 0.194 | 0.396 | 0.459 | 0.502 | 0.151*** | 0.358 |
| Textiles | 0.015 | 0.122 | 0.014 | 0.116 | 0.015 | 0.123 |
| Apparel & footwear | 0.119 | 0.324 | 0.054 | 0.228 | 0.129* | 0.336 |
| Wood & paper | 0.121 | 0.326 | 0.108 | 0.313 | 0.123 | 0.328 |
| Publishing & printing | 0.024 | 0.155 | 0.027 | 0.163 | 0.024 | 0.153 |
| Chemicals, rubber, plastic | 0.015 | 0.122 | 0.054 | 0.228 | 0.009*** | 0.093 |
| Non-metallic minerals | 0.087 | 0.282 | 0.054 | 0.228 | 0.092 | 0.289 |
| Fabricated metal products | 0.209 | 0.407 | 0.122 | 0.329 | 0.223** | 0.417 |
| Machinery, etc. | 0.023 | 0.149 | 0.014 | 0.116 | 0.024 | 0.153 |
| Furniture & other mfg. | 0.194 | 0.396 | 0.095 | 0.295 | 0.210** | 0.408 |
| Observations | 531 | | 74 | | 457 | |

Note: Stars indicate statistically significant differences in member and non-member attributes. ***, ** and * indicate significance at the 10, 5, and 1 percent level, respectively.

ciations and chambers of commerce in all regions of the country. The rest of the business associations reported by the firms (19 percent) are regarded as informal business associations. Micro enterprises account for the vast majority of the firms registered under an informal business association. Both formal and informal business associations are included in the subsequent empirical analysis. Table A2 includes names of the business associations reported by the firms, and whether they are registered under CTA. Robustness is addressed in Section 5.5.[13]

Only 13 firms report being a member of more than one business association. The lack of overlapping multiple membership suggests that the market for associations is less complex compared to developed countries (Bennett, 1998).[14] On average, 78 percent of the firms state that the association represents the interest of the firm, and larger enterprises tend to be more satisfied with the interest representation. One possible explanation is that business associations better represent the interests of larger enterprises, as these firms have more capacity to formulate and communicate problems and more power to affect the selection of the association's focal points.

The survey also asked owners about the benefits of association membership and what topics they normally discuss with co-members. Answers to these questions are used to guide the subsequent empirical analysis. Overall, more than 80 percent of those surveyed said that associations bring benefits to the firm. As pointed out previously, the most important benefit reported is that associations function as a forum for firms to interact. Other highly ranked benefits include lobbying government, providing commercial and technical information, and lastly, enforcement of norms and quality standards.

According to Table 2, some 65 percent of the surveyed owners that are members of a business association said that they talk to co-members about new technologies and business practices. The topics discussed include new government legislation, access to resources, customers, and new innovations. The least discussed topic, independent of firm size, is informal credit opportunities.

[13]There is no reason at this stage to expect informal business associations to be different from formal business associations registered under CTA, except with respect to their ability to engage in the public–private dialogue.

[14]Results are unaffected by the choice to only use the main business association reported by the firm.

Table 2: Descriptive statistics: Membership firms

| | All members | Micro | Small | Medium |
|---|---|---|---|---|
| Member of a formal BA (registered under CTA) | 81.1 | 31.3 | 96.4 | 93.3 |
| Member of more than one BA | 17.6 | 6.3 | 14.3 | 26.7 |
| Pay membership fee | 81.6 | 76.5 | 85.7 | 80.6 |
| Does the BA represent the interest of the firm? | 77.6 | 70.6 | 78.6 | 80.6 |
| Does the BA bring benefits to the firm? | 82.2 | 68.8 | 81.5 | 90.0 |
| *Type of benefits the association brings* | | | | |
| Lobbying the government | 52.8 | 31.3 | 55.6 | 62.1 |
| Organizing commercial and technical fairs | 36.1 | 29.6 | 48.3 | 36.1 |
| Lobbying banks to facilitate access to credit | 23.6 | 22.2 | 24.1 | 23.6 |
| Access to key inputs | 33.3 | 29.6 | 41.4 | 33.3 |
| Providing commercial and technical information | 43.1 | 37.0 | 51.7 | 43.1 |
| Enforcing norms and quality standards | 47.2 | 44.4 | 55.2 | 47.2 |
| Resolving business disputes | 36.1 | 40.7 | 34.5 | 36.1 |
| Providing a “moral guarantee” to foreign partners | 23.6 | 6.3 | 22.2 | 34.5 |
| Forum to interact with other firms | 55.6 | 25.0 | 63.0 | 65.5 |
| Place to identify trading partners | 34.7 | 18.8 | 29.6 | 48.3 |
| Do you talk to co-members about new tech. and business practices? | 64.9 | 56.3 | 57.1 | 76.7 |
| *If yes, what do you talk about?* | | | | |
| New innovations | 56.3 | 22.2 | 56.3 | 69.6 |
| Suppliers | 56.3 | 22.2 | 50.0 | 73.9 |
| Customers | 64.6 | 55.6 | 50.0 | 78.3 |
| Access to bank credit | 35.4 | 22.2 | 25.0 | 47.8 |
| Informal credit opportunities | 10.4 | 11.1 | 0.0 | 17.4 |
| New government legislation | 60.4 | 22.2 | 56.3 | 78.3 |
| Access to resources | 60.4 | 44.4 | 50.0 | 73.9 |
| Observations | 74 | 16 | 28 | 30 |

Note: Percentages. Provides statistics exclusively for firms that are members of a business association (74 firms). BA is short for business association, and CTA is short for Confederation of Business Associations. All questions are related to the most important BA if the firm is member of more than one BA.

# 3 Empirical methodology

As discussed in Section 2.1, firms may obtain new knowledge both through interaction with co-members and through information disseminated by the business association. Business practices might therefore be similar between co-members as a result of information dissemination by the association and not as a result of knowledge sharing between co-members. Given the data, this study is unable to disentangle the different sources of information. However, the qualitative interviews lend credibility to the idea that knowledge diffuses between co-members. Moreover, independent of the source of the information, the adoption decision by the firm is similar.

## 3.1 Firms' decision to adopt new knowledge

While central dissemination by the business association and firm interaction are likely causes behind collocation in business associations, a firm's decision to adopt the new knowledge depends on the profitability and actions taken by co-members. Fafchamps and Söderbom (2014) develop a theoretical framework to investigate diffusion between firms that buy inputs from a common supplier, or sell outputs to a common client. I apply the same methodology but investigate diffusion between co-members in business associations. The model provides a test of whether business practices are strategic substitutes or complements by comparing whether business practices between linked firms are more or less similar. In this setting, linked firms are firms that are members of the *same* business association.

Formally, consider two firms $i$ and $j$ in a network $g_{ij}$. Let $g_i = [g_{1i}, ...., g_{iN}]$ be a vector of $i$'s neighboring firms, where $N$ is the total number of agents, and let $y \equiv [y_1, ...., y_N]$ denote a vector of the actions of all agents. The payoff function of firm $i$ from adapting action $y_i$ is as follows:

$$\begin{aligned} max\, \pi_i &= \alpha_i y_i + \gamma g_i y + \rho y_i g_i y - \frac{1}{2} y_i^2 \\ s.t.\, y_i &\geq 0 \end{aligned} \tag{1}$$

where parameter $\alpha_i$ represents the profitability of the action, $\gamma$ denotes the direction of the externality, and $\rho$ indicates whether the actions taken by firm $i$ are similar or dissimilar to actions taken by other firms in the network. The last term in equation (1) represents the cost of taking action $y_i$. Actions are strategic complements when $\rho > 0$ (or $\alpha_i \geq 0$), strategic substitutes when

$\rho < 0$ (or $\alpha_i \leq 0$), and neither of the two when $\rho = 0$. This means that similar practices may either be due to strategic complementarity or to correlation in linked firms' profitability of taking an action, and vice versa for dissimilar practices.[15] Three important observations follow from this:

First, in the long run, adoption patterns only depend on the distribution of benefits from adoption (i.e., profitability) and on local strategic substitutes and complements. If firms have dissimilar profitability or actions are strategic substitutes, then some firms adopt new practices while others do not. In contrast, if firms have similar profitability and actions are strategic complements, then firms have similar practices independent of whether they are directly or indirectly connected. However, in the short run, where knowledge diffuses more slowly, adoption decisions in the latter case are more likely among firms that are directly linked through co-membership in business associations.

Second, different types of proximity are likely to matter differently. For instance, strategic complementarities that arise from knowledge exchange between co-members would suggest that connectedness through associations matters. In contrast, strategic complementarities that arise from competition are likely to be stronger if co-members compete in the same market for the same pool of workers and customers. Hence, if business associations are composed of firms belonging to the same sector, all firms may choose to adopt more advanced technologies in order to keep up in terms of productivity or quality. The competition argument is increasingly relevant if the same-sector firms also operate in the same locality. For instance, if firms compete for the same pool of workers, they may be inclined to introduce worker training to become more attractive. If some firms adopt training practices, other firms might follow in order not to lose out on the best workers.

Finally, the patterns underlying knowledge diffusion depend on individual firm characteristics. Firms with a high profitability of an action will adopt new knowledge independently of what others do, whereas firms with low profitability are unlikely to adopt new knowledge, irrespective of other firms. Thus, strategic complements and substitutes are more relevant for firms with intermediate values of profitability. For them, adoption may only be beneficial if linked firms adopt (strategic complements) or do not adopt (strategic substitutes).

These observations translate into several important observations for the empirical analysis. The composition of the business associations is important

[15] This is also what Manski (1993) call contextual effects.

for the prediction of diffusion of new innovations and business practices. This means that the adoption decision depends on whether associations are restricted to certain types of members. If diffusion is slow enough and practices are strategic complements, then we expect that linked co-members are more similar compared to non-members. In addition, firm heterogeneity impacts adoption decisions, and thus the empirical analysis should allow the adoption decision between members to vary with certain observable firm characteristics, such as geographical distance and sectoral affiliation.

# 4 Testing strategy

## 4.1 Similarity in firms' outcomes

It is observed whether an enterprise $i$ has adopted a specific practice $y_i$. Moreover, define $b_{ijk} = 1$ if both firm $i$ and $j$ are members of the same business association $k$, and zero otherwise. If co-members share similar business practices, then observations on $y_i$ are not independent within groups: Firms have innovative practices that are more similar if they belong to the same association. This suggests investigating the presence of diffusion by testing:

$$E[|y_i - y_j||b_{ijk} = 1] < E[|y_i - y_j||b_{ijk} = 0]$$

It is possible that similarity in business practices is due to factors other than group membership. For instance, firms operating in the same sector may be more likely to have similar business practices. Further suppose that there is assortative matching by sector in the formation of business associations, such that firms belonging to the same sector are more likely to belong to the same business association. To allow for this possibility, a vector of dyadic controls $(x_{ij})$ is included to reduce omitted variable bias. The estimated model takes the form:

$$|y_i - y_j| = \alpha + \theta b_{ijk} + \gamma_1 x_{ij} + \kappa d_{ij} + e_{ij} \tag{2}$$

where $\theta$ is the coefficient of interest explaining the relationship between comembership and firms' business practices, and $u_{ij}$ is the error term. If $\theta < 0$, then firms $i$ and $j$ are more similar in $y$ when they are members of the same business association. This is consistent with business practices being strategic complements. Conversely, $\theta > 0$ implies that co-members are more dissimilar, suggesting that the adoption decision by different firms is consistent with strategic substitutes. Since our dependent variable is symmetric by construction ($|y_i - y_j| = |y_j - y_i|$), the dyadic controls must be constructed in such a way that $x_{ij} = x_{ji}$. To achieve this, I follow Fafchamps and Gubert (2007) and construct controls of the form $|x_i - x_j|$, where $x_i$ and $x_j$ are observable characteristics of $i$ and $j$. A positive coefficient on $|x_i - x_j|$ means that firms with similar attributes $x$ tend to be more similar in $y$. Finally, link-specific attributes are included in $d_{ij}$.

Another possibility is that business practices are similar because of self-selection on unobservables. The correction method applied follows a standard selection correction procedure summarized by Wooldridge (2002, ch. 18). A selection equation of the form is estimated:

$$b_{ijk} = \phi(\varphi w_{ij} + \varphi_3 d_{ij} + \eta z_{ij}) \tag{3}$$

where $w_i$ is attributes of firm $i$ that affect the likelihood of belonging to the same business association, and $z_{ij}$ is the instrument used for identification. Equation (3) can be used to correct for self-selection, but it also allows us to investigate assortative matching into business associations.[16]

Two kinds of regressors are included: sums $(w_i + w_j)$ and absolute differences $|w_i - w_j|$. Combining these types of regressors enables a distinction between the case where members have a higher/lower $w$ than non-members from cases where members have a more similar/dissimilar $w$ than non-members. The coefficient on $|w_i - w_j|$, $\varphi_1$, identifies negative and positive assortative matching. A negative $\varphi_1$ indicates positive assortative matching: The more dissimilar $i$ and $j$, the less likely they are to be in the same group. Similarly, a positive $\varphi_1$ indicates negative assortative matching: Members of the same business association are less similar to each other compared to the rest of the population. The coefficient on $(w_i + w_j)$, $\varphi_2$, captures the propensity for a firm to join a business association conditional on $w$. This means that the interpretation is similar to that of standard linear regression estimates: A positive $\varphi_2$ indicates that this characteristic is associated with membership in the larger group.

To rule out the possibility that characteristics of business association members are a consequence of membership in equation (3), the attributes of the firm ($w_i$) to be included are reasonably time-invariant, such as year of establishment, nationality, and the educational level of the owner. In a second step,

[16] Arcand and Fafchamps (2012) use the same approach to investigate matching of households in community-based organizations in Senegal and Burkina Faso.

firm size and the legal status under which the firm is currently registered are also included. Measures concerning firms' economic status or physical assets are excluded from the analysis. The two distance measures included in $d_{ij}$ are outlined below.

From equation (3) two inverse Mills ratios of the form $b_{ijk}\frac{\hat{\phi}}{\hat{\Phi}}$ and $(1 - b_{ijk})\frac{\hat{\phi}}{1-\hat{\Phi}}$ are computed and included in equation (2) as control variables for selection:

$$|y_i - y_j| = \alpha + \theta b_{ijk} + \gamma x_{ij} + \rho\,(x_{ij} - \bar{x})\,b_{ijk} + \kappa d_{ij} + \delta_1 b_{ijk}\frac{\hat{\phi}}{\hat{\Phi}} + \delta_2\,(1 - b_{ijk})\,\frac{\hat{\phi}}{1-\hat{\Phi}} + \xi_{ij} \quad (4)$$

where the inverse Mills ratios is nonzero one at a time since $b_{ijk} = \{0, 1\}$. This allows us to investigate unobserved heterogeneity considering both members and non-members. Joint significance of $\delta_1$ and $\delta_2$ implies the presence of unobserved heterogeneity. As $\delta_1$ and $\delta_2$ relate to co-members and non-members, respectively, these can be used to investigate assortative matching on firms unobservable ability to introduce new business practices. $\delta_1 > 0$ and $\delta_2 < 0$ imply that membership firms in the same association are less similar in terms of unobserved ability to introduce new business practices than if they were matched at random, i.e., negative assortative matching. In contrast, $\delta_1 < 0$ and $\delta_2 > 0$ indicate positive assortative matching, which means that firms belonging to the same association are more similar in their unobserved ability to introduce new practices. In the case where business associations develop around firms with higher unobserved ability, the estimated parameter of $\theta$ is expected to increase once selection on unobservables is controlled for.

The firm attributes included as absolute differences ($|x_i - x_j|$) are the educational level of the owner and the size and age of the firm. The additional control variables are included as link-specific ($d_{ij}$) characteristics. The first is a dummy equal to one if the nationality of the owner differs between firm $i$ and $j$, and zero otherwise. Second, I construct a dummy variable equal to one if the gender of the owners differs, and zero otherwise. Finally, two distance measures are constructed and included in both equation (3) and (4). The first distance measure is the physical distance between firm $i$ and $j$. The distance is computed as the Euclidean distance between firms based on the latitude and longitude obtained using GPS receivers. The second measure is a dummy variable that takes the value one if both firms belong to the same sector using 2-digit ISIC categories. Summary statistics of the dyadic variables are reported in Table A1.

One important caveat to keep in mind when interpreting the results is that dissimilar practices between firms may be due to either strategic substitution or to negative correlation in the profitability of taking action $y$ between linked firms. If unobserved profitability of a given practice is more strongly correlated across linked firms, it would bias $\theta$ below 0. Hence, a negative and statistically significant estimate of $\theta$ may be due either to diffusion or to unobserved profitability of a given practice. However, if $\theta$ is positive or not statistically significantly different from zero, then the net effect of diffusion and unobserved profitability of a given practice are likely to be positive (Fafchamps and Söderbom, 2014). Observing $\theta > 0$ is therefore evidence in support of some level of diffusion between firms.

Equation (3) and (4) to be estimated are both dyadic regressions. The dependent and independent variables are defined for every pair of firms $ij$ in the data. This implies that there are $n \times (n-1)$ observations underlying the regression, where $n$ denotes the number of firms. This means that the empirical analysis is performed on $531 \times 530/2 = 140,715$ unique enterprise pairs, also called dyads. Based on the unique pairs $(ij)$ I create different network measures. The most direct network measure is whether firm $i$ and $j$ are both members of the *same* business association. In the data, some 253 pairs of firms are directly linked through co-membership. A broader network measure applied is whether firm $i$ and $j$ are both members of a business association. This type of link is more common: 2,701 links are identified between membership firms.

Dyadic observations are generally not independent, as residuals containing the same firm are correlated. To compute standard errors that are robust to correlation in the error terms across firms, standard errors are bootstrapped following the procedure described by Fafchamps and Söderbom (2014). To clarify, I briefly outline the bootstrapping method applied. First, a random sample of $n$ firms is drawn from the firm-level survey (531 observations) with replacement. Second, a dyadic dataset containing $n \times (n-1)$ observation is constructed. Third, the dyadic regressions are estimated based on the new sample, and parameter estimates are stored. I repeat this process 1,000 times. Finally, an average of the standard deviations of the estimated parameters is used as the standard error.

## 4.2 Identification

The instrument included in the selection equation ($z_{ij}$) should affect collocation but not similarity/dissimilarity between firms' business practices once we control for self-selection. The geographical distance to headquarters of a relevant set of business associations serves this purpose. The underlying idea of the identification strategy is to use the geographical location of firms relative to that of the business association as a source of variation in firms' information about the existence of a specific business association and time costs of participating in association activities. Since the level of analysis is the pair of firms ($ij$), it is possible to use the location information on headquarters to create a predictor for the probability that two firms collocate. Building on the argument for a single firm, two firms ($i$ and $j$) are more likely to join the same business association if they are both geographically close to the headquarters. This approach differs from previous studies using geographical distances, for example, between firms and tax registration offices in cities (i.e., McKenzie and Sakho, 2010), as the predictor developed here is based on a pool of potential business associations.

Formally, I construct a predictor to describe the probability for two firms to join the same business association ($z_{ij} = P\left(ij \in b_{ijk}\right)$). Under the assumption that firms can only collocate in one business association, the predictor is constructed as follows:

$$\begin{aligned} P\left(ij \in b_{ijk}\right) &= P(ij \in k = 1) + ... + P(ij \in k = N) \\ &= P(i \in k = 1) \times P(j \in k = 1) + ... + P(i \in k = N) \times \\ &\quad P(j \in k = N) \end{aligned}$$

where $k$ is the business association ($k \in 1, ...N$) and $b_{ijk}$ equals 1 if firm $i$ and $j$ are members of the same business association. The maximum number of relevant business associations $N$ varies across firms because membership in some business associations is restricted to certain members based on sector. $P(i \in k = 1) = |\hat{\psi}_{ik}| \times db_{ik}$ where $\hat{\psi}_{ik}$ is obtained from estimating $b_{ik} = \alpha + \psi db_{ik} + \mu_{ik}$ in which $b_{ik}$ equals 1 if firm $i$ is a member of association $k$, and $db_{ik}$ is the geographical distance between firm $i$ and the headquarters of business association $k$.[17] The composition of the predictor changes slightly when the broader network measure is used: whether firm $i$ and $j$ are members of any business association. In this case, membership in all relevant associations

[17]This method relies on two assumptions. First, two firms cannot be co-members of two business associations at the same time. Second, firm $i$'s decision to join association $k$ is independent of $j$'s decision to join association $k$.

for firm $i$ is treated as mutually exclusive. Next, the probability that firm $i$ joins a business association is treated as independent from firm $j$'s decision.[18]

Since membership in business associations is not restricted based on location, all potential associations need to be considered. In collaboration with CTA, information on the location of formally registered business association headquarters was obtained. Based on the headquarters' addresses, GPS coordinates were attained using Google Earth (earth.google.com). Unfortunately, the online maps available for Mozambique do not include individual household numbers. For this reason, the midpoint of the road was used as a proxy for the location of headquarters of business associations registered under CTA. For business associations not registered under CTA, firms' survey answers were used to identify in which province the headquarters are located.[19] To proxy for headquarters location within the province, a focal point in the main city of the province was chosen. Table A3 summarizes whether the location of the headquarters is precisely identified using the formal address or whether a proxy is applied for the location of headquarters. The relevant set of business associations for each firm is identified based on firm answers. Fortunately, the firm survey asked membership firms whether membership in the business association is restricted to specific members. The sectoral affiliation of members in the restricted business association was then used to identify the sector(s) from which firms can potentially obtain membership.[20] Table A3 indicates whether the business association reported by the firm is restricted and to which type of firms.

For the predicted collocation probability to be a valid instrument, it must satisfy both the inclusion and exclusion restrictions. The subsequent estimation results of equation (3) (shown in Table 3) confirm a strong correlation between observed co-membership and the predicted probability of co-membership based on the distance to association headquarters. Regarding the exclusion restriction, we may be concerned that the predicted probability also influences a firm's adoption decision: If firms $i$ and $j$ are located geographically close to each other, they are both more likely to join the same association and adopt the same business practices under the assumption that diffusion decreases over geographical distance. To overcome this, the geographical distance between the

[18]Formally, the construction of the probability changes to: $P\left(ij \in b_{ij}\right) = (P(i \in k = 1) + P(i \in k = 2) + ... + P(i \in k = N) \times (P(j \in k = 1) + P(j \in k = 2) + ... + P(j \in k = N))$

[19]The survey included the following question: *"In what province is the headquarters of the business association located?"*

[20]2-digit classification in the International Standard Industrial Classification (ISIC) rev. 3 (ISIC 15-37).

production facilities of firm $i$ and $j$ are included in both the selection and outcome regression. This means that identification relies on the difference in distance between firms $i$ and $j$, and the distance to association headquarters. Figure A1 illustrates this graphically.

There are two main limitations of this identification strategy. First, the set of business associations is restricted to the set of associations reported by the surveyed firms. Thus, the set of business associations is the lower bound of the potential set of business associations available to manufacturing firms in Mozambique. The total number of formal business associations in Mozambique amounts to 70 business associations; however, not all of these associations are open to manufacturing firms. A crude estimation based on the names of the formal business associations registered under CTA indicate that around 27 business associations are open to manufacturing firms. Out of these, 14 associations are represented in the firm survey (in addition to CTA), while the rest of the business associations are sector-specific, representing firms not surveyed in the IIM 2012 survey. Exclusion of relevant business associations will bias the predicted probability of co-membership ($P\left(ij \in b_{ijk}\right)$) downwards and thus underestimate the probability of collocation. This will lead to downward bias of $\theta$ in equation 4. Finding that the predictor helps explain co-membership even in the presence of this potential issue can therefore be considered as evidence that an effect is present.

Second, the location of business associations may be endogenous. For instance, it may be that business associations are formed in places where the private sector has grown rapidly. The larger mass of firms that face similar challenges is likely to increase the probability that a business association is formed. Attempts to account for endogenous location are controlled for in the empirical analysis by means of town fixed effects. However, it is not possible to account for shocks, such as a sudden change in the business environment, that lead firms to create a business association.

# 5 Results

## 5.1 Assortative matching

To address assortative matching into business associations based on observables, equation (2) is estimated to identify the factors that affect the likelihood that any two firms belong to the same business association. The estimation results are presented in Table 3. In column 1 the dependent variable $b_{ijk}$ equals one if firm $i$ and $j$ belong to the *same* business association $k$, and zero otherwise. Remember that the estimated dyadic regression includes all firms. Hence, the dependent variable in column 1 takes the value zero either when $i$ and $j$ do not belong to the same association *or* when they do not belong to any association. This is meaningful on the basis that the question of interest is how firms sort into individual business associations. Robustness is addressed in Section 5.5. In columns 2 and 3, the dependent variable is generalized to determine whether a firm joins any business association. The dependent variable equals one if firms $i$ and $j$ belong to the same *or* to any other business association, and zero otherwise.

Several interesting patterns emerge from Table 3. First, the predicted co-membership probability based on distances to association headquarters is statistically significant and positive: Higher predicted co-membership increases the probability of observed co-membership. To determine whether the pair $ij$ belong to any business association, column 2 control the membership predictor based on the largest probability that a firm belong to a business association $P(ij \in b_{ij})_{max}$, while column 3 control for the predictor based on mutual exclusion in all possible business associations $P(ij \in b_{ij})$. The predictor in column 2 is positive and statistically significant on a 5 percent level, while the broader defined predictor in column 3 is insignificant. In the remaining part of the paper, $P(ij \in b_{ij})_{max}$ is used to correct for self-selection into business associations.

Second, firms assort into the *same* business associations according to the sector in which they operate (i.e., positive assortative matching), thought the estimate is only significant on a 10 percent level. This confirms the intuition that firms collocate in associations based on sector, and suggests that business associations are specialized and sector specific. As expected, the sign of the coefficient estimate changes when the dependent variable is $b_{ij}$ in column 2. However, the estimate become statistically insignificant. Hence, when we match firms depending on whether they belong to any business association, we find no evidence of positive assortative matching into associations.

Third, the geographical distance between firms is only significant in column 1. The negative sign implies negative assortative matching on the geographical distance between firms. Hence, co-membership does not travel over long distances: Geographically close firms are more likely to assort into the *same* business association. In contrast, geographical proximity does help explain why firms join a business association more broadly. The lack of assortative matching on geographical distance in columns 2 and 3 provide support for the observation that membership is not location specific as suggested in the descriptive statistics Table 1.

Turning to the other regressors, the sign of the coefficient estimates, $\varphi_1$ and $\varphi_2$, are consistent across the three columns. Recall that a negative coefficient on $|w_i - w_j|$ implies positive assortative matching. The estimation result suggests some level of positive assortative matching on firm size, age, and legal ownership status. This means that firms of similar size, age, and legal status are more likely to be co-members. Furthermore, the positive estimate on $(w_i + w_j)$ for firm size and age indicates that larger and older firms are more likely to join a business association. In addition, column 2 also suggests that membership is not restricted to owners of a specific nationality (i.e., negative assortative matching).

## 5.2 Similarity and dissimilarity in firms' outcomes

The objective is to find quantitative evidence to support the motivated collocation benefits of business associations. I test whether the practices related to innovativeness, technology, and benefits from tax exemptions are more similar among firms that are close to each other, either in a network sense through association membership or geographically. To address self-selection into business associations based on unobservables, inverse Mills ratios are calculated based on estimation of equation (3). The estimation technique used to estimate equation (4) is linear (OLS), and standard errors are bootstrapped to be robust to heteroskedasticity and correlation in errors across firms.

First, benefits from tax exemptions are investigated. Dyadic dependent variables are constructed from dummy variables measuring whether firms enjoy tax exemptions related to VAT on domestic purchases, VAT on imports, and custom duties. The regression results are shown in Table 4. Each panel has a different dependent variable. Columns 1, 2, and 3 investigate diffusion patterns using co-membership ($b_{ijk}$), while potential diffusion in columns 4, 5, and 6 is measured in terms of general association membership ($b_{ij}$). Columns 1 and 4 in Panels A, B, and C are not corrected for self-selection on unobservables. Across the three panels in columns 2 and 5, the estimated coefficients on membership using Mills ratios as controls are presented. Finally, columns 3 and 6, in addition to the Mills ratios, also include the regressors $(x_{ij} - \bar{x})\, b_{ijk}$. Hence, $\hat{\theta}$ in columns 2 and 5 corresponds to the average treatment effects, whereas $\hat{\theta}$ in columns 3 and 6 corresponds to the average treatment effect of

Table 3: Assortative matching

| | 1 | 2 | 3 |
|---|---|---|---|
| Predictor co-members: $P(ij \in b_{ijk})$ | 1.286*** | | |
| | (0.301) | | |
| Predictor both members: $P(ij \in b_{ij})_{max}$ | | 1.075** | |
| | | (0.435) | |
| Predictor both members: $P(ij \in b_{ij})$ | | | 2.182 |
| | | | (1.689) |
| Belong to the same sector | 0.390* | -0.089 | 0.252 |
| | (0.239) | (0.128) | (0.137) |
| Geographical distance (log) | -0.210*** | 0.021 | 0.005 |
| | (0.039) | (0.023) | (0.013) |
| *Absolute difference in:* | | | |
| Educational level of owner | 0.010 | 0.007 | 0.006 |
| | (0.082) | (0.043) | (0.043) |
| Nationality of owner | 0.019 | 0.062** | 0.071* |
| | (0.269) | (0.025) | (0.025) |
| Firm size (log) | -0.203** | -0.053 | -0.057 |
| | (0.096) | (0.059) | (0.059) |
| Firm age (log) | -0.011* | -0.007* | -0.007* |
| | (0.007) | (0.004) | (0.004) |
| Registered legal status | -0.338** | -0.192** | -0.183* |
| | (0.139) | (0.093) | (0.094) |
| *Sum of:* | | | |
| Educational level of owner | -0.017 | -0.053 | -0.063 |
| | (0.069) | (0.057) | (0.057) |
| Nationality of owner | -0.286 | -0.257* | -0.272 |
| | (0.333) | (0.165) | (0.166) |
| Firm size (log) | 0.258*** | 0.339*** | 0.364*** |
| | (0.087) | (0.072) | (0.071) |
| Firm age (log) | 0.017*** | 0.018*** | 0.018*** |
| | (0.005) | (0.004) | (0.004) |
| Registered legal status | 0.125 | 0.098 | 0.096 |
| | (0.117) | (0.097) | (0.098) |
| Observations | 140,715 | 140,715 | 140,715 |

Note: Probit. The dependent variable in column 1 is $b_{ijk}$, while the dependent variable in columns 2 and 3 is $b_{ij}$. A constant is included in all specifications. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

the treated.

Using either method, the coefficient on $b_{ijk}$ is statistically insignificant across all columns in the three panels. However, coefficient estimates on $b_{ij}$ are positive and statistically significant for the dependent variable of whether the firm enjoys tax exemptions from VAT on domestic purchases (Panel A, columns 4, 5, and 6). Hence, access to VAT exemptions on domestic purchases is dissimilar across association members. This is in line with the fact that exemptions on domestic purchases are limited to some sectors and depend on membership in a formal business association (see Section 2.1). Moreover, the coefficient estimate in columns 5 and 6 of Panel A are significantly larger than the estimate reported in column 4. This suggests that the methodology reveals the presence of self-selection into business associations above and beyond that captured by selection on observables. This is also confirmed by the fact that the coefficients on the inverse Mills ratios are jointly significantly different from zero. The two Mills ratios implies positive assortative matching in association membership, $b_{ij}$: Firms belonging to business associations are more similar in their unobserved ability to introduce new business practices.

Second, I investigate practices related to innovativeness and technology upgrading. Estimation results are presented in Tables 5 and 6. For co-membership in the same business associations, only the coefficient estimate for R&D is statistically significant at the 10 percent level. The positive sign suggests that co-members are dissimilar in R&D investments. The estimated effect implies that the likelihood that co-members report the same answer to whether they employ staff exclusively for R&D is 47 percentage points lower for firms that are co-members (column 3, Panel A). This result is not consistent with the notion that network proximity tends to result in similar practices. The obvious reason is that firms are able to free-ride on other firms' investments in R&D, and so there is no evidence supporting the idea that if some firms invest in R&D, others will invest in R&D in order to stay competitive. The estimation results including a variable for whether both firms are members of any business association ($b_{ij}$) support this finding (Panel A, columns 4, 5, and 6) at a 5 percent level.

Whereas collocation does not lead to similar R&D practices, innovativeness in terms of the likelihood to either improve existing products, introduce new products or a new technology suggests that that membership firms tend to be more similar (Table 6: Panel B, columns 4, 5, and 6); however, the coefficient estimate is only statistically significant at the 10 percent level. The estimated

Table 4: Diffusion: Tax exemptions

| Panel A | Tax exemptions from VAT on domestic purchases | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| *Co-members* | 0.086 | (0.078) | 0.347 | (0.229) | 0.436 | (0.418) |
| Same sector | 0.012 | (0.012) | 0.003 | (0.012) | 0.003 | (0.012) |
| Distance (log) | -0.006** | (0.003) | -0.004 | (0.003) | -0.004 | (0.003) |
| IMR1 | | | -0.115 | (0.091) | -0.116 | (0.123) |
| IMR2 | | | 0.798** | (0.322) | 0.797** | (0.322) |
| Panel B | Tax exemptions from VAT on imports | | | | | |
| | 1 | | 2 | | 3 | |
| *Co-members* | 0.027 | (0.080) | -0.240 | (0.223) | -0.386 | (0.406) |
| Same sector | 0.013 | (0.011) | 0.011 | (0.011) | 0.011 | (0.011) |
| Distance (log) | -0.007*** | (0.003) | -0.006** | (0.003) | -0.006** | (0.003) |
| IMR1 | | | 0.125 | (0.091) | 0.151 | (0.124) |
| IMR2 | | | 0.199 | (0.350) | 0.200 | (0.350) |
| Panel C | Tax exemptions on customs duties | | | | | |
| | 1 | | 2 | | 3 | |
| *Co-members* | 0.007 | (0.080) | -0.038 | (0.262) | 0.350 | (0.542) |
| Same sector | 0.001 | (0.011) | 0.000 | (0.012) | 0.000 | (0.012) |
| Distance (log) | -0.011*** | (0.003) | -0.011*** | (0.003) | -0.011*** | (0.003) |
| IMR1 | | | 0.021 | (0.107) | -0.068 | (0.159) |
| IMR2 | | | 0.059 | (0.327) | 0.057 | (0.327) |

| | Tax exemptions from VAT on domestic purchases | | | | | |
|---|---|---|---|---|---|---|
| | 4 | | 5 | | 6 | |
| *Both BA member* | 0.076** | (0.043) | 0.302*** | (0.108) | 0.301*** | (0.112) |
| Same sector | 0.011 | (0.012) | 0.004 | (0.013) | 0.004 | (0.013) |
| Distance (log) | -0.006** | (0.003) | -0.006** | (0.003) | -0.006** | (0.003) |
| IMR1 | | | -0.113** | (0.050) | -0.112** | (0.054) |
| IMR2 | | | 0.418*** | (0.153) | 0.418*** | (0.153) |
| | Tax exemptions from VAT on imports | | | | | |
| | 4 | | 5 | | 6 | |
| *Both BA member* | 0.076 | (0.055) | 0.021 | (0.144) | 0.031 | (0.164) |
| Same sector | 0.012 | (0.011) | 0.010 | (0.011) | 0.011 | (0.012) |
| Distance (log) | -0.007** | (0.003) | -0.007** | (0.003) | -0.007** | (0.003) |
| IMR1 | | | 0.033 | (0.070) | 0.030 | (0.079) |
| IMR2 | | | 0.126 | (0.163) | 0.125 | (0.163) |
| | Tax exemptions on customs duties | | | | | |
| | 4 | | 5 | | 6 | |
| *Both BA member* | 0.020 | (0.053) | -0.012 | (0.135) | -0.007 | (0.160) |
| Same sector | 0.001 | (0.011) | -0.001 | (0.012) | -0.001 | (0.012) |
| Distance (log) | -0.011*** | (0.003) | -0.011*** | (0.003) | -0.011*** | (0.003) |
| IMR1 | | | 0.020 | (0.066) | 0.019 | (0.080) |
| IMR2 | | | 0.105 | (0.160) | 0.105 | (0.160) |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. BA is short for business association. A constant and control variables are included in all specifications. In addition, columns 3 and 6 include demeaned interactions times co-membership and both BA members, respectively. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

effect suggests that members of business associations are 17 percentage points more likely to report innovative activities compared to non-members. These results suggest that for innovativeness, strategic complementarity effects dominate strategic substitution for membership firms.

Turning to the control variables ($x_{ij}$ in equation (4)), including the same-sector variable and the geographical distance between firms, some interesting results emerge. For the dependent variable on whether firms enjoy tax exemptions, the estimated coefficients on same sector are not statistically significant in any of the estimations. However, the positive sign in Panels A, B, and C are in line with the field observation that only selected sectors enjoy access to tax exemptions through their membership in a formal business association. In contrast, the sign of the same-sector dummy is negative in all specifications for the dependent variables related to firms' innovativeness (Tables 5 and 6). This is likely to be explained by the fact that firms in the same sector tend to follow similar patterns of innovativeness.

Next, the role of geographical proximity between firms is considered. With only a few exceptions, the distance coefficient is negative in all specifications shown in Tables 4, 5, and 6, and is statistically significant, at least at the 5 percent level. Hence, geographical proximity tends to be associated with greater differences in innovation practices and access to tax exemptions. These results are similar to Fafchamps and Söderbom (2014) and suggest that strategic substitution effects dominate strategic complementarities for firms located near each other. One exception to this finding is that geographical proximity increase co-members probability to improve existing products (i.e. stratetic complements). Competition is likely to explain the observed similarity in firms' decision to upgrade existing products: Firms that compete within a given region keep up in terms of productivity by improving their products.

### 5.3 Co-membership diffusion between geographically close firms

To investigate whether co-membership diffusion is faster between geographically close co-members, an interaction between co-membership $b_{ijk}$ and geographical distance is included in equation (4). By this, I wish to establish whether co-members operating in the same area are more likely to take up the same business practices. The results of this analysis are reported in the Appendix, Tables A3-A5. Conditional on being co-members in the same business association, geographical proximity does not explain firms decision to innovate

Table 5: Diffusion: Innovation 1

| Panel A | Employ staff exclusively for R&D | | | | | | | Employ staff exclusively for R&D | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | | | 4 | | 5 | | 6 | |
| Co-members | 0.133* | (0.074) | 0.141 | (0.239) | 0.468* | (0.282) | Both BA member | 0.113* | (0.060) | 0.309** | (0.138) | 0.317** | (0.162) |
| Same sector | -0.003 | (0.009) | -0.011 | (0.009) | -0.010 | (0.009) | Same sector | -0.003 | (0.009) | -0.013 | (0.010) | -0.012 | (0.010) |
| Distance (log) | -0.006*** | (0.002) | -0.004*** | (0.002) | -0.004** | (0.002) | Distance (log) | -0.006*** | (0.002) | -0.006*** | (0.002) | -0.006*** | (0.002) |
| IMR1 | | | 0.002 | (0.099) | -0.103 | (0.133) | IMR1 | | | -0.094 | (0.067) | -0.098 | (0.071) |
| IMR2 | | | 0.753** | (0.370) | 0.751** | (0.370) | IMR2 | | | 0.532 *** | (0.187) | 0.531*** | (0.180) |
| Panel B | Introduced a new product | | | | | | | Introduced a new product | | | | | |
| | 1 | | 2 | | 3 | | | 4 | | 5 | | 6 | |
| Co-members | 0.040 | (0.065) | 0.068 | (0.231) | 0.317 | (0.422) | Both BA member | 0.040 | (0.039) | 0.034 | (0.104) | 0.036 | (0.104) |
| Same sector | -0.013 | (0.012) | -0.011 | (0.012) | -0.011 | (0.012) | Same sector | -0.013 | (0.012) | -0.015 | (0.012) | -0.015 | (0.011) |
| Distance (log) | -0.006*** | (0.002) | -0.006*** | (0.002) | -0.006*** | (0.002) | Distance (log) | -0.006*** | (0.002) | -0.006*** | (0.002) | -0.006*** | (0.002) |
| IMR1 | | | -0.014 | (0.096) | -0.088 | (0.128) | IMR1 | | | 0.007 | (0.051) | 0.006 | (0.056) |
| IMR2 | | | -0.137 | (0.276) | -0.138 | (0.279) | IMR2 | | | 0.151 | (0.125) | 0.151 | (0.128) |
| Panel C | Improved existing products | | | | | | | Improved existing products | | | | | |
| | 1 | | 2 | | 3 | | | 4 | | 5 | | 6 | |
| Co-members | -0.038 | (0.066) | 0.092 | (0.223) | 0.045 | (0.416) | Both BA member | -0.029 | (0.031) | -0.135 | (0.100) | -0.138 | (0.095) |
| Same sector | 0.002 | (0.006) | 0.006 | (0.007) | 0.006 | (0.007) | Same sector | 0.003 | (0.006) | 0.003 | (0.007) | 0.003 | (0.006) |
| Distance (log) | 0.006** | (0.003) | 0.006** | (0.003) | 0.006** | (0.003) | Distance (log) | 0.006** | (0.003) | 0.006** | (0.003) | 0.006** | (0.003) |
| IMR1 | | | -0.063 | (0.096) | -0.051 | (0.129) | IMR1 | | | 0.058 | (0.048) | 0.059 | (0.041) |
| IMR2 | | | -0.370 | (0.258) | -0.369 | (0.258) | IMR2 | | | -0.028 | (0.063) | -0.027 | (0.063) |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. BA is short for business association. A constant and control variables are included in all specifications. In addition, columns 3 and 6 include demeaned interactions times co-membership and both BA members, respectively. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Table 6: Diffusion: Innovation 2

| Panel A | Introduced a new product or improved existing product | | | | | | | Introduced a new product or improved existing product | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | | | 4 | | 5 | | 6 | |
| *Co-members* | 0.036 | (0.091) | 0.199 | (0.318) | 0.394 | (0.654) | *Both BA member* | 0.041 | (0.038) | -0.069 | (0.121) | -0.071 | (0.125) |
| Same sector | -0.011 | (0.014) | -0.008 | (0.014) | -0.007 | (0.014) | Same sector | -0.012 | (0.015) | -0.015 | (0.015) | -0.015 | (0.014) |
| Distance (log) | 0.002 | (0.004) | 0.001 | (0.004) | 0.002 | (0.004) | Distance (log) | 0.002 | (0.004) | 0.003 | (0.004) | 0.003 | (0.004) |
| IMR1 | | | -0.078 | (0.136) | -0.139 | (0.199) | IMR1 | | | 0.066 | (0.059) | 0.066 | (0.063) |
| IMR2 | | | -0.376 | (0.331) | -0.376 | (0.331) | IMR2 | | | 0.201 | (0.135) | 0.201 | (0.135) |
| Panel B | Introduced a new product, improved existing product or introduced a new technology | | | | | | | Introduced a new product, improved existing product or introduced a new technology | | | | | |
| | 1 | | 2 | | 3 | | | 4 | | 5 | | 6 | |
| *Co-members* | -0.029 | (0.068) | 0.025 | (0.228) | -0.002 | (0.412) | *Both BA member* | -0.046* | (0.037) | -0.167* | (0.100) | -0.174* | (0.105) |
| Same sector | 0.004 | (0.007) | 0.010 | (0.008) | 0.010 | (0.008) | Same sector | 0.004 | (0.007) | 0.008 | (0.008) | 0.007 | (0.008) |
| Distance (log) | 0.010** | (0.004) | 0.008** | (0.004) | 0.008** | (0.004) | Distance (log) | 0.010** | (0.004) | 0.010** | (0.004) | 0.010** | (0.004) |
| IMR1 | | | -0.029 | (0.098) | -0.026 | (0.130) | IMR1 | | | 0.061 | (0.051) | 0.063 | (0.045) |
| IMR2 | | | -0.613* | (0.314) | -0.612* | (0.314) | IMR2 | | | -0.213** | (0.099) | -0.212** | (0.099) |
| Panel C | Introduced a new technology | | | | | | | Introduced a new technology | | | | | |
| | 1 | | 2 | | 3 | | | 4 | | 5 | | 6 | |
| *Co-members* | 0.095 | (0.069) | 0.195 | (0.234) | 0.253 | (0.470) | *Both BA member* | 0.093* | (0.037) | 0.096 | (0.099) | 0.093 | (0.110) |
| Same sector | -0.042*** | (0.015) | -0.049*** | (0.015) | -0.049*** | (0.015) | Same sector | -0.042*** | (0.015) | -0.050*** | (0.015) | -0.050*** | (0.015) |
| Distance (log) | -0.007** | (0.003) | -0.006** | (0.003) | -0.006** | (0.003) | Distance (log) | -0.008*** | (0.003) | -0.007** | (0.003) | -0.007** | (0.003) |
| IMR1 | | | -0.042 | (0.099) | -0.044 | (0.145) | IMR1 | | | 0.011 | (0.046) | 0.013 | (0.050) |
| IMR2 | | | 0.646** | (0.287) | 0.645** | (0.287) | IMR2 | | | 0.494*** | (0.133) | 0.494*** | (0.133) |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. BA is short for business association. A constant and control variables are included in all specifications. In addition, columns 3 and 6 include demeaned interactions times co-membership and both BA members, respectively. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

or their access tax exemptions.

## 5.4 Co-membership diffusion between same-sector firms

To further investigate whether diffusion is stronger among co-members that operate in the same sector, the exercise above is repeated by interacting co-membership with a dummy variable for whether firms $i$ and $j$ are both members of the same sector. The estimation results are shown in the Appendix, Tables A6-A8. The co-membership–sector interaction term is only statistically significant for R&D expenditures: Co-members in the same sector have differential attitudes towards R&D. Thus, strategic substitution seems to be equally strong within and across sectors.

## 5.5 Robustness checks

### Assortative matching

The dyad regression reported in Table 3, column 1 includes all firms, even those who do not belong to a business association. In this regression the dependent variable is zero either when $i$ and $j$ do not belong to the same business association or when $i$ and $j$ are not members of a business association. This is a meaningful approach since we are trying to understand how firms join together in the same business association. However, one concern is that the factors that affect membership in any association differ from those that affect which specific association a firm joins conditional on joining. To investigate this further, column 1 of Table 3 is re-estimated using only firms that belong to a business association. To correct for possible self-selection, a probit regression using firm-level data is estimated, and from these individual firm-specific Mills ratios, $R_i$ are created. Regressors of the form $|R_i - R_j|$ and $(R_i + R_j)$ are then created and included in the dyad regression of equation (2).[21]

Regression results shown in Table 7, column 1 indicate that both inverse Mills ratios are statistically significant. The negative coefficient estimate on $|R_i - R_j|$ suggests that large differences in the predicted firm probabilities decrease co-membership probability. Hence, different factors seem to drive whether firms choose to join any association and which specific association

[21]The identifying selection variables include the number of association members known (for association members, the number refers to members known prior to membership), ethnicity dummies, and whether the owner is member of the ruling party, Frelimo. Due to missing values, some 30 observations are lost.

a firm joins. This is meaningful on the basis that the sample of business associations includes associations open to different members (for instance, the Federation of Industry, AIMO-FE), as well as sector-specific business associations such as the bakery association, AMOPAO.

### Formal vs. informal business associations

As previously summarized, some 19 percent of the business associations are not registered under the CTA, and thus are regarded as informal business associations. As above, one may worry that selection into formal and informal business associations differs. Table 3 is re-estimated allowing only for membership firms in formal business associations, while membership in informal business associations is set to zero. Estimation results are shown in Table 7, columns 2 and 3. The sign and magnitude of the estimated coefficients are similar to the baseline estimation. Only the significance level of some of the control variables has changed slightly.

### Association activity

It is conceivable that assortative matching is driven by the type of activity undertaken by the business association. If certain activities are related to certain groups of firms, such as exporting firms, this may drive the correlations reported in Table 3. Certain activities are more likely to be related to certain groups if a business association is restricted to specific members, and the activities in turn are largely determined by its members. Fortunately, the survey included answers to questions regarding membership restriction and activity choice by the business association. Table 8 reports survey answers depending on whether the business association is formally registered under CTA. The null hypothesis that the organization of business associations is similar across formality status cannot be rejected. The summary statistics suggest that the business associations considered in this paper largely differ in terms of organization and selection of focal points.

Table 8: Business associations

| | Yes | No | All | Total obs. |
|---|---|---|---|---|
| Formally registered under CTA | 60 (81%) | 14 (19%) | | 74 |
| | Yes | Yes | Yes | |
| Membership is restricted | 23 (49%) | 6 (60%) | 29 (51%) | 57 |
| BA activity chosen by members | 38 (78%) | 8 (80%) | 46 (78%) | 59 |

Table 7: Robustness: Assortative matching

| | Robustness of the dependent variable | | Sample restricted to registered members under CTA | | | |
|---|---|---|---|---|---|---|
| | Belong to the same BA | | Belong to the same BA | | Belong to a BA | |
| | 1 | | 2 | | 3 | |
| Pred. co-membership probability | 0.559 | (0.474) | 0.955*** | (0.301) | 1.206*** | (0.369) |
| Belong to same sector | 0.715** | (0.344) | 0.476** | (0.239) | -0.164 | (0.121) |
| Geographical distance (log) | -0.260*** | (0.054) | -0.215*** | (0.039) | 0.022 | (0.021) |
| *Absolute difference in:* | | | | | | |
| Educational level of owner | 0.044 | (0.121) | 0.016 | (0.082) | 0.013 | (0.040) |
| Nationality of owner | -0.066 | (0.299) | 0.033 | (0.270) | 0.061*** | (0.024) |
| Firm size (log) | -0.129 | (0.135) | -0.186* | (0.096) | -0.051 | (0.056) |
| Firm age | -0.012 | (0.008) | -0.011 | (0.007) | -0.007** | (0.004) |
| Registered legal status | -0.413 | (0.265) | -0.312** | (0.139) | -0.203** | (0.097) |
| IMR1 | -0.290* | (0.175) | | | | |
| *Sum of:* | | | | | | |
| Educational level of owner | 0.020 | (0.084) | -0.002 | (0.069) | -0.049 | (0.052) |
| Nationality of owner | -0.119 | (0.331) | -0.290 | (0.333) | -0.290* | (0.158) |
| Firm size (log) | -0.002 | (0.107) | 0.312*** | (0.087) | 0.375*** | (0.074) |
| Firm age | 0.011* | (0.006) | 0.017*** | (0.005) | 0.019*** | (0.004) |
| Registered legal status | 0.172 | (0.189) | 0.105 | (0.117) | 0.105 | (0.105) |
| IMR2 | -0.339** | (0.157) | | | | |
| Observations | 20,970 | | 136,969 | | 136,969 | |

Note: Probit. Column 1 is restricted to association members. Columns 2 and 3 are restricted to business associations that are registered under the CTA. The dependent variable in columns 1 and 2 is $b_{ijk}$, while the dependent variable in columns 3 is $b_{ij}$. A constant is included in all specifications. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

# 6 Conclusion

Using network data from Mozambique to indicate whether any two firms are members of the same business association, this paper investigates assortative matching into business associations and whether co-members are more similar compared to non-members. In order to distinguish members' adoption decisions from self-selection into business associations, I use distance between firms and association headquarters. Residential proximity to association headquarters can reasonably be expected to reduce the cost of joining the same business association. To control for the fact that nearby firms are also more likely to hear about the same business practices, which in turn might influence their adoption decisions, the paper also exploits data on firm location.

First, the results reveal positive assortative matching into the same business association by sector and firm size and age, as well as physical distance. However, membership of business associations more broadly defined is not found to be determined by firms sectoral status and geographical location. This finding is consistent with the wide diversity of business associations in Mozambique. These results are robust to the definition of association membership, status, and organization of the business association.

Second, the results show that the main benefit associated with membership is that business associations create a forum for firms to interact and talk about new technologies and business practices. This is in line with the firm innovation literature that small firms, particularly in developing countries, acquire economic knowledge from interaction with business peers. This suggests that business associations can play a central role in development by adding to society's stock of social capital.

Third, I find limited evidence in support of knowledge diffusion of business practices between co-members. There is no strong evidence that network proximity in terms of co-membership is associated with similarity in innovations or technology upgrading. Rather, there is strong evidence to suggest that firms located near each other differ more with respect to innovation and technology upgrading. This somewhat surprising result is in line with the previous findings by Fafchamps and Söderbom (2014) in Ethiopia and Sudan. Moreover, investments in R&D are found to be strategic substitutes: If firm $i$ invests in R&D, then co-members of the same business association are less likely to adopt similar practices. This finding is robust within and across sectors, as well as between geographically distant firms and consistent with Fafchamps and Söderbom (2014). Moreover, results confirm the findings in the ethnographic study that membership firms gain access to tax exemptions delegated to business associations by the government. This implies that associations can help provide services normally provided by the state, and so Mozambican business associations might be a valuable place to divert other state assignments.

Though the findings related to assortative matching are clear, their interpretation is less straightforward. For instance, it is unknown whether members were initially more resourceful and more alike compared to non-members before they joined a business association, or whether they became so after they joined. The results related to assortative matching should therefore be interpreted with caution as they are only valid if reverse causality is not present. Furthermore, it is uncertain whether some activities by the business association also benefit non-members. If this were the case, it would be less surprising that co-members' business practices are not more correlated compared to non-members. Whether this is the case in Mozambique is doubtful. From interviews with directors of three formal business associations as well as membership firms, it was confirmed that associations have limited power to engage in the public–private dialogue. This is partly due to the lack of legal obligation on the part of the government to consult the relevant parties in the reforming process of the private sector.

In light of the limited evidence in support of diffusion, a few caveats should be pointed out. First, as noted by Fafchamps and Söderbom (2014), correlation in business practices does not imply diffusion, as unobserved contextual effects may be driving firms' decision to take up certain business practices. However, the evidence on tax exemptions of domestic goods in favor of membership firms confirms the appropriateness of the empirical methodology. Along the same lines, we cannot distinguish adoption due to possible diffusion between firms and dissemination by business associations. Second, finding limited result of diffusion does not mean that diffusion does not take place between co-members. Though the choice of business practices considered was guided by survey answers, we might have considered the wrong business practices. This would cause the diffusion effects to be lower or even absent. Third, despite the focus on the development of business associations by donors and the government, business associations are still in their infancy in Mozambique when it comes to delivering business services and lobbying the government. Thus, the majority of them are somewhat limited in performing many of the functions normally undertaken by their counterparts in other countries. At this stage of development it therefore might not be possible to detect differences in business practices across members and non-members.

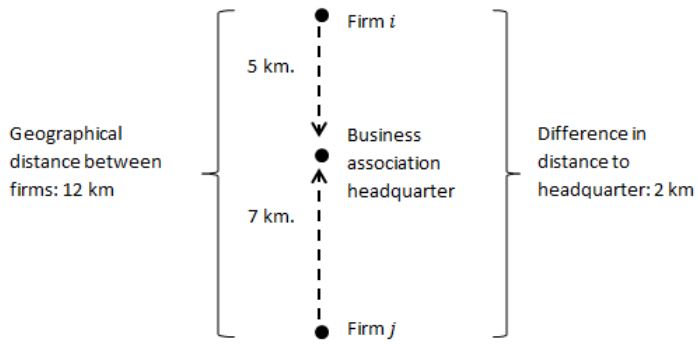
Acknowledging these caveats, the paper complements and confirms previous findings that diffusion is limited between African entrepreneurs. The study further confirms previous evidence that geographically proximate firms are more dissimilar compared to firms located farther apart. This is consistent with the large heterogeneity of African firms and the slow convergence of productivity observed both across and within sectors in sub-Saharan Africa (Gelb, Meyer and Ramachandran, 2014).

# Appendix A

Table A1: Descriptive statistics, dyad level

| | Mean | Std. Dev. | Min. | Max. |
|---|---|---|---|---|
| Both member of BA | 0.019 | 0.137 | 0 | 1 |
| Co-membership in same BA | 0.002 | 0.042 | 0 | 1 |
| Same sector | 0.150 | 0.357 | 0 | 1 |
| Distance betwen $i$ and $j$ (kilometers) | 580.54 | 507.90 | 0.064 | 3964.58 |
| *Absolute difference in:* | | | | |
| Educational level | 1.498 | 1.154 | 0 | 4 |
| Nationality of owner | 0.227 | 0.419 | 0 | 1 |
| Gender | 0.171 | 0.376 | 0 | 1 |
| Firm size (log) | 1.276 | 1.023 | 0 | 5.53 |
| Firm age | 12.535 | 11.826 | 0 | 79 |
| Legal ownership form | 0.512 | 0.887 | 0 | 6 |
| *Sum of:* | | | | |
| Educational level | 5.559 | 1.887 | 2 | 10 |
| Nationality of owner | 0.260 | 0.475 | 0 | 2 |
| Gender | 0.188 | 0.413 | 0 | 2 |
| Firm size (log) | 4.377 | 1.632 | 0 | 10.94 |
| Firm age | 29.879 | 17.201 | 4 | 152 |
| Legal ownership form | 2.644 | 1.022 | 2 | 14 |
| Number of unique firm pairs | 140.715 | | | |

Figure A1: Identification on distance to headquarters




Note: The dotted lines indicates the difference in geographical distance to business association headquarters. Distance is measured in kilometers (km).

Table A2: Relevant business associations

| Acronym | Name (Portuguese) | Translated name | CTA member | Member of AIMO |
|---|---|---|---|---|
| ACB | Associação Comercial da Beira | Commercial Association of Beira | x | |
| ACIS | Associação Comercial e Industríal de Sofala | Commercial and Industrial Association of Sofala | x | |
| ACIANA | Associação Comercial, Industríal e Agricola de Nampula | Commercial, Industrial and Agricultural Association of Nampula | x | |
| ASSOCIET | Associação dos Empresários de Tete | Businessmen Association Tete | x | |
| AIOPA | Associação dos Industríais de Óleo e Produtos Afins | Association of Manufacturers of Oil and Related Products | x | x |
| AIGM | Associação dos Industríais Gráficos de Moçambique | Association of Industrial Graphics Mozambique | x | |
| AEM | Associação Empresarial de Manica | Business Association of Manica | x | |
| ACTIVA | Associação Moçambicana de Mulheres Empresarias e Executivas | Mozambican Association of Women Entrepreneurs and Executives | x | |
| CCM | Câmara de Comércio de Moçambique | Chamber of Commerce of Mozambique | x | |
| CCMOBRA | Câmara de Comércio Moçambique-Brasil | Chamber of Commerce Mozambique-Brazil | x | |
| CCMUSA | Câmara de Comércio Moçambique-EUA | Chamber of Commerce Mozambique-US | x | |
| AMA | Associação Moçambicana de Avicultores | Mozambican Association of Poultry Farmers | x | |
| AMOPAO | Associação Moçambicana dos Panificadores | Mozambican Association of Bakers | x | x |
| ADCMN | Associação para Desenvolvinmento de Marien Ngouabi | Association for Desenvolvinmento Marien Ngouabi | | |
| APEMA | Associação Provincial dos Empreiteiros de Manica | Provincial Association of Contractors of Manica | | |
| MANICA | Associação das Moageiras de Manica | Association of Millers Manica | | |
| AIMO-FI | Associação Industríal de Moçambique - Federação das Industríais | Industrial Association of Mozambique - Federation of Industry | x | |
| CTA | Confederação das Associações Económicas de Moçambique | Confederation of Business Associations of Mozambique | | |

Note: The last column indicates whether the association is member of CTA (i.e., registered association): x indicates that the association is registered under CTA. It was not possible to obtain the full names of five informal business associations: ADEMA, API, ASSEMA, MABINDZO, and AZOFI.

Table A3: Location of headquarters

| Acronym | Location | Location proxy | Membership restricted | ISIC 2-digit code | Description of ISIC 2-digit code |
|---|---|---|---|---|---|
| ACB | * | | No | | |
| ACIS | * | | No | | |
| ACIANA | * | | No | | |
| ASSOCIET | ** | Bus terminal in Tete city | No | | |
| AIOPA | * | | Yes | 15, 24 | Food products and beverages; chemicals products |
| AIGM | * | | Yes | 21, 22 | Paper, paper products, printing, and publishing |
| AEM | * | | No | | |
| ACTIVA | * | | Yes | | Female owners |
| CCM | * | | No | | |
| CCMOBRA | * | | No | | |
| CCMUSA | * | | No | | |
| AMA | * | | Yes | 15 | Food products and beverages |
| AMOPAO | * | | Yes | | Food products and beverages |
| ADCMN | ** | Xai-Xai Public Library | Yes | 17, 18, 19 | Textiles, apparel etc., and dressing leather |
| ADEMA | ** | Engineering Faculty in Chimoio | No | | |
| APEMA | ** | Engineering Faculty in Chimoio | Yes | 36 | Furniture, jewelery, etc. |
| API | *** | The Cathedral of Beira | No | | |
| ASSEMA | ** | Engineering Faculty in Chimoio | Yes | 26 | Non-metallic mineral products |
| MABINDZO | ** | The Central Hospital of Maputo | Yes | 15 | Food products and beverages |
| AZOFI | *** | The Cathedral of Beira | Yes | 21, 22 | Paper, paper products, printing and publishing |
| MANICA | ** | Engineering Faculty in Chimoio | No | | |
| AIMO-FI | * | | No | | |
| CTA | * | | No | | |

Note: * indicates that the headquarters location is exact, ** indicates that the location is based on the survey answer, while *** indicates that the location is equal to the location of the firm that reported being a member of that business association (information missing in the survey answers). Column (2) reports location proxies for non-registered business associations reported in the survey.

Table A4: Heterogeneity: geographical distance

| Panel A | Employ staff exclusively for R&D | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| *Co-members* | 0.227* | (0.124) | 0.225 | (0.248) | 0.489* | (0.321) |
| Same sector | -0.003 | (0.009) | -0.011 | (0.009) | -0.010 | (0.009) |
| Distance (log) | -0.006*** | (0.002) | -0.004 | (0.003) | -0.004 | (0.003) |
| Co-members*distance | -0.030 | (0.028) | -0.031 | (0.030) | -0.004 | (0.034) |
| IMR1 | | | 0.008 | (0.100) | -0.103 | (0.133) |
| IMR2 | | | 0.753** | (0.373) | 0.751** | (0.374) |
| **Panel B** | **Introduced a new product** | | | | | |
| | 1 | | 2 | | 3 | |
| *Co-members* | 0.088 | (0.131) | 0.108 | (0.246) | 0.280 | (0.318) |
| Same sector | -0.013* | (0.012) | -0.011 | (0.012) | -0.011 | (0.012) |
| Distance (log) | -0.006*** | (0.002) | -0.006** | (0.003) | -0.006** | (0.003) |
| Co-members*distance | -0.015 | (0.041) | -0.015 | (0.041) | 0.007 | (0.050) |
| IMR1 | | | -0.011 | (0.096) | -0.088 | (0.129) |
| IMR2 | | | -0.137 | (0.278) | -0.138 | (0.277) |
| **Panel C** | **Improved existing products** | | | | | |
| | 1 | | 2 | | 3 | |
| *Co-members* | -0.029 | (0.131) | 0.094 | (0.247) | 0.096 | (0.331) |
| Same sector | 0.002 | (0.006) | 0.006 | (0.007) | 0.006 | (0.006) |
| Distance (log) | 0.006** | (0.003) | 0.006** | (0.003) | 0.006** | (0.003) |
| Co-members*distance | -0.003 | (0.037) | -0.001 | (0.037) | -0.009 | (0.045) |
| IMR1 | | | -0.063 | (0.099) | -0.051 | (0.129) |
| IMR2 | | | -0.370 | (0.243) | -0.369 | (0.250) |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. A constant and control variables are included in all specifications. In addition, columns (3) and (6) include de-meaned interactions times co-membership and both BA-members, respectively. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Table A5: Heterogeneity: geographical distance

| Panel A | Introduced a new product or improved existing product | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| *Co-members* | 0.099 | (0.192) | 0.247 | (0.351) | 0.420 | (0.505) |
| Same sector | -0.011 | (0.014) | -0.008 | (0.014) | -0.007 | (0.014) |
| Distance (log) | 0.002 | (0.004) | 0.002 | (0.004) | 0.002 | (0.004) |
| Co-members*distance | -0.020 | (0.065) | -0.018 | (0.065) | -0.005 | (0.074) |
| IMR1 | | | -0.075 | (0.138) | -0.139 | (0.203) |
| IMR2 | | | -0.376 | (0.327) | -0.376 | (0.327) |

| Panel B | Introduced a new product, improved existing product or introduced a new technology | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| *Co-members* | -0.034 | (0.136) | 0.016 | (0.260) | 0.034 | (0.330) |
| Same sector | 0.004 | (0.007) | 0.010 | (0.008) | 0.010 | (0.008) |
| Distance (log) | 0.010** | (0.004) | 0.008** | (0.004) | 0.008** | (0.004) |
| Co-members*distance | 0.002 | (0.039) | 0.003 | (0.039) | -0.007 | (0.046) |
| IMR1 | | | -0.030 | (0.100) | -0.026 | (0.125) |
| IMR2 | | | -0.613** | (0.306) | -0.612** | (0.316) |

| Panel C | Introduced a new technology | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| *Co-members* | 0.046 | (0.132) | 0.151 | (0.261) | 0.154 | (0.354) |
| Same sector | -0.042*** | (0.015) | -0.049*** | (0.015) | -0.049*** | (0.015) |
| Distance (log) | -0.008** | (0.003) | -0.006** | (0.003) | -0.006** | (0.003) |
| Co-members*distance | 0.016 | (0.036) | 0.016 | (0.037) | 0.018 | (0.042) |
| IMR1 | | | -0.045 | (0.100) | -0.044 | (0.135) |
| IMR2 | | | 0.645** | (0.303) | 0.645** | (0.299) |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. A constant and control variables are included in all specifications. In addition, columns (3) and (6) include de-meaned interactions times co-membership and both BA-members, respectively. The number in parenthesis are bootstrapped standard errors that are robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent level, respectively.

Table A6: Heterogeneity: geographical distance

| Panel A | Tax examptions from VAT on domestic purchases | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| *Co-members* | 0.053 | (0.129) | -0.208 | (0.232) | -0.267 | (0.313) |
| Same sector | 0.013 | (0.011) | 0.011 | (0.011) | 0.011 | (0.011) |
| Distance (log) | -0.007** | (0.003) | -0.006** | (0.003) | -0.006** | (0.003) |
| Co-members*distance | -0.008 | (0.033) | -0.012 | (0.033) | -0.022 | (0.037) |
| IMR1 | | | 0.127 | (0.095) | 0.151 | (0.125) |
| IMR2 | | | 0.200 | (0.350) | 0.200 | (0.347) |
| **Panel B** | **Tax examptions from VAT on imports** | | | | | |
| | 1 | | 2 | | 3 | |
| *Co-members* | -0.018 | (0.141) | 0.251 | (0.235) | 0.229 | (0.331) |
| Same sector | 0.012 | (0.011) | 0.003 | (0.011) | 0.003 | (0.011) |
| Distance (log) | -0.006** | (0.003) | -0.004 | (0.003) | -0.004 | (0.003) |
| Co-members*distance | 0.033 | (0.035) | 0.036 | (0.036) | 0.038 | (0.043) |
| IMR1 | | | -0.122 | (0.087) | -0.116 | (0.119) |
| IMR2 | | | 0.798** | (0.321) | 0.797** | (0.313) |
| **Panel C** | **Tax examptions on customs duties** | | | | | |
| | 1 | | 2 | | 3 | |
| *Co-members* | -0.109 | (0.135) | -0.137 | (0.259) | 0.066 | (0.396) |
| Same sector | 0.001 | (0.011) | 0.000 | (0.012) | 0.000 | (0.012) |
| Distance (log) | -0.011*** | (0.003) | -0.011*** | (0.003) | -0.011*** | (0.003) |
| Co-members*distance | 0.037 | (0.041) | 0.037 | (0.042) | 0.052 | (0.052) |
| IMR1 | | | 0.014 | (0.107) | -0.068 | (0.157) |
| IMR2 | | | 0.058 | (0.324) | 0.057 | (0.328) |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. A constant and control variables are included in all specifications. In addition, columns (3) and (6) include de-meaned interactions times co-membership and both BA-members, respectively. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Table A7: Heterogeneity: same sector

| Panel A | Employ staff exclusively for R&D | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| *Co-members* | 0.269*** | (0.102) | 0.551* | (0.285) | 0.510 | (0.446) |
| Same sector | -0.002 | (0.009) | -0.010 | (0.009) | -0.010 | (0.009) |
| Distance (log) | -0.006*** | (0.002) | -0.004 | (0.003) | -0.004 | (0.003) |
| Co-members*sector | -0.241* | (0.141) | -0.298* | (0.163) | -0.275 | (0.189) |
| IMR1 | | | -0.110* | (0.102) | -0.103 | (0.133) |
| IMR2 | | | 0.752* | (0.376) | 0.751* | (0.375) |

| Panel B | Introduced a new product | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| *Co-members* | 0.102 | (0.078) | 0.276 | (0.278) | 0.341 | (0.451) |
| Same sector | -0.012 | (0.012) | -0.011 | (0.012) | -0.011 | (0.012) |
| Distance (log) | -0.006*** | (0.002) | -0.006** | (0.003) | -0.006** | (0.003) |
| Co-members*sector | -0.109 | (0.121) | -0.151* | (0.146) | -0.164 | (0.176) |
| IMR1 | | | -0.071 | (0.103) | -0.088 | (0.129) |
| IMR2 | | | -0.138 | (0.279) | -0.138 | (0.278) |

| Panel C | Improved existing products | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| *Co-members* | -0.094 | (0.109) | -0.012 | (0.330) | 0.036 | (0.432) |
| Same sector | 0.002 | (0.006) | 0.006 | (0.006) | 0.006 | (0.006) |
| Distance (log) | 0.006** | (0.003) | 0.006** | (0.003) | 0.006** | (0.003) |
| Co-members*sector | 0.098 | (0.125) | 0.076 | (0.161) | 0.060 | (0.172) |
| IMR1 | | | -0.034 | (0.114) | -0.051 | (0.129) |
| IMR2 | | | -0.369 | (0.250) | -0.369 | (0.250) |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. A constant and control variables are included in all specifications. In addition, columns (3) and (6) include de-meaned interactions times co-membership and both BA-members, respectively. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Table A8: Heterogeneity: same sector

| Panel A | Introduced a new product or improved existing product | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| *Co-members* | 0.043 | (0.122) | 0.304 | (0.455) | 0.409 | (0.688) |
| Same sector | -0.011 | (0.014) | -0.007 | (0.014) | -0.007 | (0.014) |
| Distance (log) | 0.002 | (0.004) | 0.001 | (0.004) | 0.002 | (0.004) |
| Co-members*sector | -0.012 | (0.182) | -0.076 | (0.234) | -0.101 | (0.266) |
| IMR1 | | | -0.107 | (0.166) | -0.139 | (0.203) |
| IMR2 | | | -0.377 | (0.328) | -0.376 | (0.327) |

| Panel B | Introduced a new product, improved existing product or introduced a new technology | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| *Co-members* | -0.076 | (0.108) | -0.083 | (0.320) | -0.010 | (0.420) |
| Same sector | 0.003 | (0.007) | 0.010 | (0.008) | 0.010 | (0.008) |
| Distance (log) | 0.010** | (0.004) | 0.008** | (0.004) | 0.008** | (0.004) |
| Co-members*sector | 0.083 | (0.134) | 0.078 | (0.167) | 0.054 | (0.183) |
| IMR1 | | | 0.000 | (0.110) | -0.026 | (0.125) |
| IMR2 | | | -0.612* | (0.316) | -0.612* | (0.316) |

| Panel C | Introduced a new technology | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| *Co-members* | 0.065 | (0.110) | 0.136 | (0.321) | 0.250 | (0.460) |
| Same sector | -0.042*** | (0.015) | -0.049*** | (0.015) | -0.049*** | (0.015) |
| Distance (log) | -0.007** | (0.003) | -0.006** | (0.003) | -0.006** | (0.003) |
| Co-members*sector | 0.053 | (0.144) | 0.043 | (0.164) | 0.018 | (0.185) |
| IMR1 | | | -0.025 | (0.112) | -0.044 | (0.135) |
| IMR2 | | | 0.646** | (0.299) | 0.645** | (0.299) |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. A constant and control variables are included in all specifications. In addition, columns (3) and (6) include de-meaned interactions times co-membership and both BA-members, respectively. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Table A9: Heterogeneity: same sector

| Panel A | Tax exemptions from VAT on domestic purchases | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| *Co-members* | 0.051 | (0.113) | -0.295 | (0.208) | -0.407 | (0.431) |
| Same sector | 0.013 | (0.010) | 0.011 | (0.011) | 0.011 | (0.011) |
| Distance (log) | -0.007** | (0.003) | -0.006** | (0.003) | -0.006** | (0.003) |
| Co-members*sector | -0.042 | (0.154) | 0.040 | (0.171) | 0.140 | (0.197) |
| IMR1 | | | 0.140 | (0.107) | 0.151 | (0.125) |
| IMR2 | | | 0.200 | (0.348) | 0.200 | (0.348) |
| **Panel B** | **Tax exemptions from VAT on imports** | | | | | |
| | 1 | | 2 | | 3 | |
| *Co-members* | 0.040 | (0.108) | 0.306* | (0.306) | 0.437 | (0.435) |
| Same sector | 0.011 | (0.011) | 0.003 | (0.011) | 0.003 | (0.011) |
| Distance (log) | -0.006** | (0.003) | -0.004 | (0.003) | -0.004 | (0.003) |
| Co-members*sector | 0.082 | (0.135) | 0.029 | (0.157) | -0.004 | (0.179) |
| IMR1 | | | -0.104 | (0.098) | -0.116 | (0.119) |
| IMR2 | | | 0.798** | (0.313) | 0.797** | (0.313) |
| **Panel C** | **Tax exemptions on customs duties** | | | | | |
| | 1 | | 2 | | 3 | |
| *Co-members* | 0.017 | (0.117) | -0.030 | (0.381) | 0.363 | (0.569) |
| Same sector | 0.001 | (0.011) | 0.000 | (0.012) | 0.000 | (0.012) |
| Distance (log) | -0.011*** | (0.003) | -0.011*** | (0.003) | -0.011*** | (0.003) |
| Co-members*sector | -0.018 | (0.160) | -0.006 | (0.201) | -0.087 | (0.217) |
| IMR1 | | | 0.019 | (0.128) | -0.068 | (0.157) |
| IMR2 | | | 0.059 | (0.329) | 0.057 | (0.329) |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. A constant and control variables are included in all specifications. In addition, columns (3) and (6) include de-meaned interactions times co-membership and both BA-members, respectively.Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance on a 10, 5, and 1 percent levels, respectively.

Table A10: Diffusion: co-membership

| | Employ staff exclusively for R&D | | | | | | Introduced a new product | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | |
| Co-membership | 0.133* | (0.074) | 0.141 | (0.239) | 0.468* | (0.282) | 0.040 | (0.065) | 0.068 | (0.231) | 0.317 | (0.422) |
| Same sector | -0.003 | (0.009) | -0.011 | (0.009) | -0.010 | (0.009) | -0.013 | (0.012) | -0.011 | (0.012) | -0.011 | (0.012) |
| Geographical distance (log) | -0.006*** | (0.002) | -0.004*** | (0.002) | -0.004** | (0.002) | -0.006*** | (0.002) | -0.006*** | (0.002) | -0.006*** | (0.002) |
| Owners nationality differ | 0.038 | (0.036) | 0.035 | (0.036) | 0.035 | (0.036) | 0.013 | (0.027) | 0.014 | (0.027) | 0.014 | (0.027) |
| Owners gender differ | 0.003 | (0.025) | 0.002 | (0.025) | 0.003 | (0.025) | -0.016 | (0.028) | -0.016 | (0.028) | -0.016 | (0.028) |
| *Absolute difference:* | | | | | | | | | | | | |
| Educational level | -0.006 | (0.004) | -0.005 | (0.004) | -0.005 | (0.004) | 0.002 | (0.005) | 0.002 | (0.005) | 0.002 | (0.005) |
| Firm size (log) | 0.028** | (0.012) | 0.029** | (0.013) | 0.029** | (0.013) | 0.011 | (0.010) | 0.011 | (0.011) | 0.011 | (0.011) |
| Firm age | 0.003*** | (0.001) | 0.003*** | (0.001) | 0.003*** | (0.001) | -0.001 | (0.001) | -0.001 | (0.001) | -0.001 | (0.001) |
| *De-meaned difference:* | | | | | | | | | | | | |
| Same sector | | | | | -0.275 | (0.190) | | | | | -0.164 | (0.176) |
| Geographical distance (log) | | | | | -0.004 | (0.035) | | | | | 0.007 | (0.048) |
| Owners nationality differ | | | | | -0.074** | (0.035) | | | | | -0.016 | (0.046) |
| Owners gender differ | | | | | -0.168** | (0.082) | | | | | -0.020 | (0.105) |
| Educational level | | | | | -0.026 | (0.088) | | | | | -0.041*** | (0.004) |
| Firm size (log) | | | | | -0.034 | (0.107) | | | | | 0.048 | (0.137) |
| Firm age | | | | | -0.002 | (0.004) | | | | | 0.000 | (0.120) |
| IMR1 | | | 0.002 | (0.099) | -0.103 | (0.133) | | | -0.014 | (0.096) | -0.088 | (0.128) |
| IMR2 | | | 0.753** | (0.370) | 0.751** | (0.370) | | | -0.137 | (0.276) | -0.138 | (0.279) |
| Observations | 140,715 | | 140,715 | | 140,715 | | 140,715 | | 140,715 | | 140,715 | |
| R-squared | 0.043 | | 0.046 | | 0.047 | | 0.002 | | 0.002 | | 0.002 | |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. A constant and control variables are included in all specifications. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Table A11: Diffusion: co-membership

| | Improved existing products | | | | | | Introduced a new product or improved existing product | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | |
| Co-membership | -0.038 | (0.066) | 0.092 | (0.223) | 0.045 | (0.416) | 0.036 | (0.091) | 0.199 | (0.318) | 0.394 | (0.654) |
| Same sector | 0.002 | (0.006) | 0.006 | (0.007) | 0.006 | (0.007) | -0.011 | (0.014) | -0.008 | (0.014) | -0.007 | (0.014) |
| Geographical distance (log) | 0.006** | (0.003) | 0.006** | (0.003) | 0.006** | (0.003) | 0.002 | (0.004) | 0.001 | (0.004) | 0.002 | (0.004) |
| Owners nationality differ | -0.009 | (0.006) | -0.008 | (0.006) | -0.008 | (0.006) | 0.010 | (0.027) | 0.011 | (0.028) | 0.011 | (0.028) |
| Owners gender differ | 0.000 | (0.004) | 0.001 | (0.004) | 0.001 | (0.004) | -0.021 | (0.028) | -0.021 | (0.028) | -0.021 | (0.028) |
| *Absolute difference:* | | | | | | | | | | | | |
| Educational level | 0.004 | (0.004) | 0.003 | (0.004) | 0.003 | (0.004) | 0.008 | (0.006) | 0.008 | (0.007) | 0.007 | (0.007) |
| Firm size (log) | 0.011 | (0.007) | 0.010* | (0.006) | 0.010* | (0.006) | 0.019 | (0.014) | 0.019 | (0.014) | 0.019 | (0.014) |
| Firm age | -0.000 | (0.000) | -0.000 | (0.000) | -0.000 | (0.000) | -0.001 | (0.001) | -0.000 | (0.001) | -0.000 | (0.001) |
| *De-meaned difference:* | | | | | | | | | | | | |
| Same sector | | | | | 0.060 | (0.178) | | | | | -0.101 | (0.269) |
| Geographical distance (log) | | | | | -0.009 | (0.046) | | | | | -0.005 | (0.074) |
| Owners nationality differ | | | | | -0.022 | (0.107) | | | | | -0.042 | (0.176) |
| Owners gender differ | | | | | 0.061 | (0.146) | | | | | 0.046 | (0.214) |
| Educational level | | | | | 0.089* | (0.069) | | | | | 0.045 | (0.090) |
| Firm size (log) | | | | | 0.019 | (0.076) | | | | | 0.069 | (0.143) |
| Firm age | | | | | 0.004 | (0.005) | | | | | 0.003 | (0.007) |
| IMR1 | | | -0.063 | (0.096) | -0.051 | (0.129) | | | -0.078 | (0.136) | -0.139 | (0.199) |
| IMR2 | | | -0.370 | (0.258) | -0.369 | (0.258) | | | -0.376 | (0.331) | -0.376 | (0.331) |
| Observations | 140,715 | | 140,715 | | 140,715 | | 140,715 | | 140,715 | | 140,715 | |
| R-squared | 0.001 | | 0.001 | | 0.001 | | 0.001 | | 0.001 | | 0.001 | |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. A constant and control variables are included in all specifications. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Table A12: Diffusion: co-membership

| | Introduced a new product, improved existing product or introduced a new technology | | | | | | Introduced a new technology | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | |
| Co-membership | -0.029 | (0.068) | 0.025 | (0.228) | -0.002 | (0.412) | 0.095 | (0.069) | 0.195 | (0.234) | 0.253 | (0.470) |
| Same sector | 0.004 | (0.007) | 0.010 | (0.008) | 0.010 | (0.008) | -0.042*** | (0.015) | -0.049*** | (0.015) | -0.049*** | (0.015) |
| Geographical distance (log) | 0.010** | (0.004) | 0.008** | (0.004) | 0.008** | (0.004) | -0.007** | (0.003) | -0.006** | (0.003) | -0.006** | (0.003) |
| Owners nationality differ | -0.008** | (0.004) | -0.005 | (0.004) | -0.005 | (0.004) | 0.044*** | (0.006) | 0.042*** | (0.006) | 0.042*** | (0.006) |
| Owners gender differ | 0.007 | (0.006) | 0.007 | (0.006) | 0.007 | (0.006) | 0.016 | (0.014) | 0.015 | (0.014) | 0.015 | (0.014) |
| *Absolute difference:* | | | | | | | | | | | | |
| Educational level | 0.004*** | (0.000) | 0.003*** | (0.000) | 0.003*** | (0.000) | 0.003*** | (0.001) | 0.003*** | (0.001) | 0.003*** | (0.001) |
| Firm size (log) | 0.004 | (0.010) | 0.003 | (0.010) | 0.003 | (0.010) | 0.036 | (0.037) | 0.037 | (0.037) | 0.037 | (0.037) |
| Firm age | 0.000 | (0.011) | 0.000 | (0.011) | 0.000 | (0.011) | 0.000 | (0.034) | 0.000 | (0.034) | 0.000 | (0.034) |
| *De-meaned difference:* | | | | | | | | | | | | |
| Same sector | | | | | 0.054 | (0.191) | | | | | 0.018 | (0.192) |
| Geographical distance (log) | | | | | -0.007 | (0.047) | | | | | 0.018 | (0.043) |
| Owners nationality differ | | | | | -0.042 | (0.111) | | | | | -0.003 | (0.151) |
| Owners gender differ | | | | | -0.014 | (0.147) | | | | | 0.042 | (0.152) |
| Educational level | | | | | 0.062 | (0.070) | | | | | -0.008 | (0.060) |
| Firm size (log) | | | | | 0.035 | (0.077) | | | | | 0.024 | (0.111) |
| Firm age | | | | | 0.006 | (0.006) | | | | | -0.003 | (0.004) |
| IMR1 | | | -0.029 | (0.098) | -0.026 | (0.130) | | | -0.042 | (0.099) | -0.044 | (0.145) |
| IMR2 | | | -0.613* | (0.314) | -0.612* | (0.314) | | | 0.646** | (0.287) | 0.645** | (0.287) |
| Observations | 140,715 | | 140,715 | | 140,715 | | 140,715 | | 140,715 | | 140,715 | |
| R-squared | 0.002 | | 0.002 | | 0.002 | | 0.012 | | 0.013 | | 0.013 | |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. A constant and control variables are included in all specifications. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Table A13: Diffusion: co-membership

| | Tax exemptions from VAT on imports | | | | | | Tax exemptions from VAT on domestic purchases | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | |
| Co-membership | 0.027 | (0.080) | -0.240 | (0.223) | -0.386 | (0.406) | 0.086 | (0.078) | 0.347 | (0.229) | 0.436 | (0.418) |
| Same sector | 0.013 | (0.011) | 0.011 | (0.011) | 0.011 | (0.011) | 0.012 | (0.012) | 0.003 | (0.012) | 0.003 | (0.012) |
| Geographical distance (log) | -0.007*** | (0.003) | -0.006** | (0.003) | -0.006** | (0.003) | -0.006** | (0.003) | -0.004 | (0.003) | -0.004 | (0.003) |
| Owners nationality differ | 0.051 | (0.039) | 0.050 | (0.040) | 0.050 | (0.040) | 0.028 | (0.036) | 0.025 | (0.036) | 0.025 | (0.036) |
| Owners gender differ | 0.054 | (0.044) | 0.054 | (0.044) | 0.054 | (0.044) | 0.066 | (0.042) | 0.065 | (0.041) | 0.065 | (0.042) |
| *Absolute difference:* | | | | | | | | | | | | |
| Educational level | -0.009* | (0.005) | -0.009* | (0.005) | -0.009* | (0.005) | -0.010* | (0.006) | -0.009 | (0.006) | -0.009 | (0.006) |
| Firm size (log) | 0.017 | (0.012) | 0.017 | (0.012) | 0.017 | (0.012) | 0.020 | (0.014) | 0.022 | (0.014) | 0.022 | (0.014) |
| Firm age | -0.001 | (0.001) | -0.001 | (0.001) | -0.001 | (0.001) | 0.003*** | (0.001) | 0.003*** | (0.001) | 0.003*** | (0.001) |
| *De-meaned difference:* | | | | | | | | | | | | |
| Same sector | | | | | 0.140 | (0.195) | | | | | -0.004 | (0.180) |
| Geographical distance (log) | | | | | -0.022 | (0.038) | | | | | 0.038 | (0.044) |
| Educational level | | | | | 0.031 | (0.040) | | | | | 0.060* | (0.058) |
| Firm size (log) | | | | | -0.219** | (0.087) | | | | | -0.014 | (0.093) |
| Firm age | | | | | -0.020 | (0.003) | | | | | 0.038*** | (0.004) |
| Owners nationality differ | | | | | 0.025 | (0.141) | | | | | -0.061 | (0.143) |
| Owners gender differ | | | | | -0.004 | (0.109) | | | | | -0.004 | (0.167) |
| IMR1 | | | 0.125 | (0.091) | 0.151 | (0.124) | | | -0.115 | (0.091) | -0.116 | (0.123) |
| IMR2 | | | 0.199 | (0.350) | 0.200 | (0.350) | | | 0.798** | (0.322) | 0.797** | (0.322) |
| Observations | 140,715 | | 140,715 | | 140,715 | | 140,715 | | 140,715 | | 140,715 | |
| R-squared | 0.010 | | 0.010 | | 0.010 | | 0.013 | | 0.015 | | 0.015 | |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. A constant and control variables are included in all specifications. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Table A14: Diffusion: co-membership

| | Tax exemptions on customs duties | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| Co-membership | 0.007 | (0.080) | -0.038 | (0.262) | 0.350 | (0.542) |
| Same sector | 0.001 | (0.011) | 0.000 | (0.012) | 0.000 | (0.012) |
| Geographical distance (log) | -0.011*** | (0.003) | -0.011*** | (0.003) | -0.011*** | (0.003) |
| Owners nationality differ | 0.064 | (0.040) | 0.064 | (0.040) | 0.064 | (0.040) |
| Owners gender differ | 0.077* | (0.045) | 0.077* | (0.045) | 0.078* | (0.045) |
| *Absolute difference:* | | | | | | |
| Educational level | -0.007 | (0.006) | -0.007 | (0.006) | -0.007 | (0.006) |
| Firm size (log) | 0.002 | (0.012) | 0.002 | (0.012) | 0.002 | (0.012) |
| Firm age | -0.000 | (0.001) | -0.000 | (0.001) | -0.000 | (0.001) |
| *De-meaned difference:* | | | | | | |
| Same sector | | | | | -0.087 | (0.216) |
| Geographical distance (log) | | | | | 0.052 | (0.053) |
| Educational level | | | | | -0.070* | (0.037) |
| Firm size (log) | | | | | -0.161 | (0.127) |
| Firm age | | | | | -0.055*** | (0.004) |
| Owners nationality differ | | | | | 0.115 | (0.131) |
| Owners gender differ | | | | | -0.001 | (0.104) |
| IMR1 | | | 0.021 | (0.107) | -0.068 | (0.159) |
| IMR2 | | | 0.059 | (0.327) | 0.057 | (0.327) |
| Observations | 140,715 | | 140,715 | | 140,715 | |
| R-squared | 0.012 | | 0.012 | | 0.012 | |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. A constant and control variables are included in all specifications. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Table A15: Diffusion: BA members

| | Employ staff exclusively for R&D | | | | | | Introduced a new product | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | |
| Both BA member | 0.113* | (0.060) | 0.309** | (0.138) | 0.317** | (0.162) | 0.040 | (0.039) | 0.034 | (0.104) | 0.036 | (0.104) |
| Same sector | -0.003 | (0.009) | -0.013 | (0.010) | -0.012 | (0.010) | -0.013 | (0.012) | -0.015 | (0.012) | -0.015 | (0.011) |
| Geographical distance (log) | -0.006*** | (0.002) | -0.006*** | (0.002) | -0.006*** | (0.002) | -0.006*** | (0.002) | -0.006*** | (0.002) | -0.006*** | (0.002) |
| Owners nationality differ | 0.036 | (0.036) | 0.023 | (0.038) | 0.023 | (0.037) | 0.013 | (0.027) | 0.009 | (0.030) | 0.009 | (0.028) |
| Owners gender differ | 0.003 | (0.025) | 0.001 | (0.025) | 0.001 | (0.025) | -0.016 | (0.028) | -0.017 | (0.027) | -0.017 | (0.029) |
| *Absolute difference:* | | | | | | | | | | | | |
| Educational level | -0.005 | (0.004) | -0.002 | (0.004) | -0.002 | (0.004) | 0.002 | (0.005) | 0.003 | (0.005) | 0.003 | (0.005) |
| Firm size (log) | 0.028** | (0.012) | 0.026** | (0.013) | 0.026** | (0.012) | 0.011 | (0.010) | 0.010 | (0.011) | 0.010 | (0.010) |
| Firm age | 0.003*** | (0.001) | 0.003*** | (0.001) | 0.003*** | (0.001) | -0.001 | (0.001) | -0.001 | (0.001) | -0.001 | (0.001) |
| *De-meaned difference:* | | | | | | | | | | | | |
| Same sector | | | | | -0.197** | (0.079) | | | | | -0.080 | (0.079) |
| Geographical distance (log) | | | | | -0.027*** | (0.008) | | | | | -0.010 | (0.007) |
| Owners nationality differ | | | | | -0.070 | (0.061) | | | | | 0.005 | (0.064) |
| Owners gender differ | | | | | -0.174*** | (0.033) | | | | | -0.024 | (0.058) |
| Educational level | | | | | -0.029*** | (0.011) | | | | | -0.046*** | (0.017) |
| Firm size (log) | | | | | -0.053** | (0.023) | | | | | 0.021 | (0.021) |
| Firm age | | | | | -0.001 | (0.002) | | | | | 0.000 | (0.001) |
| IMR1 | | | -0.094 | (0.067) | -0.098 | (0.071) | | | 0.007 | (0.051) | 0.006 | (0.056) |
| IMR2 | | | 0.532 *** | (0.187) | 0.531*** | (0.180) | | | 0.151 | (0.125) | 0.151 | (0.128) |
| Observations | 140,715 | | 140,715 | | 140,715 | | 140,715 | | 140,715 | | 140,715 | |
| R-squared | 0.046 | | 0.066 | | 0.066 | | 0.002 | | 0.003 | | 0.003 | |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. BA is short for business association. A constant and control variables are included in all specifications. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Table A16: Diffusion: BA members

| | Improved existing products | | | | | | Introduced a new product or improved existing product | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | |
| Both BA member | -0.029 | (0.031) | -0.135 | (0.100) | -0.138 | (0.095) | 0.041 | (0.038) | -0.069 | (0.121) | -0.071 | (0.125) |
| Same sector | 0.003 | (0.006) | 0.003 | (0.007) | 0.003 | (0.006) | -0.012 | (0.015) | -0.015 | (0.015) | -0.015 | (0.014) |
| Geographical distance (log) | 0.006** | (0.003) | 0.006** | (0.003) | 0.006** | (0.003) | 0.002 | (0.004) | 0.003 | (0.004) | 0.003 | (0.004) |
| Owners nationality differ | -0.009 | (0.006) | -0.007 | (0.006) | -0.007 | (0.007) | 0.009 | (0.027) | 0.004 | (0.028) | 0.004 | (0.028) |
| Owners gender differ | 0.001 | (0.004) | 0.000 | (0.004) | 0.000 | (0.004) | -0.021 | (0.028) | -0.022 | (0.028) | -0.022 | (0.028) |
| *Absolute difference:* | | | | | | | | | | | | |
| Educational level | 0.003 | (0.004) | 0.003 | (0.004) | 0.003 | (0.004) | 0.008 | (0.006) | 0.009 | (0.007) | 0.009 | (0.006) |
| Firm size (log) | 0.011 | (0.007) | 0.010 | (0.007) | 0.010 | (0.006) | 0.019 | (0.014) | 0.018 | (0.014) | 0.018 | (0.014) |
| Firm age | -0.000 | (0.000) | 0.000 | (0.000) | -0.000 | (0.000) | -0.001 | (0.001) | -0.001 | (0.001) | -0.001 | (0.001) |
| *De-meaned difference:* | | | | | | | | | | | | |
| Same sector | | | | | 0.077 | (0.067) | | | | | -0.000 | (0.119) |
| Geographical distance (log) | | | | | -0.001 | (0.013) | | | | | -0.012 | (0.014) |
| Owners nationality differ | | | | | -0.003 | (0.024) | | | | | -0.000 | (0.063) |
| Owners gender differ | | | | | 0.073 | (0.085) | | | | | 0.055 | (0.076) |
| Educational level | | | | | 0.086** | (0.035) | | | | | 0.038 | (0.043) |
| Firm size (log) | | | | | 0.010 | (0.034) | | | | | 0.033 | (0.042) |
| Firm age | | | | | 0.003 | (0.002) | | | | | 0.003 | (0.002) |
| IMR1 | | | 0.058 | (0.048) | 0.059 | (0.041) | | | 0.066 | (0.059) | 0.066 | (0.063) |
| IMR2 | | | -0.028 | (0.063) | -0.027 | (0.063) | | | 0.201 | (0.135) | 0.201 | (0.135) |
| Observations | 140,715 | | 140,715 | | 140,715 | | 140,715 | | 140,715 | | 140,715 | |
| R-squared | 0.001 | | 0.001 | | 0.001 | | 0.001 | | 0.002 | | 0.002 | |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. BA is short for business association. A constant and control variables are included in all specifications. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Table A17: Diffusion: BA members

| | Introduced a new product, improved existing product or introduced a new technology | | | | | | Introduced a new technology | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | |
| Both BA member | -0.046* | (0.037) | -0.167* | (0.100) | -0.174* | (0.105) | 0.093* | (0.037) | 0.096 | (0.099) | 0.093 | (0.110) |
| Same sector | 0.004 | (0.007) | 0.008 | (0.008) | 0.007 | (0.008) | -0.042*** | (0.015) | -0.050*** | (0.015) | -0.050*** | (0.015) |
| Geographical distance (log) | 0.010** | (0.004) | 0.010** | (0.004) | 0.010** | (0.004) | -0.008*** | (0.003) | -0.007** | (0.003) | -0.007** | (0.003) |
| Owners nationality differ | -0.007 | (0.010) | -0.001 | (0.010) | -0.001 | (0.010) | 0.043 | (0.037) | 0.031 | (0.037) | 0.031 | (0.038) |
| Owners gender differ | 0.007 | (0.011) | 0.007 | (0.011) | 0.007 | (0.011) | 0.015 | (0.034) | 0.013 | (0.034) | 0.013 | (0.034) |
| *Absolute difference:* | | | | | | | | | | | | |
| Educational level | 0.003 | (0.004) | 0.002 | (0.004) | 0.002 | (0.004) | 0.003 | (0.006) | 0.006 | (0.006) | 0.006 | (0.006) |
| Firm size (log) | 0.004 | (0.006) | 0.005 | (0.006) | 0.005 | (0.006) | 0.036** | (0.014) | 0.034** | (0.014) | 0.034** | (0.014) |
| Firm age | 0.000* | (0.000) | 0.000 | (0.000) | 0.000 | (0.000) | 0.000 | (0.001) | -0.000 | (0.001) | -0.000 | (0.001) |
| *De-meaned difference:* | | | | | | | | | | | | |
| Same sector | | | | | 0.080* | (0.047) | | | | | 0.070 | (0.060) |
| Geographical distance (log) | | | | | -0.005 | (0.014) | | | | | 0.009 | (0.009) |
| Owners nationality differ | | | | | -0.027 | (0.033) | | | | | 0.020 | (0.049) |
| Owners gender differ | | | | | -0.011 | (0.083) | | | | | 0.039 | (0.069) |
| Educational level | | | | | 0.060* | (0.033) | | | | | -0.014 | (0.017) |
| Firm size (log) | | | | | 0.022 | (0.035) | | | | | 0.009 | (0.021) |
| Firm age | | | | | 0.005** | (0.002) | | | | | -0.002 | (0.002) |
| IMR1 | | | 0.061 | (0.051) | 0.063 | (0.045) | | | 0.011 | (0.046) | 0.013 | (0.050) |
| IMR2 | | | -0.213** | (0.099) | -0.212** | (0.099) | | | 0.494*** | (0.133) | 0.494*** | (0.133) |
| Observations | 140,715 | | 140,715 | | 140,715 | | 140,715 | | 140,715 | | 140,715 | |
| R-squared | 0.002 | | 0.003 | | 0.003 | | 0.013 | | 0.018 | | 0.019 | |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. BA is short for business association. A constant and control variables are included in all specifications. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Table A18: Diffusion: BA members

| | Tax exemptions from VAT on imports | | | | | | Tax exemptions from VAT on domestic purchases | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | |
| Both BA members | 0.076 | (0.055) | 0.021 | (0.144) | 0.031 | (0.164) | 0.076** | (0.043) | 0.302*** | (0.108) | 0.301*** | (0.112) |
| Same sector | 0.012 | (0.011) | 0.010 | (0.011) | 0.011 | (0.012) | 0.011 | (0.012) | 0.004 | (0.013) | 0.004 | (0.013) |
| Geographical distance (log) | -0.007** | (0.003) | -0.007** | (0.003) | -0.007** | (0.003) | -0.006** | (0.003) | -0.006** | (0.003) | -0.006** | (0.003) |
| Owners nationality differ | 0.050 | (0.040) | 0.047 | (0.041) | 0.047 | (0.041) | 0.027 | (0.036) | 0.016 | (0.038) | 0.016 | (0.038) |
| Owners gender differ | 0.054 | (0.044) | 0.053 | (0.044) | 0.053 | (0.045) | 0.066 | (0.042) | 0.065 | (0.041) | 0.065 | (0.041) |
| *Absolute difference:* | | | | | | | | | | | | |
| Educational level | -0.009 | (0.005) | -0.008 | (0.005) | -0.008 | (0.005) | -0.009 | (0.006) | -0.007 | (0.006) | -0.007 | (0.006) |
| Firm size (log) | 0.017 | (0.013) | 0.016 | (0.013) | 0.016 | (0.013) | 0.020 | (0.014) | 0.019 | (0.014) | 0.019 | (0.014) |
| Firm age | -0.001 | (0.001) | -0.001 | (0.001) | -0.001 | (0.001) | 0.003** | (0.001) | 0.002** | (0.001) | 0.002** | (0.001) |
| *De-meaned difference* * $b_{ij}$: | | | | | | | | | | | | |
| Same sector | | | | | 0.021 | (0.069) | | | | | 0.059 | (0.049) |
| Geographical distance (log) | | | | | 0.004 | (0.011) | | | | | 0.020** | (0.009) |
| Educational level | | | | | 0.018 | (0.054) | | | | | 0.056 | (0.042) |
| Firm size (log) | | | | | -0.214*** | (0.078) | | | | | -0.015 | (0.101) |
| Firm age | | | | | -0.017 | (0.026) | | | | | 0.037 | (0.032) |
| Owners nationality differ | | | | | 0.065** | (0.027) | | | | | -0.076*** | (0.020) |
| Owners gender differ | | | | | -0.004*** | (0.001) | | | | | -0.004** | (0.002) |
| IMR1 | | | 0.033 | (0.070) | 0.030 | (0.079) | | | -0.113** | (0.050) | -0.112** | (0.054) |
| IMR2 | | | 0.126 | (0.163) | 0.125 | (0.163) | | | 0.418*** | (0.153) | 0.418*** | (0.153) |
| Observations | 140,715 | | 140,715 | | 140,715 | | 140,715 | | 140,715 | | 140,715 | |
| R-squared | 0.011 | | 0.011 | | 0.011 | | 0.014 | | 0.019 | | 0.019 | |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. BA is short for business association. A constant and control variables are included in all specifications. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

Table A19: Diffusion: BA members

| | Tax exemptions on customs duties | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| Both BA members | 0.020 | (0.053) | -0.012 | (0.135) | -0.007 | (0.160) |
| Same sector | 0.001 | (0.011) | -0.001 | (0.012) | -0.001 | (0.012) |
| Geographical distance (log) | -0.011*** | (0.003) | -0.011*** | (0.003) | -0.011*** | (0.003) |
| Owners nationality differ | 0.064 | (0.040) | 0.061 | (0.041) | 0.061 | (0.041) |
| Owners gender differ | 0.077* | (0.045) | 0.077* | (0.045) | 0.077* | (0.046) |
| *Absolute difference:* | | | | | | |
| Educational level | -0.007 | (0.006) | -0.007 | (0.005) | -0.007 | (0.006) |
| Firm size (log) | 0.002 | (0.012) | 0.002 | (0.012) | 0.001 | (0.012) |
| Firm age | -0.000 | (0.001) | -0.000 | (0.001) | -0.000 | (0.001) |
| *De-meaned difference* * $b_{ij}$: | | | | | | |
| Same sector | | | | | 0.042 | (0.069) |
| Geographical distance (log) | | | | | 0.016** | (0.008) |
| Educational level | | | | | -0.044 | (0.069) |
| Firm size (log) | | | | | -0.184** | (0.086) |
| Firm age | | | | | -0.063*** | (0.018) |
| Owners nationality differ | | | | | 0.072*** | (0.027) |
| Owners gender differ | | | | | -0.000 | (0.002) |
| IMR1 | | | 0.020 | (0.066) | 0.019 | (0.080) |
| IMR2 | | | 0.105 | (0.160) | 0.105 | (0.160) |
| Observations | 140,715 | | 140,715 | | 140,715 | |
| R-squared | 0.012 | | 0.012 | | 0.013 | |

Note: Linear probability model. Dependent variables are calculated as $(y_i - y_j)$. BA is short for business association. A constant and control variables are included in all specifications. Standard errors are bootstrapped, and thus robust to heteroskedasticity and cross-observation correlation in the error terms involving the same firms. ***, **, and * indicate significance at the 10, 5, and 1 percent levels, respectively.

# Appendix B

The fieldwork was carried out through semi-structured interviews with the owner. All interviews was carried out between the 13th and 24th of August 2012, and each interview lasted between 1-2 hours. The interview was structured by a list of questions prepared in close connection to the survey answers. The main questions used to structure the interview on the topic of business associations and firms business network are outlined below. To validate the survey data a number of broader questions was also included (not outlined here).

- Why did the firm choose to join a business association?
- In what way does your membership bring benefits to the firm?
- Does the business association help the firm overcome constraints?
- If yes, can you give a concrete example of how the association has assisted you?
- Do you participate in arrangements/meetings/fairs organized by the business association?
  - If yes, how often?
  - If no, what is the main reason for not participating?
- Can the function of the business association be improved in order to better assist your firm, and how?
- Do you ever join forces with other suppliers/producers in order to serve larger customers, whom you would not be able to serve alone?
- Does the presence of business association contribute to a more or less competitive market?
- Is it important to know somebody in the business association in order to join? And what type of connections is needed?
- Do you actively participate in the operation of the business association?
- Would you consider cancelling your membership in the near future?
- Have you recently introduced a new product or improved existing ones?

  - If yes, how did you obtain the information related to the production of this product?

- Do you share or discuss information on new innovations with your business partners?

- In what way is the business network of the firm important for the operation of the firm?

# References

ARCAND, J.-L. AND M. FAFCHAMPS (2012): "Matching in community-based organizations," *Journal of Developing Economics*, 98, 203–219.

AUDRETSCH, D. B. (1998): "Agglomeration and the location of innovative activity," *Oxford Review of Economic Policy*, 14(2), 18–29.

BRÄUTIGAM, D., L. RAKNER AND S. TAYLOR (2002): "Business Associations and Growth Coalitions in Sub-Saharan Africa," *The Journal of Modern African Studies*, 40, 519–547.

DANIDA (2012): "Denmark-Mozambique Partnership Policy 2012-2015," http://mozambique.um.dk/da/ /media/Mozambique/Documents/Content

DEVARAJAN, S., D. DOLLAR AND T. HOLMGREN (2001): *Aid and reform in Africa: Lessons from ten case studies*. Washington, DC: The World Bank.

DI (2013): "Sub-Saharan Africa," http://di.dk/SiteCollectionDocuments/DIBD/Global

DNEAP (2006): "Enterprise Development in Mozambique: Results Based on Manufacturing Surveys Conducted in 2002 and 2006," Discussion paper, National Directorate of Studies and Policy Analysis, Ministry of Planning and Development, Mozambique.

DONER, R. F. AND B. R. SCHNEIDER (2000): "Business Associations and Economic Development: Why Some Associations Contribute More Than Others," *Business and Politics*, 2(3), 261–288.

FAFCHAMPS, M. AND F. GUBERT (2007): "The formation of risk sharing networks," *Journal of Development Economics*, 83, 326–350.

FAFCHAMPS, M. AND S. QUINN (2012): "Networks and Manufacturing Firms in Africa: Initial Results from a Randomised Experiment," Working paper.

FAFCHAMPS, M. AND S. QUINN (2013): "Social Networks and Business Practices: Evidence form a Randomized Experiment with Microentrepreneurs," *Working paper*.

FAFCHAMPS, M. AND M. SÖDERBOM (2014): "Network Proximity and Business Practices in African Manufacturing," *World Bank Economic Review*, 28, 99–129.

FRANCO, A. M. AND D. FILSON (2006): "Spin-outs: knowledge diffusion through employee mobility," *RAND Journal of Economics*, 37(4), 841–860.

GELB, A., C. J. MEYER AND V. RAMACHANDRAN (2014): "Development as diffusion: Manufacturing productivity and sub-Saharan Africa's missing middle," WIDER Working Paper 2014/042.

GOLDSMITH, A. A. (2002): "Business Associations and Better Governance in Africa," *Public Administration and Development*, 22, 39–49.

Heilbrunn, J. R. (1997): "Commerce, Politics, and Business Associations in Benin and Togo," *Comparative Politics*, 29, 473–492.

Heilman, B. and J. Lucas (1997): "A Social Movement for African Capitalism? A Comparison of Business Associations in Two African Cities," *African Studies Review*, 40, 141–171.

ICA (2003): "Mozambique Investment Climate Assessment," Discussion paper, the World Bank, Regional Program for Enterprise Development (RPED), Africa Finance and Private Sector (AFTFP).

——— (2009): "Mozambique Investment Climate Assessment," Discussion paper, the World Bank, Regional Program for Enterprise Development (RPED), Africa Finance and Private Sector (AFTFP).

IIM (2012): "2012 Survey of Mozambican Manufacturing Firms," Discussion paper, National Directorate of Studies and Policy Analysis, Ministry of Planning and Development, Mozambiaque.

Kraay, A., I. Soloaga and J. Tybout (2002): "Product Quality, Productive Efficiency, and International Technology Diffusion: Evidence from Plant-Level Panel Data," Policy Research Working Paper Series 2759. Washington D.C: The World Bank.

Lucas, J. (1994): "The State, Civil Society and Regional Elites: A Study of Three Associations in Kano, Nigeria," *African Affairs*, 93, 21–38.

McMillan, J. and C. Woodruff (2000): "Private Order under Dysfunctional Public Order," paper presented for conference on Private and Public Ordering of Commercial Transactions, University of Michigan Law School, Feb. 11-12.

Moore, M. and L. Hamalai (1993): "Economic Liberalization, Political Pluralism and Business Associations in Development Countries," *World Development*, 21, 1895–1912.

Muendler, M.-A., J. Rauch and O. Tocoiand (2012): "Employee Spinoff and Other Entrants: Evidence from Brazil," *International Journal of Industrial Organization*, 30(5), 447–458.

Nadvi, K. (1999): "Facing the new competition: Business associations in developing country industrial clusters," Discussion paper 103/1999. International Institute for Labour Studies, Geneva.

Olson, M. (1965): *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press.

Sakar, S. (2004): "Mozambique: Creating a Productive Wheel in a Competitive Cart," Discussion paper, Technical repoort by Nathan Associates Inc.

USAID (2008): “Improving Institutional Management and Sustainability of the Confederation of Mozambican Business Associations (CTA),” Discussion paper.