Risk Preferences and Risk Pooling in Networks: Theory and Evidence from Community Detection in Ghana

By Daniel Putman*

Click Here For Most Current Version

Do individuals share risk with others who have similar risk preferences or different risk preferences? That is, do individuals assortatively match on risk preferences? I investigate this question in both risk sharing networks and risk pooling communities. Using risk sharing network data in Ghana, I construct a bilateral risk sharing network and import community detection (clustering) algorithms from network science to partition this risk sharing network into risk pooling communities. I estimate that individuals prefer to assortatively match on risk preferences in risk sharing networks. While assortative matching holds in risk pooling communities, the magnitude is reduced as compared to earlier estimates. In other words, an individual's risk pooling community features more diverse risk preferences than their network neighborhood. How does this impact the welfare of agents in the network? I build a theoretical model of covariate risk pooling with heterogeneous risk aversion. Constructing a planners problem, I compare the actual allocation of agents in communities to three benchmarks: optimal risk pooling, no covariate risk pooling, and the "preferred" diversity of risk preferences as implied by network connections. While the actual allocation of agents to communities is less diverse than optimal risk pooling, it lies closer to optimal than to the worst case scenario.

JEL: O12, 016, 017, L14
Keywords: Risk Pooling, Networks, Assortative Matching, Risk Preferences, Insurance

* University of California, Davis Ag. & Resource Economics. This paper has benefited immensely from conversations with Travis Lybbert, Ashish Shenoy, Laura Meinzen-Dick, Thomas Walker, Tor Tolhurst, Arman Rezaee, as well as other comments from Michael R. Carter, Steve Boucher, Jamie Hansen-Lewis and attendees at UC Davis Development Workshop and Development Tea. All errors herein are my own.

I. Introduction

The classic story of informal risk sharing is of two people who are subject to uncorrelated shocks. This includes risks like illness, loss of employment, and theft: if you lose your job, I insure you; If I lose my job, you insure me. Informal risk sharing has been long studied as a way to smooth this idiosyncratic risk. Indeed, we often see evidence consistent with a high degree of idiosyncratic risk sharing (Townsend, 1994; Kinnan, 2014). In contrast, sharing of correlated risks like output price and weather has received much less attention. Despite this, when risk preferences di er, sharing covariate risk can lead to welfare improvements by shifting covariate risks from more risk averse to less risk averse agents in exchange for a \premium" (Chiappori et al., 2014). In this story of informal risk sharing, in a bad year the less risk averse agent takes the hit. In a good year, they receive the prize; and in all years they are rewarded by the more risk averse agent for taking on this risk.

However, the ability to share covariate risk in this way depends crucially on the ability of agents with di erent risk preferences to pool risk. Homophily, or the preference to connect to those similar to oneself, is a common feature of social and economic networks (Miller McPherson et al., 2001). In the setting of risk sharing, Attanasio et al. (2012) demonstrates that individuals prefer to share risk with those who have similar risk preferences in a lab in eld setting. This pattern as *assortative matching* on risk preferences arises as a barrier to this type of risk sharing. This leads me to the question of interest: given the degree of assorative matching on risk preferences found in real world risk sharing networks, how good of insurance can covariate risk sharing deliver? Empirically, I study this question by asking if individuals form connections with others who have similar or di erent risk preferences. To translate these estimates into concrete welfare estimates, I construct a theoretical model of optimal risk pooling.

In Section III, I build a theoretical model of risk pooling in a village setting. This features agents with high and low risk aversion who will be assigned to one of two risk pooling groups. According to this model, optimal risk pooling happens when the composition of the groups reflects the composition of the village. For example, if the village is made up of 50% less risk averse individuals, you would prefer each group to also be made up of 50% less risk averse individuals. This result implies that optimal risk pooling should feature no assortative matching on risk preferences.

To measure the degree of assorative matching on risk preferences, I bring econometric models of network formation to rich microdata featuring income shocks, networks, and risk preferences from a survey in four villages in rural southern Ghana, described in Section IV. This setting features prominent correlated income risk (for example, price risk in Pineapple markets), and the data includes a detailed social networks module, and a set of hypothetical gambles. I look carefully at networks and use two main measures of the risk pooling network. First, I connect people who have exchanged gifts in the past and trust each other in a bilateral risk sharing network. Second, I arrange individuals in risk pooling groups using *community detection*, clustering methods which are sensitive to the details of networks (Newman, 2011). I argue this measure of risk pooling groups accounts for the possible scope of risk pooling in networks (i.e., the relevant set of individuals). For risk preferences, in order to match the theory as closely as possible, I back out coefficients of absolute risk aversion using these hypothetical gambles.

Using dyadic regression, I estimate that individuals prefer to assortatively match on risk preferences in the risk sharing network. That is, they prefer to match with individuals who have a similar degree of risk aversion. Moreover, this tends to be driven by connections between family members. That these results mirror those of Attanasio et al. (2012) is notable given that the current paper differs in both in country context and research design. In addition, I estimate a Subgraph Generation Model (or SUGM), which yields similar results. Additionally, this model allows for further exploration of who matches with who: I find that assortative matching is driven by less risk averse individuals, who tend to have higher degree overall and harbor a preference connect to their own type. When I treat risk pooling communities as the relevant unit of risk sharing at the subvillage level, however, I do not find evidence of assortative matching on risk preferences. When estimating the SUGM, I find that assortative matching still holds in risk pooling communities, at a significantly reduced magnitude. In other words, risk pooling communities feature more diverse preferences than bilateral risk sharing relationships. Section V desribes this research design and Section VI presents these findings.

What are the welfare impacts of the degree of (and reduction in) assortative matching? In Section VII, I divide individuals into more and less risk averse types and quantify the welfare implications of the allocation of types in communities. To do this, I simulate four scenarios: (A) an optimal scenario, (B) a community scenario, (C) the bilateral scenario, and (D) a worst case scenario. I find substantial differences between the optimal and worst case scenario, with the community and bilateral scenarios falling much closer to optimum and quite close together. First, I find that the observed networks tend to be close to optimal networks already. If 0% is the worst case scenario, and 100% is the best case scenario, the observed networks are close to 75%. As you might expect based on the results, the risk pooling groups function better for this than the bilateral networks. However, if we use full covariate insurance as a benchmark, even the optimal scenario has losses equal to 16.5% of per capita consumption. While individuals may be able to do well with the tools they are given, there are still large gains to be had in risk management.

Finally, how do we explain the di erences between bilateral (or *preferred*) risk sharing and the community level (or *effective*) risk pooling? Several mechanisms might account for the di erence. In section VIII I conclude with a discussion of the results, welfare implications, and possible explanations for the di erences in assortative matching in risk sharing networks v. risk pooling communities.

II. Background

A. Graphs and Risk Sharing Networks

Throughout this paper, I will draw on graph theory to describe intuition, formalize intuition and visually represent risk sharing networks. A graph g is a set of *nodes* and an *edgelist* (which naturally contains *edges*). I refer to these nodes and edges by their subscripts. I subscipt nodes by i. For edges, I use the combination of subscripts i and j to refer to that edge: if there is a connection between i and j, I say ij 2 g, hence ij is in the edgelist. An adjacency matrix represents these nodes and edges in an n n matrix $A = A(g)$. For the scope of this paper, I work with unweighted and undirected graphs, choosing to work with reciprocal relationships. Thus $a_{ij} = 1$ if ij 2 g and 0 if not. The adjacency matrix is also symmetric: $a_{ij} = a_{ji}$ for all $i; j$. The diagonal $a_{ii} = 0$ by construction.

Nodes and edges go by many other names. In the case of risk sharing, nodes represent agents and edges represent the social connections between those agents. I will use \agents" and \individuals" interchangeably when referring to nodes in the network. Likewise, I will use \links" and \connections" interchangeably when referring to edges. *Dyads* are not interchangeable, however: dyads are all possible combinations ij regardless of whether that edge exists in the network.

How do risk sharing networks relate to risk sharing arrangements? I take the view that risk sharing arrangements (and by extension, risk pooling arrangements) are *ex ante* informal contracts specifying transfers based on *ex post* states of the world. Given the complexity of such a contract, the edge itself does not tell the story of the risk sharing arrangement happening over that edge. For parameterized contracts, however, we can list these parameters as edge characteristics. Depending on whether the risk sharing network is *ex ante* or *ex post*, these social connections may reflect agreements of transfers or past transfers (or some mix of the two).

Some node level measures I will use are *neighborhood* and *degree*. An agent i's *neighborhood* is all agents j such that $ij \in g$. An agent i's *degree* is the size of agent i's neighborhood. We can compute degree $d(i) = \sum_{j=1}^{n} a_{ij}$. On the network level, some measures of network structure are *density* and *clustering*. The *density* of a network is the total number of connections divided by the number of potential connections. The clustering coefficient of a network takes (three times) the number of triangles and divides it by the potential number of triangles.

B. Related Literature

Covariate and Idiosyncratic Risk Sharing. — Formal insurance markets are often missing in developing economies (Mccord et al., 2007). In the absence of formal insurance markets, informal risk sharing, mediated through social networks, helps fill this void (Fafchamps, 2003, 2008). In theory, covariate risk sharing has great potential to benefit those who face risk (and lack formal insurance markets). Using panel data from Thailand, Chiappori et al. (2014) structurally estimates risk preferences in villages when risk sharing arrangements are complete. The current work rst di ers in modeling approach by using constant absolute risk aversion preferences as opposed to constant relative risk aversion. This choice relates more the the convenience of presenting welfare results. The solution of the two modeling approaches share an intuition: less risk averse agents might take on more covariate risk in exchange for some increase in consumption over the long term. In their model, individuals who have risk preferences near to risk neutral make a market in taking on additional covariate risk. This intuition is further shared by the model in Wang (2015). More substantially, the current work relaxes the assumption of covariate risk pooling at the village level to get at the relationship between network structure, network formation and the viability of covariate risk sharing.

Assortative Matching in Risk Sharing Networks. — A tension fundamental to covariate risk sharing is between the observed propensity for *homophily* in risk sharing networks and the need for diversi cation of risk and risk preferences. Homophily { or the principle that similarity breeds connection { has long been studied in social networks more broadly (Miller McPherson et al., 2001). A number of theoretical papers make predictions about assortativity in risk sharing networks. In the context of risk sharing, Attanasio et al. (2012) constructs a model of network formation that predicts those with similar risk aversion tend to connect in a network formation process. Wang (2015) also builds a model of risk sharing with heterogeneity in risk preferences and predicts assortative matching in risk preferences under certain conditions. Notably, these conditions hold in the experiment done by Attanasio et al. (2012). Agents are not only predicted to match assortatively on risk preferences. Gao and Moon (2016) build a model of risk sharing which predicts that agents will match assortatively on income shock variance.

Empirically, we also see evidence of assortative matching on various dimensions. Attanasio et al. (2012) uses a risk pooling experiment to test for assortative matching in 70 municipalities in Colombia. In this study, the authors measure risk aversion by offering a set of gambles. Then they give their experimental subjects the ability to form risk pooling groups to pool risk. They find close friends and relatives match assortatively on risk attitudes, whereas less familiar dyads tend not to match at all. My results qualitatively mirror these results, providing a scientific replication of the results of Attanasio et al. (2012) using observational data from Ghana. In addition, Fafchamps and Gubert (2007) finds evidence of geographic assortative matching in risk sharing networks (though this is likely correlated with kinship).

The Scope of Risk Pooling. — The *scope* of risk pooling is the various set of individuals a given individual might share risk with, given the need. Very often the implied scope of risk pooling is all other agents in the village, as much of the literature takes the assumption of a common pot at this level. For example, Townsend (1994) develops a model to test for the presence of efficient risk sharing at the village level. Contrary to this assumption of village level risk sharing, transfers tend to occur over reciprocal, bilateral relationships (Fafchamps, 2003; Weerdt and Dercon, 2006; Blumenstock et al., 2016). Here the scope of risk pooling is only the individuals directly connected to each other with each connected pair forming a risk pool. Specific contexts allow us to observe the scope of informal or quasi-formal risk pooling groups, including mutual fire insurance in Andorra (Cabrales et al., 2003), funeral societies in Ethiopia and Tanzania (Dercon et al., 2006), and kinship groups in Malawi (Fitzsimons et al., 2018). However, these contexts tend to be the exception. In general, informal insurance groups do not coincide with the village.

What is the relevant scope of risk pooling? Absent these quasi-formal groups, is it only to ones (network) neighbors, or are some members of village beyond neighbors contribute to risk pooling? Theory suggests reasons for relevant individuals to lie beyond the network neighborhood. First, Ambrus et al. (2014) models the e ect of risk sharing network structure on *ex post* consumption risk sharing. After shocks have been realized, they hypothesize the emergence of risk sharing islands where consumption is smoothed. The scope of risk sharing extends beyond the immediate network neighborhood. These risk sharing islands di er from the risk pooling communities described later in this paper, but are conceptually related. In particular, risk pooling communities map the *ex ante* structure of village risk sharing networks while risk sharing islands map *ex post* consumption smoothing conditional on existing networks and realized shocks. Risk shairng islands arise *ex post* where risk pooling communities exist *ex ante*. Second, Bourles et al. (2017), model altruism in networks and nd that intermediaries are important to ows of transfers through networks when networks are intransitive.

Empirically, Jackson et al. (2012) emphasizes the role of network *support* in favor exchange (including risk sharing transfers), i.e., that stable gift giving relationships will have a third person connected to both \supporting" the relationship. In contrast the theoretical results, this might imply a restriction to the network neighborhood to only those individuals featuring support. Fitzsimons et al. (2018) predicts that larger risk sharing groups will result in worse risk sharing and nds evidence for this using data from rural Malawi. Weerdt and Dercon (2006) nds that rst order connections matter for illness related risk sharing, but second order connections do not. However, in co-current work (using replication data from Attanasio et al. (2012)) I nd that constructing risk pooling communities using a friends and family network can help predict joining an experimental risk pooling above and beyond the network itself. Moreover, rst order connections within community are strongly predictive of joining the same risk sharing group. Additionally, second order connections within community and rst order connections outside of community also tend to predict co-membership in experimental risk pooling groups. This empirical and theoretical work motivates the use of community detection algorithms.

Mobile Money and Covariate Risk. | What is and isn't covariate risk is also evolving with the adoption of mobile nancial technologies. The rapid spread of mobile money payment systems has reduced transaction costs for transferring money long distances (Jack and Suri, 2014). For those connected to mobile money, canonical covariate shocks (drought, ooding, earthquakes) can become idiosyncratic (Blumenstock et al., 2016; Riley, 2018).

III. Theoretical Model

Consider a case where a risk neutral planner constructs a community in a village to maximize expected utility within a community. Here I leave aside community size and its impact on community composition and focus on optimal community composition. In this case, we think about community composition with regard to risk aversion, with relatively less and more risk averse individuals. We set up this problem in two steps. First, we characterize how risk is shared in a community according to its composition. Second, using the solutions and value functions from the rst optimization problem, we write a planners problem maximizing aggregate expected utility of consumption in a village with communities conditional on the composition of those communities.

A. Risk Sharing in Communities

To model risk sharing in communities, we start from a baseline of perfect idiosyncratic risk sharing. This means that all shocks that are above and below a villager's mean income are smoothed to their mean income (we will assume these are zero for the purposes of this problem). After this set of transfers takes place, a round of risk shifting takes place. Less risk averse individuals may take on more of covariate risk. This covariate risk derives from both the average idiosyncratic shock, which in general is not zero, and the (perfectly correlated) covariate shock. More risk averse agents are able to take on less of the covariate risk over time, shifting them on to less risk averse individuals. However, less risk averse individuals are still risk averse, so they require some compensation for the risk they take on. Thus, recurring transfers are made to these individuals regardless of the covariate shock.

Setup. | Suppose a community of xed size N. Villagers have exponential utility functions with coe cient of absolute risk aversion .

$$u(c) = e^{c}$$

However, suppose we have low and high risk aversion households, where type and indexed by ` $= 1; 2$. Then we assume $_2 > \ _1 > 0$. $N_{`}$ is number of individuals of type `, and $p = N_1 = N$ characterizes the composition of the group. All households face a perfectly correlated shock $\tilde{y}_V$ and an idiosyncratic shock $\tilde{y}_i$. Risk is symmetric between households and between types: Household level shocks, $\tilde{y}_i \overset{iid}{\sim} N(0; \;^2)$ and type level shocks $\tilde{y}_v \overset{iid}{\sim} N(0; \;^2)$.

$$y_{\ell i} = \tilde{y}_i + \tilde{y}_v$$

Taking account of risk sharing process, we write the consumption of household i of type ℓ as a weighted sum of the idiosyncratic and covariate shock in the community. For type $\ell = 1$

$$c_{1i} = \frac{}{p} \left(\frac{1}{N} \sum_{i=1}^{N} \tilde{y}_i + \tilde{y}_v \right) \quad {}_{1i}$$

and for type $\ell = 2$

$$c_{2i} = \frac{1}{1 \quad p} \left(\frac{1}{N} \sum_{i=1}^{N} \tilde{y}_i + \tilde{y}_v \right) \quad {}_{2i}:$$

The proportion of covariate risk that is borne by the less risk averse individuals in the community is regulated by the parameter $\in [0; 1]$. When $= 1$, all risk is taken on by less risk averse individuals, when $= p$, risk is shared equally among all members of the community (i.e., only idiosyncratic risk is pooled), and when $= 0$ all risk is taken on by more risk averse households. Conversely, ${}_{\ell i}$ regulates the recurring transfers from the more risk averse to the less risk averse. Thus, it is the case that the aggregate transfer into the pot exceeds the aggregate transfer out:

$$N_1 \; {}_{1i} \quad N_2 \; {}_{2i}:$$

Due to the exponential utility function and normal distribution of shocks, we are able to represent expected utility as a mean-variance decomposition

$$E(U(c_{`i})) = E(c_{`i}) \quad \frac{`_i}{2} Var(c_{`i})$$

We will refer to the special case of $EU_0 = E(U(c_1 i j \quad = p)))$, which simply describes the utility from idiosyncratic risk sharing in absence of risk sharing.

Optimization Problem. | We maximize expected utility of less risk averse agents subject to several constraints. Constraint 2 is an incentive compatibility constraint: more risk averse agents cannot being worse o than in the case where they only perfectly pool idiosyncratic risk. Constraints 3 and 4 serve as individual budget constraints for each type, and nally 5 serves to ensure feasibility of the recurring transfers.

(1) $$\max_{1;\ 2;} E(U_1(c_{1i}))$$

(2) subject to $$EU_0 \quad E(U(c_{2i}))$$

(3) $$c_{1i} = \quad \frac{}{p} \quad \frac{1}{N} \sum_{i=1}^{N} y_i + y_v \quad ! \quad {}_{1i}$$

(4) $$c_{2i} = \frac{1}{1 \quad p} \quad \frac{1}{N} \sum_{i=1}^{N} y_i + y_v \quad ! \quad {}_{2i}$$

(5) $$0 \quad p \;{}_1 + (1 \quad p) \;{}_2$$

Solutions and Value Functions. | How much covariate risk is shifted to the less risk averse agents? Recall, if = 1, all covariate risk shifts to less risk averse inviduals, and if = p, the baseline of perfect idiosyncratic risk sharing is maintained.

$$(6) \qquad (p) = \frac{p \ _2}{(1 \quad p) \ _1 + p \ _2}$$

Since $_2 >$ $_1$, $> p$. This means some degree of covariate risk is shifted to less risk averse individuals. Likewise, unless $_1 = 0$ (we assumed it doesn't) or $p = 1$ not all risk is taken on by the less risk averse. This brings us to an important result of the model: group composition matters for the degree of risk sharing. What are more risk averse agents willing to pay to shift risk away? Since $_2$ is paid into the community pot, Type 2's willingness to pay depends on their own risk aversion, type 2's risk aversion, and group composition

$$(7) \qquad _2(p) = \frac{^2}{2} \quad 1 \quad \frac{^2_1}{((1 \quad p) \ _1 + p \ _2)^2}$$

where the expression in parentheses lies between 0 and 1. Because risk is symmetric in this model, the transfer does not depend on covariate risk. Presumably, were this not the case, it would appear in the above solutions.

Finally, type 1 will maximize their utility and hence the payments they receive from type 2. We can write $_1$ by converting type 2's willingness to pay into type 1's average payment:

$$(8) \qquad _1(p) = \frac{1 \quad p}{p} \ _2(p):$$

Once we have these solutions we are able to compute the value functions we will use in order to solve the planner's problem. These are

(9) $$V_1(p;\ _1;\ _2) = \frac{1}{2}\ \frac{1\quad p}{p}\quad 1\quad \frac{1}{(1\quad p)\ _1 + p\ _2}\ _2^{!}$$
$$\frac{1}{2}\ \frac{2}{(1\quad p)\ _1 + p\ _2}\ ^2\ \frac{2}{n} +\ ^2$$

(10) $$V_2(p;\ _1;\ _2) = \frac{2}{2}\quad 1 + \frac{1}{(1\quad p)\ _1 + p\ _2}\ ^2\quad 1 + \frac{2}{n} +\ ^2\ !$$

B. Planner's Problem

Setup. | The risk neutral planner seeks to maximize aggregate expected utility of consumption conditional on the composition of communities. We have two communities, $g = A; B$. We will update the notation from the rst stage slightly. For a given community g, N_g is the community size and $N_A + N_B = N$. Then $N_{\ell g}$ is the number of individuals of type ℓ in community g and $p_{\ell g} = \frac{N_{\ell g}}{N_g}$.

Optimization Problem. | We can state the planner's problem as an optimization problem where the planner

(11) $$\max_{N_{1A}} N_{1A} V_1(p_{1A}) + N_{2A} V_2(p_{1A}) + N_{1B} V_1(p_{1B}) + N_{2B} V_2(p_{1B})$$

(12) subject to $$N_1 = N_{1A} + N_{1B}$$

(13) $$N_2 = N_{2A} + N_{2B}$$

(14) $$N_A = N_{1A} + N_{2A}$$

(15) $$N_B = N_{1B} + N_{2B}$$

To simplify this problem, we consider the simple case where we have an equal number of more and less risk averse types. That is, $N_1 = N_2$. This implies that we can encompass the entire problem just by looking at one choice parameter, p_{1A}, and conditioning on the size of the smaller community, N_A. $p_{1A} = \frac{N_{1A}}{N_A}$, and we can express $p_{2A} = 1 - p_{1A}$, $p_{1B} = \frac{2N_{1B}}{N} = \frac{2(N_1 - N_{1A})}{N}$ and $p_{2B} = 1 - p_{1B}$. Setting $N_1 = N_2$ reduces the set of constraints to three, and our quick computations takes account of these three constraints. Hence we have the problem

$$
\begin{aligned}
(16) \qquad \max_{N_{1A}} \; & N_{1A} V_1\left(\frac{N_{1A}}{N_A}\right) + N_{2A} V_2\left(\frac{N_{1A}}{N_A}\right) \\
& + N_{1B} V_1\left(\frac{2(N_1 - N_{1A})}{N}\right) + N_{2B} V_2\left(\frac{2(N_1 - N_{1A})}{N}\right)
\end{aligned}
$$

Solving this planners problem for an analytic solution is relatively difficult. However, it is relatively easy to characterize the optimal allocation of types numerically. In Figure 1, I plot the objective in Problem 16 against p_{1A}, the new choice variable. To construct this example, I set $\sigma_c^2 = 100^2$, $N = 100$, $N_1 = N_2 = 50$, and set $\gamma_1 \approx 0.0016$, $\gamma_2 \approx 0.0037$, the means of these for the type 1 and type 2 in my data (see Section IV.A for construction of coefficients of risk aversion and types).

Looking at Figure 1, we see that welfare is maximized when $p_{1A} = 0.5$, when diversity of types is maximized. Likewise, we see that welfare is minimized when as p_{1A} approaches 0 or 1, when diversity of types is minimized. Interestingly, unequal sized suboptimal communities improve over more equally sized suboptimal communities holding p_{1A} equal. Also interesting, welfare is not symmetrically suboptimal when the proportion of less risk averse agents strays from zero. If a community is overfilled with a type (i.e., $p_{1A} \neq 0.5$) it is better to "overfill" the smaller community with type 1 (less risk averse) agents as opposed to overfilling

Figure 1. Optimal Allocation of Types Between Unequally Sized Communities

the larger community. For another way to look at this, in the appendix I plot the proportion of risk taken on by both groups in a risk pooling frontier (see Figure D1). Finally as a demonstration that this is not an artifact of equal sized communities, in 2, I vary the composition of types in the population. In this gure we see that welfare is maximized when $p_{1A} = p_1$, the proportion of type 1 agents in the population.

IV. Data and Context

A. Data

The data I use comes from AMA-CRSP Survey of Risk Coping and Social Networks in southern rural Ghana. In particular, these data feature information about assets, income, consumption (positive and negative) shocks data for households in four villages in Ghana. The survey ties this to network data collected to measure risk sharing networks (Walker, 2011a). The survey instrument and

Figure 2. Optimal Allocation of Types Between Unequally Size Communities with varying numbers of Type 1 and Type 2 agents.

technical details can be found in Walker (2011b).[1]

After processing, the data includes 631 individuals split into four villages. The survey is of spouses in households, though I keep the data at the individual level. However, sample sizes for each empirical exercise vary, since the unit of analysis is not the individual. In the dyadic regressions, the sample size is the number of dyads, or the number of potential connections between individuals within villages. In the dyadic regressions, sample size = 71062. In Subgraph Generation Models, the sample size is determined by which feature is being estimated, and are reported in the estimation tables.

To test hypotheses about assortative matching in risk sharing networks and do welfare simulations, I construct the data to match the theoretical model presented as closely as possible. In particular, I construct risk preferences assuming constant absolute risk aversion preferences and use community detection to construct risk

[1] Walker (2011a) and Walker (2011b) as well as the raw data can be downloaded at http://barrett.dyson.cornell.edu/research/datasets/ghana.html

pooling groups that serve as empirical analogues to the modeled risk pooling groups.

Risk Preferences. | Leaving aside risk neutrality, using a set of four hypothetical gambles we use exponential preferences to measure individuals risk aversion.The first two menus presented are in the gains domain. In the first menu, the risky gamble Y_B is held fixed while increasing the sure payment. In the second, the sure payment is held fixed while reducing the upside of the gamble. We assume that if an individual reaches a point of indifference between two gambles, we assign them to the midpoint between the two gambles. Hence, if the mean differs, we take the average of the mean of the two gambles and assign this value to the point of indifference. If the variance differs, we take the average of variances and assign this value to the point of indifference. The second two menus are reflections of the first two onto the domain of losses.

(17) $$u(c) = \frac{1 \quad e^{\;c}}{}$$

For mathematical tractability, we assume Y_B is normally distributed. Using this, we represent expected utility as a mean variance utility function (Sargent, 1987). We write

(18) $$E(Y) \quad \frac{}{2} V(Y)$$

I look at a menu of choices between hypothetical gambles.Y_A is a set of sure payments and Y_B is a set of risky income sources. Respondents are able to choose

Figure 3. Risk Sharing Networks with Risk Aversion Plotted: Darmang (top) and Pokrom (bottom)

Figure 4. Risk Sharing Networks with Risk Aversion Plotted: Oboadaka (top) and Konkonuru (bottom)

between these will be indi erent between the two when

(19) $$E(Y_B) \quad \frac{}{2} V(Y_B) = Y_A$$

Observing this indi erence point, we can write

(20) $$= \frac{2(E(Y_B) \quad Y_A)}{V(Y_B)}$$

and recover the coe cient of absolute risk aversion. We compute ^ for each lottery and individual. To combine these into measures of risk aversion, we average over individuals and use the mean risk aversion.

Table 1|Risk Sharing Network Summary Statistics by Risk Preferences. Standard errors in parentheses.

	More Risk Averse	Less Risk Averse	Risk Loving	Not Surveyed
Isolates	0.09	0.09	0.16	0.38
	(0.001)	(0.001)	(0.004)	(0.005)
Degree	4.62	6.61	4.79	2.72
	(0.02)	(0.03)	(0.06)	(0.05)
Clustering	0.25	0.23	0.23	0.17
	0.00	0.00	0.00	0.00
Betweenness	85.98	119.53	99.85	43.23
	(0.69)	(0.90)	(2.56)	(1.27)
Closeness	31.03	39.15	32.14	21.08
	(0.09)	(0.10)	(0.26)	(0.25)
N	236	217	96	82

Risk Sharing Network. | To construct the risk sharing network, I use the intersection of a gift network and a trust network. To construct the gift network, a link occurs if individual i has received a gift from j and also if i has given a gift to j. In the trust network, I report a link if both i and j report trusting each other (in previous uses of this data, this has been a \strong ties" network). If a tie occurs in both networks, I record a tie between i and j and use this as my risk sharing network.

For an alternative network, I aggregate the trusted *ex post* gift network and a trusted kin network. In the kin network a link occurs if i reports being related to j or vice-versa and the trust network is constructed as above. To aggregate these networks, I take the intersection of the trust and the kin networks rst. Then, I include a link in the risk sharing network if a link occurs between i and j in either the trusted kin or the trusted gift network.

Table 1 presents risk sharing network summary statistics, aggregated across villages. Degree is a measure of centrality, which counts the number of individuals who are directly connected to an individual. Here we present the average degree. Isolates is computed as the proportion of individuals who are not connected to the network. Clustering is the average local clustering coe cient. This measures, for individual i connected to j and k, what proportion of the time are j and k also connected (also called transitivity). Betweenness and closeness are both measures of network centrality. Betweenness centrality measures how betweenness measures how central an individual is in a network by counting how many nodes it lies on the shortest path between. Closeness centrality measures the distance to the rest of the network. Individuals who are short distances on average from others in the network, will have high closeness centrality.[2] When comparing less and more risk averse individuals we can see di erences in almost all measures of centrality. Less

[2] In particular, I use harmonic closeness centrality, computed

$$C(i) = \sum_j \frac{1}{\text{distance}(i,j)}.$$

risk averse individuals have lower degree, are closer on average to the rest of the network, and hold network position in between more other individuals. Despite this, it's interesting to note that the di erence in clustering between less and more risk averse individuals would appear to be economically small.[3]

V. Empirical Strategy

A. Community Detection

Intuition. | Community detection is unsupervised learning. Because we're not using a measure of risk sharing to validate the algorithm here, we need to be careful in choosing algorithm that relates in a reasonable way to our understanding of risk sharing in networks. To give intuition to the algorithm choice, consider a potential risk sharing process. A large gift is given to a randomly chosen individual in a risk sharing network. The household gives a gift to a (network) neighboring individual who is less well o than they are, sharing their positive shock equally. Ex ante, all the individual's neighbors should be equally likely to receive this gift.[4] Having received this gift, this individual is also is obligated to share with their worse o neighboring individual, provided they are not worse o than all of their network neighbors. This process of risk sharing continues. Eventually, the individual receiving the transfer ends with an e ective shock which is still worse than their neighbors and gives no gifts. We end up with progressively smaller transfers \walking" randomly through the network. Individuals who are close to

[3] Though it is not the topic of the current work, one might interpret this as a di erence in linking social capital without an accompanying di erence in bonding social capital. In terms of communities discussed later, this might also suggest that less risk averse individuals might be more likely to serve as liaisons between risk pooling communities, while being similarly integrated into the dense and clustered risk pooling communities as their more risk averse counterparts.

[4] There might be many reasons the household would feel obliged which might range from self-enforcing contracts, i.e., the threat of losing future access to their network (Coate and Ravallion, 1993; Ligon et al., 2002; Ambrus et al., 2014), or altruism (Boures et al., 2017).

the individual who received the prize will have a high probability of receiving a transfer, while those far will have a low probability. Likewise, even if one does not receive a transfer in the first step, if an individual is connected to many of the same households as the one receiving the prize, they get additional chances for a gift. We would expect that within a dense and clustered portion of the network, most often the gains from the positive shock will not make it far, instead getting "trapped" in the local network.

Similarly, we can present this same intuition for negative shocks: a randomly chosen individual a negative shock. Based on a similar rule, they might request a favor from one of their neighbors, who might request a favor from one of their neighbors, and so on. We end up with progressively smaller transfers being drawn through the network. Close individuals will more likely receive a request for a gift. Likewise, requests will remain trapped within dense and clustered portions of the network.[5]

Walktrap Algorithm. — This process functions as a way risk is effectively pooled and mirrors the intuition of the walktrap algorithm. The walktrap algorithm uses random walks over the network to measure the similarity of nodes. The algorithm exploits the fact that a random walker in a graph tends to get "trapped" within tightly knit sections of the graph. A random walker starts at a random node i and moves to an adjacent node with probability $1/d(i)$ (where $d(i)$ is the degree of i). Then for each of a fixed number of periods, this process is repeated from the landing node, k, moving to an adjacent node with probability $1/d(k)$. If nodes are in the same community a random walk of fixed length from

[5]The validity of this process isn't essential for the success of the algorithm of course. Many parallel reasons might point the same direction. For example, tightly knit networks might serve as more useful to overcome information asymmetries in networks if information is passed along the same links (Bloch et al., 2008).

node i should often land on node j.

However, walks from node i frequently landing on node j is not a perfect condition for nodes to reside within the same community. First, nodes with high degree tend to receive more random walks, so far o individuals with high degree might be labeled as within the same community. For this reason, the measure of distance should take into account the degree of the receiving node. Second, individuals in the same community should see other nodes in the the network similarly. Pons and Latapy (2004) uses this second fact to build a measure of similarity.

$$r_{ij} = \sqrt{\frac{\sum_{k=1}^{n} (P_{ik}^{s} - P_{jk}^{s})^2}{d(k)}}$$

where P_{ik}^{s} is the probability of a random walk of length s from node i landing on node k. The algorithm works by iteratively merging nodes who are similar into communities.

The algorithm starts by assigning each node into its own community in the network. Similarities are computed between adjacent nodes using the measure above and the closest two communities are merged. Then, the merged community remains in the network as a single node (with all its out of community links preserved). Similarities are updated and the process repeats until all nodes are merged into one community. This builds a *dendrogram*, or a tree depicting the merges of nodes into communities. Each merge is depicted by a split in the tree. This gives us many possible community assignments. To choose an optimal community assignment, a tuning statistic called *modularity* is used to cut the dendrogram at one of the splits.

Modularity measures the internal quality of the community by looking at how many links exist within the community compared to a how many we would expect at random. The measure follows from a thought experiment: suppose you were to take a graph and randomly \rewire" it. This rewiring preserves the degree of individual nodes, while destroying the community structure. We take the average number of within community links from rewiring as the counterfactual. Having many more links within the community than the counterfactual implies a good community detection, fewer implies a poor community structure. For more about modularity, including the formal de nition, please see Appendix B.B2.

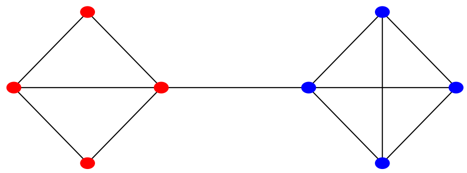

Figure 5. A Stylzied Risk Sharing Network: what we observe, with (latent to the econometrican) communities denoted by red and blue.

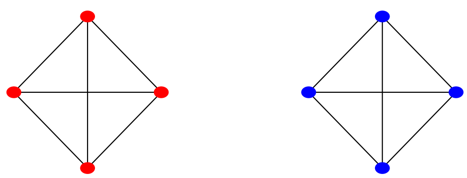

Figure 6. Community Network: After detecting the communities, we connect all who are within a community and disconnect those between communities. This is the Network Arising from Community Detection.

Networks Arising From Community Detection. | After we have assigned nodes to communities, we construct an additional risk sharing network using these community assignments. Assuming that e ective risk pooling takes place at the community level, all nodes assigned to a given community are linked within the network. Additionally, we assume no risk sharing takes place between com-

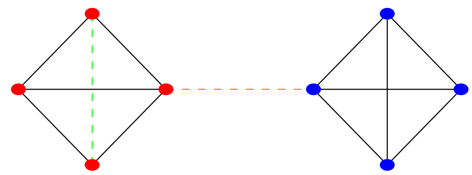

Figure 7. Difference in Networks, community network less risk sharing network. Green is added, orange is removed. Here we see one connection added to the red community and one removed between the red and blue communities.

munities, hence we \delete" links that cross community boundaries. We represent the community graph using an adjacency matrix C where $c_{ij} = 1$ if i and j are in the same community and 0 if not. Like the adjacency matrix, C is symmetric. I use this network in dyadic regression and subgraph generation models to measure who you pool risk with in effective terms. This is depicted in Figures 5, 6, and 7.

B. Dyadic Regression

Before diving into the Subgraph Generation Models, it is useful to estimate dyadic regressions. While dyadic regressions cannot handle \higher order" features like stars or triangles, they are familiar, interpretable as regressions, and comparable with past literature. In particular, I note similarities and differences between these results and those found in Attanasio et al. (2012).

In these regressions, each pair of nodes is treated as an observation. Let $a_{ij} = 1$ if individual i has given and received gifts from j and both i and j trust each other. Hence our dependent variable is the risk sharing network defined by using trusted gift givers. This construction requires that $a_{ij} = a_{ji}$ so the adjacency matrix A is symmetric. In addition, as required by Fafchamps and Gubert (2007) all explanatory variables also enter symmetrically. In particular, I estimate a set of four linear probability models. The most parsimonious model regresses risk sharing connections on differences in measured risk aversion,

$$a_{ij} = \beta_0 + \beta_1|\gamma_i - \gamma_j| + \varepsilon_{ij} \tag{21}$$

where γ_i is the risk aversion of individual i and ε_{ij} is the error term. Here (as in the following specifications) a negative estimate β_1 is evidence of assortative matching, i.e., that individuals prefer to share risk with individuals who have similar risk preferences to their own. In our sample, less risk averse agents tend to be more well connected and have higher degree relative to more risk averse and risk loving agents. A second specification includes the sum of degrees d_i and d_j to control for the correlation between risk aversion and degree.[6]

$$a_{ij} = \beta_0 + \beta_1|\gamma_i - \gamma_j| + \beta_2(d_i + d_j) + \varepsilon_{ij} \tag{22}$$

A positive estimate of β_2 suggests that individuals who have more total links are more likely to link to each other. Given this endogenous construction, positive estimate will not come as a surprise and is closer to an exercise in accounting. A third specification examines kin ties as a predictor of risk sharing connections.

$$a_{ij} = \beta_0 + \beta_1|\gamma_i - \gamma_j| + \beta_3 f_{ij} + \varepsilon_{ij} \tag{23}$$

where family is $f_{ij} = 1$ when i and j report being family members. A positive estimate of β_3 suggest that family are more likely to be connected within the risk

[6]The results, here and in the community network, are robust to computing the degree as a "leave-out" degree for each dyad, i.e. leaving out the dyad in question.

sharing network. A fourth specification combines specifications 1 and 2.

$$a_{ij} = \beta_0 + \beta_1 |\gamma_i - \gamma_j| + \beta_2 (d_i + d_j) + \beta_3 f_{ij} + \varepsilon_{ij} \tag{24}$$

Finally, a 5th specification introduces interactions between the difference in coefficients of risk aversion and family ties.

$$\begin{aligned} a_{ij} = {} & \beta_0 + \beta_1 |\gamma_i - \gamma_j| + \beta_2 (d_i + d_j) + \beta_3 f_{ij} \\ & + \beta_4 f_{ij} |\gamma_i - \gamma_j| + \varepsilon_{ij} \end{aligned} \tag{25}$$

A negative estimate of β_4 is evidence that assortative matching is stronger among family members. Moreover, if $\beta_1 + \beta_4$ is negative, this provides evidence that within family members, risk aversion is an important determinant of risk sharing connections.

In addition, I re-estimate regressions 21, 22, 23, 24, and 25 with detected communities as the network (as opposed to the network adjacency matrix). In all of the above specifications, I replace a_{ij} with c_{ij}, the ij th entry of the community matrix C. $c_{ij} = 1$ if i and j are in the same detected community, and 0 otherwise. Degree sum is included in the re-estimation of 22, 24 and 25. While these coefficient estimates are comparable, note that in the community network this ends up being 2 × (community size − 1). Importantly, as we consider the community network in dyadic regression, we still view it as possible that a given agent could join any community in a village and have access to their network, hence set of dyads under consideration does not change from the risk sharing network models.

As mentioned above, I estimate these regressions as linear probability models (though results are similar estimating logistic regression). Note however, that the errors are non-independent. In particular, the residuals of dyads involving a particular node might be arbitrarily correlated, i.e., $Cov(\varepsilon_{ij}, \varepsilon_{lk}) \neq 0$ if $i = l$, $i = k$, $j = l$, or $j = k$. To correct standard errors for this non-independence, I use dyadic robust standard errors as proposed by Fafchamps and Gubert (2007), discussed in Cameron and Miller (2014). The asymptotic properties of this estimator are described in Tabord-Meehan (2019).[7]

C. Subgraph Generation Models

Intuition. — A useful tool for understanding risk sharing networks and communtities is called a Subgraph Generation Model (SUGM). SUGMs treat networks as emergent properties of their constituent subgraphs.[8] A subgraph (or induced subgraph) of a graph is the graph obtained from taking a subset of nodes in the graph and all edges connecting those nodes to each other. For example, if we select a subset of two nodes of a graph, the subgraph will be either a link or two unconnected nodes. If we take three points, we might observe a triangle (a trio of nodes all connected by edges), a line (one node connected to the two others), a pair and an isolate (only two nodes connected), or an empty subgraph (three unconnected nodes). In general, we focus on connected subgraphs for the SUGM. In three node example above, the means we leave aside the pair and isolate and the empty subgraph, focusing on the triangle and the line. Likewise, while a link is an interesting subgraph, two unconnected nodes is not. When describing these models, I will use "features" of the SUGM interchangably with subgraph.

[7] When a natural unit of clustering is present, as in Attanasio et al. (2012), that can be used to cluster standard errors as opposed to dyadic robust. Unfortunately, in our case, the data does not feature such a unit.

[8] While Exponential Random Graph Models (ERGMs) have similar motivation, they do not succeed at reconstructing graphs with any success. They depend on an assumption of independence of links, if this does not hold they are not consistent. To the contrary, much of the theory of risk sharing would expect links are dependent on each other (Chandrasekhar and Jackson, 2018).

While SUGMs can be estimated using GMM, I am able to directly estimate the parameters using an algorithm given by Chandrasekhar and Lewis (2016) and Chandrasekhar and Jackson (2018). Estimating a SUGM directly is essentially estimating the relative frequency of various subgraphs in a network. However, we can't stop at simply estimating the features. Because networks are the union of many subgraphs, subgraphs might overlap and incidentally generate new subgraphs. For example, three links placed bewteen ij, jk, and ik would incidentally generate a triangle. We want to estimate the true rate of subgraph generation. To do this, I order subgraphs by number of links involved in their construction. Then, I compute the number of subgraphs generated of that type, but only if they are not a portion of a \larger" subgraph. For subgraphs of same size, order is arbitrary, but must exclude occurrences of this subgraph incidentally generated by other subgraphs who are further along in the order. For example, for a SUGM featuring links and triangles, I order links rst, triangles second, etc. While counting links and potential links, I neglect pairs of nodes ij if jk and ik are in the graph.[9] This algorithm yields consistent estimates of the true rate that these subgraphs are generated at.

Links and Isolates Subgraph Generation Model with Types. | I start by estimating three simple SUGMs on the risk sharing network, using only links and isolates as features. The SUGMs seek to understand how individuals of di erent risk preferences connect to each other. For various reasons, a small subset of individuals in the network did not participate in the survey module that recovered risk preferences.[10] Hence, we start with a 2-type model to diagnose the di erences

[9] If we to add lines of 3 nodes, I could order these before or after triangles. Ordering lines before triangles I would look at potential links ij and jk where ik is not in the graph. Likewise, I would need to remove pairs of nodes ij if jk or ik are in the graph.

[10] Some of these individuals were not surveyed at all, but appear in the network, others may be part of the sample who missed that particular round or module.

between those individuals who participated and those who did not. I estimate five features: Isolates of surveyed nodes, isolates of nodes who were not surveyed, and links within surveyed nodes, links between surveyed nodes and non-surveyed nodes.

Of those in the network who have a risk aversion coefficient, I split these individuals into three groups: risk loving, less risk averse, and more risk averse. Risk loving are those with $\rho < 0$. This accounts for about 20% of the individuals with preferences. I split the remaining risk averse individuals into evenly sized groups of 40% each, with more risk averse individuals being above a cut-point. The second model estimates the full model with risk averse, risk loving and non-surveyed types for a total of three types. Third, I estimate the full model splitting risk averse indivudals into more and less risk averse types.

For each model I estimate $\tilde{\beta} = \left(\{\tilde{\beta}_{I,\ell}\}_{\forall \ell}, \{\tilde{\beta}_{L,\ell,\ell}\}_{\forall \ell}, \{\tilde{\beta}_{L,\ell,r}\}_{\forall \ell, \forall r}\right)$. Coefficients for isolates of type ℓ, $\tilde{\beta}_{I,\ell}$ are estimated

$$\tilde{\beta}_{I,\ell} = \frac{\sum_{i=1}^{n} 1(\deg(i) = 0 \mid \ell_i = \ell)}{n_\ell}; \tag{26}$$

coefficients for within links of type ℓ, $\tilde{\beta}_{L,\ell,\ell}$ are estimated

$$\tilde{\beta}_{L,\ell,\ell} = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_{ij} \cdot 1(\ell_i = \ell) \cdot 1(\ell_j = \ell)}{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 1(\ell_i = \ell) \cdot 1(\ell_j = \ell)}; \tag{27}$$

and coefficients for links between type ℓ and type r, $\tilde{\beta}_{L,l,r}$ are estimated

$$\tilde{\beta}_{L,l,r} = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_{ij} \cdot \left(1(\ell_i = l) \cdot 1(\ell_j = r) + 1(\ell_i = t) \cdot 1(\ell_j = l)\right)}{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 1(\ell_i = l) \cdot 1(\ell_j = r) + 1(\ell_i = t) \cdot 1(\ell_j = l)}. \tag{28}$$

From proposition C.2 in Chandrasekhar and Jackson (2018) under some sparsity conditions[11], $^{1=2}(\tilde{}_n \quad ^n_0) \; ! \; N(0;I)$ where $^n_0$ is the true rate of subgraph generation. For a feature`, the variance of the feature is the entry on the diagonal:

$$(29) \qquad {}_{s;s} = \frac{^n_{0;s}(1 \quad ^n_{0;s})}{_s \; ^n_{m_s}}$$

where m_s is the number of nodes involved in the feature and $_s$ is number of di erent possible relabelings of the feature (note: for both isolates and links $_s = 1$). So I estimate the standard errors of the SUGM,

$$(30) \qquad \tilde{}_{s;s} = \sqrt{\frac{\tilde{}^n_s(1 \quad \tilde{}^n_s)}{_s \; ^n_{m_s}}}:$$

For the results, $_s \; ^n_{m_s}$ is the \sample size" of the feature.

Pooled Subgraph Generation Models. | Since network data I am using has four networks, I need to make choices as to how to handle these multiple networks in the Subgraph Generation Model. One approach would be to estimate a subgraph generation model for each village and average the coe cients of these. A di erent strategy, and one that relies on the same asymptotics as the single network case from Chandrasekhar and Jackson (2018) is to pool the counts and potential counts from the villages to estimate a single coe cient across the villages. This leads to an adjusted class of SUGMs I term Pooled SUGMs. I decide on these for the ability to lean on the same asymptotic theory as the single network case. However, a few considerations need to be made. Principally, we can't

[11] A rst note here is that these networks are sparse by the de nition of Chandrasekhar and Jackson (2018). If we assume a constant growth rate of the density of links, we have that they are growing at about $n^{1=3}$ or less (which is acceptable). For this particular model, none of the features chosen can incidentally generate any other feature. For example, links cannot generate isolates, nor can isolates generate links. Because the second is true, for this particular model the conditions from Chandresekhar may be cracking a walnut with a sledgehammer, so to speak.

simply combine the networks and run the SUGM. For example, it is unlikely that the dyads that would occur between villages would be reasonable potential dyads. Hence we need to collect counts of features and potential counts of features in all four villages before combining. Let count_{sv} be the count of some subgraph s in village v and potential_{sv} be the potential number of times that feature could occur. These reflect the numerator and denominator, respectively, of equations 26, 27, and 28 above. We estimate the coefficient associated with some subgraph s

$$\tilde{\beta}_s = \frac{\sum_{v=1}^{4} \text{count}_{sv}}{\sum_{v=1}^{4} \text{potential}_{sv}} \tag{31}$$

This uses only the relevant potentail occurences of the feature. Similarly, when estimating the standard errors of a feature, we cannot use the same effective sample size as we would use if we combined the networks. Let n_v be the number of nodes in the village network. If we take $\sum_s \binom{n_v}{m_s}$ we would include many combinations of nodes that in reality could not form the subgraph in question. Hence we estimate the standard errors the of pooled SUGM

$$\tilde{\sigma}_{s;s} = \sqrt{\frac{\tilde{\beta}_s(1-\tilde{\beta}_s)}{\sum_s \sum_{v=1}^{4} \binom{n_v}{m_s}}}. \tag{32}$$

Differences in Assortative Matching. — These SUGM estimates give me a way to test for assortative matching between preferred and effective risk sharing networks. However, since the the risk sharing network and the community network have different degrees of attachment, to make an apples to apples comparison, I normalize my results by taking a ratio of coe cients. In particular, I compare the estimates of the 4-type SUGM with the results of the 3-type SUGM focusing on three coe cients of interest. These are within links for type 1 agents, within links for type 2 agents, and links between type 1 and two agents. For all three, I compare these links to the coe cient on links within any risk averse agents. Doing this for both the community coe cients and the bilateral network coe cients, we can compare which correspond to within links for agents of type 1, within links for agents of type 2 and links between types. For the current results, I construct approximations for the mean and variance of these ratios using approximations for the mean and variance of a ratio[12], and using parameter estimates as stand-ins for their means, we can use an analytic expression for the variance. See appendix B.B3 for more.

VI. Results

A. Community Detection

We assign individuals to risk pooling communities using walktrap community detection with walks of 4 steps applied to our risk sharing network. I general, longer walks tend to result larger communities, whereas smaller walks result in smaller communities (For a demonstration of how communities vary by length of walk, see Figure D3). For the resulting community detection see Figures 8 and 9.

B. Dyadic Regression

Table 2 reports the results from estimating equations 21, 22, 23, 24, and 25. I include village level xed e ects in all dyad regression speci cations, though this

[12] https://www.stat.cmu.edu/ hseltman/ les/ratio.pdf

Figure 8. Risk Sharing Networks in Darmang (top) and Pokrom (bottom) with walktrap community detection (colors and numbers represent communities).

Figure 9. Risk Sharing Networks in Oboadaka (top) and Konkonuru (bottom) with walktrap community detection (colors amd numbers represent communities).

Table 2|Dyadic Regression: Bilateral Risk Sharing Network

	(1)	(2)	(3)	(4)	(5)
$j\ _i\ \ _j j$	-0.00991	-0.00761	-0.00239	-0.00197	0.00127
	(-1.11)	(-3.21)	(-0.32)	(-0.70)	(0.40)
$jd_i\ \ d_j j$		0.00878		0.00705	0.00705
		(38.21)		(25.67)	(25.66)
fam_{ij}			0.517	0.419	0.438
			(32.22)	(30.11)	(29.08)
$fam_{ij}\ \ j\ _i\ \ _j j$					-0.0197
					(-2.39)
constant	0.267	-0.144	0.171	-0.141	-0.144
	(9.73)	(-14.11)	(7.55)	(-10.87)	(-10.97)
N	71052	71052	71052	71052	71052

t statistics in parentheses

p < 0:05, p < 0:01, p < 0:001

does not e ect the magnitudes estimated in any of the speci cations. Reported t-statistics have been adjusted by clustering at the dyad level. To make results more interpretable, I transform risk aversions into z-scores before computing regressors. This changes the interpretation just a bit. For example, $_1$ estimates the e ect of a one-standard deviation absolute di erence in risk aversion.

Column 2 and 5 present my preferred speci cations. This derives from an empirical fact and my interpretation of the mechanism. First, less risk averse individuals tend to be better connected than more risk averse individuals and risk loving individuals. While we could control for the sum of risk aversion (and the sum of squared risk aversion), I control for this degree e ect directly.

Across all speci cations we see negative estimates of $_1$ or $_1 + \ _4$. However, in columns 1 and 3, when the sum of degree is omitted from the model, the estimates are small in magnitude and are not statistically signi cant (at any standard

Table 3|Dyadic Regression: Community Network

	(1)	(2)	(3)	(4)	(5)
$\|\ _i \quad _j\|$	-0.00668	-0.00161	-0.00343	0.000466	0.00145
	(-1.38)	(-1.20)	(-0.79)	(0.32)	(0.81)
$d_i + d_j$		0.00947		0.00866	0.00866
		(12.91)		(11.95)	(11.95)
fam_{ij}			0.224	0.172	0.178
			(13.56)	(12.17)	(10.96)
$fam_{ij} \quad \|\ _i \quad _j\|$					-0.00596
					(-0.68)
constant	0.135	-0.0878	0.0933	-0.101	-0.102
	(5.67)	(-6.61)	(4.26)	(-7.92)	(-7.90)
N	71052	71052	71052	71052	71052

t statistics in parentheses

$p < 0{:}05$, $p < 0{:}01$, $p < 0{:}001$

con dence level). In contrast, controlling for degree in column 2 yields a negative and signi cant (at the 1% level). I estimate a one standard deviation di erence in risk aversion leads a 0.76 percentage point reduction in the probability of linkage.

Family connections are also a strong determinant of linkage in the risk sharing network. Across speci cations 3, 4, 5, having a family connection is positively associated with linkage in the risk sharing network (statistically signi cant at the 0.1% level). In column 5, family member dyads are 43.8 percentage points more likely to form a risk sharing relationship as non-family members.

However, in columns 4 and 5, when I control for family connection and degree, the estimate of $_1$ is once again insigni cant (and positive in column 5). However, this may speak more to the mechanism of assortative matching. Similar to Attanasio et al. (2012), we would expect assortative matching on risk aversion to play a stronger role for more socially proximate individuals, who have more information about each others preferences. In column 5, we have $_1 + {}_4 =$ 0:0184, statistically signi cant at the 5% level (^{2}(1) = 6 :61). Interpreting the coe cient, between family members a one standard deviation di erence in risk aversion reduces the probability of linkage by 18.4 percentage points.

Table 3 reports the results from re-estimating equations 21, 22, 23, 24, and 25 with the community network as the outcome. The estimates of $_1$ are low in magnitude and vary in sign. None are statistically signi cantly di erent than 0 at any standard signi cance levels. Likewise, in column 5, $_1 + {}_2 =$:005 is not signi cantly di erent than zero (^{2}(1) = 0 :34). Hence, we fail provide evidence for assortative matching in the community network regressions.

C. Subgraph Generation Models with Types

The main SUGM results are presented in tables 4 and 5. While thes are abridged for clarity, full results of all SUGM models are available in Appendix C. Using this model, I estimate that individuals who are risk averse tend to form links to each other at a rate of 4.04%. The network arising from community detection tends to be denser than the risk sharing network: I estimate that individuals who are surveyed about preferences tend to form links to each other at a rate of 9.28%, more than twice the rate in the risk sharing network.

Moving past the baseline model to the main model we derive two main ndings. First, we see further evidence of assortative matching on risk preference by less risk averse individuals. Less risk averse agents form within-type links at a rate of 5.61% (compared to the base rate of 4.04%). Second, we do not see the same kind of assortative matching when looking at more risk individuals: I estimate more risk averse individuals form within-type links at a rate of 2.99%, lower than both the base rate and the rate at which less and more risk averse individuals

Model , Subgraph	Coef.	Std. Err.
Baseline SUGM		
Within: Risk Averse	0.0404	0.0009
Preferences SUGM		
Within: Less risk averse	0.0561	0.0010
Within: More risk averse	0.0299	0.0008
Between: More, less risk averse	0.0360	0.0008

Table 4|Links and Isolates Pooled Subgraph Generation Model: Risk Sharing Network. Abridged models, sample size = 49536 dyads.

Model , Subgraph	Coef.	Std. Err.
Baseline SUGM		
Within: Surveyed	0.0928	0.0013
Preferences SUGM		
Within: Less risk averse	0.1189	0.0015
Within: More risk averse	0.0713	0.0012
Between: More, less risk averse	0.0876	0.0013

Table 5|Links and Isolates Pooled Subgraph Generation Model: Community Network. Abridged models, sample size = 49536 dyads.

form links between type (3.60%). In this way, low risk aversion individuals drive assortative matching. In contrast, more risk averse types are more likely to form between links than within links.

The assortative matching in the community network mirrors the pattern the risk sharing network. First, it is driven by less risk averse individuals, who form within links at a rate of 11.89%. Second, links between low and high risk aversion individuals form at a higher rate (8.76%) than links within high risk individuals (7.13%).

However, when making an apples to apples comparison between the degree of assortative matching in the risk sharing network and the degree in the community network, we see that the degree of assortative matching falls in the community network. Measuring the degree of assorative matching as the ratio of the rate of between links to the rate of links between all risk averse individuals, we look to Tables 6 and 7. The ratio of within for less risk averse is higher in the risk sharing network, whereas the ratio of bewtween is lower. Essentially, this indicates a reduced degree of assorative matching in the e ective risk sharing communities.

Model(s) , Subgraph	Ratio	Std. Err.
Preferences =Baseline		
Within: Less risk averse	1.389	0.0336
Within: More risk averse	0.740	0.0300
Between: More, less risk averse	0.891	0.0297

Table 6|Coefficient Ratios of Links and Isolates Pooled Subgraph Generation Model: Risk Sharing Network. Sample size = 49536 dyads.

Model(s) , Subgraph	Ratio	Std. Err.
Preferences =Baseline		
Within: Less risk averse	1.281	0.0213
Within: More risk averse	0.768	0.0192
Between: More, less risk averse	0.944	0.0198
Abridged models, total sample size = 49536 dyads		

Table 7|Coefficient Ratios of Links and Isolates Pooled Subgraph Generation Model: Community Network. Sample size = 49536 dyads.

VII. Welfare Implications of Assortative Matching

To quantify the welfare e ect of assortative matching, I compare among four scenarios. I list these scenarios here from high to low aggregate welfare:

A. The optimal scenario: The planner's optimum with equal numbers of types. No assortative matching.

B. The community scenario: takes the degree of assortative matching implied by community SUGM estimates. Some assortative matching.

C. The bilateral scenario: takes the degree of assortative matching implied by bilateral SUGM estimates. slightly more assortative matching than the community scenario

D. The worst case scenario: complete assortative matching.

Qualitatively, these scenarios are depicted in Figures 10, 11, 12, and 13.

To look at these counterfactual scenarios, I simulate type allocation using a binomial data generating process. The scenarios di er by the probability of \success" in the binomial process. Using the results from our SUGMs we are able to construct implied membership of communities. In the special case of communities, where all community members form a clique, we are able to directly estimate the ratio of coe cients. This is also useful because it can give us an analytic expression for the average proportion of the majority type each community. By construction, the majority type will be type 1 in about half of the communities and type 2 in the other half.[13] Using simplifying assumptions (covered in detail in Appendix B.B4) I am able to express the average proportion of the majority type, p^U.

$$p^U = 0{:}5 + 0{:}5\sqrt{1 - \frac{N-G}{N-1}\,\frac{\tilde{}_{L;1;2}}{\tilde{}_{ra}}} \tag{34}$$

To see a visualization of equation 34 in action, see Figures 10, 11, 12, and 13.

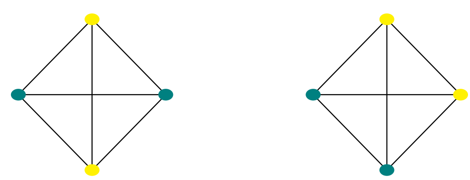

Figure 10. Scenario A, No Assortative Matching: optimal composition of risk pooling communities. Yellow is more risk averse, teal is less risk averse. $\tilde{}^{C}_{L;\,1;2} = 0{:}5$ and $p^U = 0{:}5$.

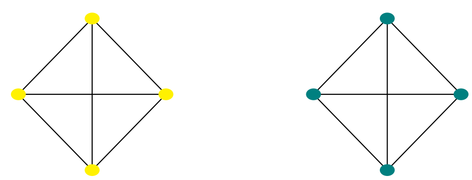

Figure 11. Scenario D, Complete Assortative Matching: a worst case composition of risk pooling communities. Yellow is more risk averse, teal is less risk averse. $\tilde{}^{C}_{L;\,1;2} = 0$ and $p^U = 1$.

Once I obtain p^U for a scenario, it becomes the basis for a simulation of com-

[13] I express this ratio as

$$\frac{\tilde{}_{1;2}}{\tilde{}_{L}} = \frac{\sum_{g=1}^{G} N_{1g}N_{2g} \Big/ N_1 N_2}{\sum_{g=1}^{G}\frac{N_g(N_g-1)}{2} \Big/ \frac{N(N-1)}{2}}: \tag{33}$$

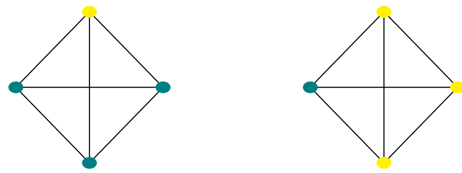

Figure 12. Scenario B, Some assortative matching: a suboptimal composition of risk pooling communities. Yellow is more risk averse, teal is less risk averse. $\tilde{\ }^{C}_{L;1;2} = 0.375$ and $p^U = 0.75$.

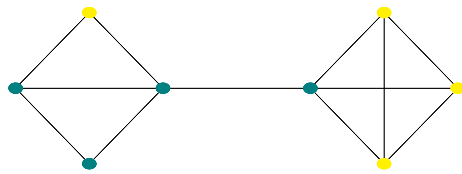

Figure 13. Scenario C, Some assortative matching in a risk sharing network. Yellow is more risk averse, teal is less risk averse. $\tilde{\ }^{C}_{L;1;2} = 0.3125$ and $p^U = 0.8061$.

munities. Each simulation proceeds as follows. First, I remove all individuals who do not have preference data, or who are not risk averse and discard resulting \communities of one." Second, I randomly sort communities into roughly equally populous type 1 or type 2 majority bins.[14] Third, after communities have been assigned to type 1 or type 2 majority bins, I simulate community membership as a N_g draws from a binomial distribution with p^U with success being de ned as a type 1 agent or a type 2 agent, respectively. Finally, I compute the value function for each community and average welfare across all four villages. I simulate community membership 50000 times, compute the value functions, and plot the results in Figure 14.

A. With no assortative matching, the optimal scenario has type 1 and type 2 agents each chosen at $.5$. The average loss per capita in this scenario is

[14] Directly minimizing the di erence in membership in type 1 and type 2 majority communities is an np-hard problem and requires a workaround. To assign communities, rst I sort the communities into a random order. I designate a bin of type 1 majority, and one for type 2 majority and a I construct a running membership sum for each bin. I add a community to bin 1 when sum 1 sum 2 and to bin 2 otherwise and proceed until all communities have been added.

	less Scenario		
Scenario	B. Community	C. Bilateral	D. Worst Case
A. Optimal	4.60	5.37	19.66
B. Community		0.76	15.06
C. Bilateral			14.29

Table 8|Differences in per capita loss from risk between scenarios. Each entry is column less row. Differences are in PPP Dollars.

136:76PPP

B. The community scenario has some assortative matching, as $R_{1;2} = 0:944$. We compute $p^U = 0:754$ under this scenario. The average loss per capita due to risk is 141:38 in this scenario.

C. The network scenario has slightly more assortative matching, as $R_{1;2} = 0:944$. We compute $p^U = 0:774$ under this scenario. The average loss due to risk is 142:13 in this scenario.

D. Finally, in the worst case scenario, there is complete assortative matching, so communities chosen as type 1 majority are completely type 1 and communities chosen as type 2 are completely type 2. The average loss due to risk is 156:43 in this scenario.

The average differences in scenarios are presented in Table 8. Due to relatively similar degrees of assortative matching in the bilateral and the community scenario, we see relatively similar degrees of welfare. However, given larger differences in the degree of assortative matching, we could see potentially large reductions in welfare. These are bounded, holding community size and risk aversion constant, by the worst case scenario.

Figure 14. Histograms of Losses from Risk from 50000 simulations: scenario means are in black optimum is green line.

VIII. Conclusion

A. Summary

The object of the paper was to characterize optimal covariate risk sharing with heterogeneous types in subvillage communities and to test that observed networks \set the table" for this type of risk sharing.

I construct a model of covariate risk sharing with heterogeneous risk preferences. In this model, agents bene t from connecting to other risk agents who have risk preferences unlike their own. In the community setting, I nd that with less and more risk averse types, the optimal allocation of types to communities re ects the population distribution of types. That is, each community should have the same roughly the same proportion of more and risk averse types as the village.

The optimal allocation gives us a situation that \sets the table" for optimal risk sharing to take place. Allocation of types to communities corresponds to assortative matching, and optimal allocation of types to communities corresponds to the case of no assortative matching. Hence, I test the hypothesis of no assortative matching using two methods, which re ect di erent dimensions of assortative matching.

Using dyadic regression, I estimate that individuals tend to match with those people who have similar degree of risk aversion. This tends to be driven by links within kinship networks. When looking at the community network, however, we do not see the same evidence of assortative matching, even within the kinshup network.

Using Subgraph Genertion Models in the risk sharing network, I nd similar results, that overall similar types tend to match with each other. However, this is driven by within links of low risk aversion types. In essence, low risk aversion types have higher degree overall but still prefer to link to their own type. In the community network, the estimates also re ect assortative matching, but with a reduction in degree.

Using the results of the Subgraph Generation models, I construct counterfactual membership in communities, and simulate welfare outcomes taking seriously the model of covariate risk sharing derived earlier. I explore four scenarios corresponding to varying levels of assortative matching: (A) an optimal scenario, (B) the community scenario, (C) the bilateral scenario, and (D) a worst case scenario. I nd large reductions in ex ante welfare due to covariate risk over a ten month panel. I estimate that on average $141.38 PPP is lost due to risk. On average, the optimal scenario averts $19.66 PPP relative to the worst case scenario. However, comparing the community scenario to the optimal scenario, I estimate the optimal scenario would avert only $4.60 PPP over the same period.

B. Limitations

It's valuable to address a few limitations which pertain largely to the estimation of the network formation models and the welfare simulations. I discuss the (lack of) causal interpretation of the the network formation models and talk through some considerations about the risk aversions.

In the estimation of network formation models, I work to handle ommited variables and pair-unobservables, by controlling for pair-degree and network density. This works to adress the impact of a third, unrelated variable causing both network connections and risk aversion. Despite this, I do not claim these network formation models are causal. In particular, I do not attempt to control reverse causality. Network position has the potential to change absolute risk aversion, especially if preferences respond to the ability to insure against risk.[15] Keeping this in mind, before considering issues with measurement of risk aversion, the estimates of the association between di erences in risk aversion accurately measure this association in equilibrium. However, a rst limitation is that these coe cients are not causal.

However, a more tricky problem is that risk aversions themselves may be underestimated or overestimated. As hypothetically framed questions, the risk aversions are not incentive compatible. It's not obvious that lack of incentive compatibility biases the risk aversion up or down, but likely would help in reducing the degree of measurement error. However, a few aspects of the survey design

[15] If we take seriously preferences as xed (as economic theory has traditionally taken them), we could take the results as casual within this story. It would be hard to take a stand towards such an assumption without bringing evidence to bear.

might still e ect this measurement. First, these estimates might be anchored by the size of the risks under consideration, which are not exceptionally large. Second, the risk aversion surveys were framed in terms of agricultural practices. Desirability bias might bias these toward being more risk averse, which tended to be the decision that would be preferred by an extension worker (using fertilizer was lower risk, for example). Third, and probably most importantly, measures of risk aversion are upper and lower bounded. About 25% of individuals who are surveyed about their preferences are as risk averse as allowed. With these under consideration, we think that these risk aversions are likely still useful. For the Subgraph Generation Models, all that matters is that the ordering of the agents is preserved. Likewise, for the dyadic regression, we should not be worried about di erences in mean or increases in variance, given the distribution of risk aversion does not change much. However, we should at least expect that due to upper boundedness, this might not be the case.

C. Future Work

Because of the novelty of the welfare simulations presented here, there seems to be room to tighten the set up to accord with the theory. While I focus on taking the theoretical model seriously, the welfare simulations still di er a bit from the planners problem. Most signi cantly, the binomial simulation approach limits the budget constraint on types. This may be okay if we are unsure about the proportion of our empirical types in the larger population, this results in a more variable distribution in welfare outcomes. This is due to di erences in the cost of risk with low risk aversion and high risk aversion.[16] At the present time, a process using sampling from the population preserving type majority communities

[16] Many low risk aversion types leads to greater average welfare.

is under consideration.

Speaking to the empirical results of the network formation models, it becomes interesting to square bilateral risk sharing preferences with e ective risk pooling groups. To do this, I can turn back to Subgraph Generation Models, which treat network structure as an emergent property of subgraph formation. For example, using a simple rule of \close matches," individuals might form chains of agents with gradients of risk preferences. These allow for agents of distant risk aversion to be present in the same community. Another possibly mechanism relies on central agents. While most agents prefer to link to agents of similar risk aversion, central agents (perhaps shopkeepers or moneylenders) may be willing to link to all types, meaning agents of various types are relatively close to many individuals within the network. In general, a mechanism where communities are more diverse than bilateral relationships depends on some emergent property.

REFERENCES

A. Ambrus, M. Mobius, and A. Szeidl. Consumption Risk-sharing in Social Networks. *American Economic Review*, 104(1):149{182, 2014. URL http://www.nber.org/papers/w15719.

O. Attanasio, A. Barr, J. C. Cardenas, G. Genicot, and C. Meghir. Risk pooling, risk preferences, and social networks. *American Economic Journal : Applied Economics*, 4(2):134{167, 2012. URL papers2://publication/uuid/6164EF9C-BD87-4E81-BB71-FC2114E70013.

F. Bloch, G. Genicot, and D. Ray. Informal insurance in social networks. *Journal of Economic Theory*, 143(143):36{58, 2008. doi: 10.1016/j.jet.2008.01.008. URL www.elsevier.com/locate/jet.

J. Blumenstock, N. Eagle, and M. Fafchamps. Airtime transfers and mobile communications: evidence in the aftermath of natural disasters. *Journal of Development Economics*, 120:157{181, 2016. ISSN 03043878. doi: 10.1016/j.jdeveco.2016.01.003. URL http://linkinghub.elsevier.com/retrieve/pii/S0304387816000109.

R. Bourles, Y. Bramoulle, and E. Perez-Richet. Altruism in Networks. *Econometrica*, 85(2):675{689, 2017. ISSN 0012-9682. doi: 10.3982/ecta13533.

A. Cabrales, A. Calvo-Armengol, and M. O. Jackson. La Crema: A case study of mutual re insurance. *Journal of Political Economy*, 111(2):425{458, 2003. ISSN 1556-5068. doi: 10.2139/ssrn.273401.

A. C. Cameron and D. L. Miller. Robust Inference for Dyadic Data. 2014. URL cameron.econ.ucdavis.edu/research[Accessed:10.09.2015].

A. G. Chandrasekhar and M. O. Jackson. A Network Formation Model Based on Subgraphs. Ssrn, 2018. doi: 10.2139/ssrn.2660381.

A. G. Chandrasekhar and R. Lewis. Econometrics of Sampled Networks. 2016.

P.-A. Chiappori, K. Samphantharak, S. Schulhofer-Wohl, and R. M. Townsend. Heterogeneity and risk sharing in village economies. Quantitative Economics, 5(1):1{27, 2014. ISSN 1759-7331. doi: 10.3982/QE131. URL http://onlinelibrary.wiley.com.ezproxy.library.wisc.edu/doi/10.3982/QE131/abstract.

S. Coate and M. Ravallion. Reciprocity without commitment insurance arrangements. Journal of Development Economics, 40:1{24, 1993.

S. Dercon, J. De Weerdt, T. Bold, and A. Pankhurst. Group-based funeral insurance in Ethiopia and Tanzania. World Development, 34(4):685{703, 2006. ISSN 0305750X. doi: 10.1016/j.worlddev.2005.09.009.

M. Fafchamps. Rural poverty, risk and development. (October):262, 2003. URL http://external.worldbankim ib.org/uhtbin/cgisirsi/x/0/0/5/?searchdata1=775749%g7Bckeyf %g7Df # gf _g.

M. Fafchamps. Risk Sharing Between Households. In Handbook of Social Economics. 2008.

M. Fafchamps and F. Gubert. Risk sharing and network formation. American Economic Review, 97(2):75{79, 2007. ISSN 0002-8282. doi: 10.1257/aer.97.2.75. URL http://www.atypon-link.com/AEAP/doi/abs/10.1257/aer.97.2.75 f %g5Cnpapers3://publication/uuid/49D0FF07-BEA7-435F-AA53-A7D77303428B.

E. Fitzsimons, B. Malde, and M. Vera-Hernandez. Group Size and the E ciency of Informal Risk Sharing. Economic Journal, 128(612):F575{F608, 2018. ISSN 14680297. doi: 10.1111/ecoj.12565.

W. Y. Gao and E. Moon. Informal Insurance Networks. B.E. Journal of Theoretical Economics, 16(2):455{484, 2016. ISSN 19351704. doi: 10.1515/bejte-2015-0006.

M. Girvan and M. E. J. Newman. Finding and evaluating community structure in networks. Cond-Mat/0308217, pages 1{16, 2003. ISSN 1063651X. doi: 10.1103/PhysRevE.69.026113. URL http://arxiv.org/abs/cond-mat/0308217 f %g 5Cnhttp://www.arxiv.org/pdf/cond-mat/0308217.pdf.

W. Jack and T. Suri. Risk Sharing and Transactions Costs : Evidence from Kenya' s Mobile Money Revolution. American Economic Review, 104(1):183{223, 2014. ISSN 00028282. doi: 10.1257/aer.104.1.183.

M. O. Jackson, T. Rodriguez-barraquer, and X. Tan. Social capital and social quilts: Network patterns of favor exchange. The American Economic Review, 102(5):1{45, 2012. ISSN 0002-8282. doi: 10.1257/aer.102.5.1857.

C. Kinnan. Distinguishing barriers to insurance in Thai villages. 2014.

E. Ligon, J. Thomas, and T. Worrall. Informal Insurance Arrangements with Limited Commitment: Theory and Evidence from Village Economies. The Review of Economic Studies, 69(1):209{244, 2002. ISSN 0034-6527. doi: 10.1111/1467-937X.00204.

M. J. Mccord, J. Roth, and D. Liber. The Landscape of Microinsurance in the World's 100 Poorest Countries. Technical Report April, 2007. URL http://www.microinsurancecentre.org/UploadDocuments/Landscapestudypaper.pdf.

Miller McPherson, Lynn Smith-Lovin, and James M. Cook. Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology 27:415{444, 2001. URL https://www.jstor.org/stable/2678628?pq-origsite=summon f &g seq=1f # gmetadataf _ginfof _gtabf _gcontents.

M. E. J. Newman. Communities, modules and large-scale structure in networks. Nature Physics, 8(1):25{31, 2011. ISSN 1745-2473. doi: 10.1038/nphys2162. URL http://dx.doi.org/10.1038/nphys2162.

P. Pons and M. Latapy. Computing communities in large networks using random walks. J. Graph Algorithms Appl., 10(2):284{293, 2004.

E. Riley. Mobile money and risk sharing against village shocks. Journal of Development Economics, 135(June):43{58, 2018. ISSN 03043878. doi: 10.1016/j.jdeveco.2018.06.015. URL https://doi.org/10.1016/j.jdeveco.2018.06.015.

T. Sargent. Macroeconomic Theory. 2nd edition, 1987.

M. Tabord-Meehan. Inference With Dyadic Data: Asymptotic Behavior of the Dyadic-Robust t-Statistic. Journal of Business and Economic Statistics, 37(4): 671{680, 2019. ISSN 15372707. doi: 10.1080/07350015.2017.1409630. URL https://doi.org/10.1080/07350015.2017.1409630.

R. M. Townsend. Risk and Insurance In Village India. Econometrica, 62(May): 539{591, 1994. ISSN 00129682. doi: 10.2307/2951659.

T. Walker. Risk coping, Social Networks, and Investment in Rural Ghana. PhD thesis, Cornell University, 2011a.

T. Walker. Technical Details on the 2009 Household Survey. Technical report, 2011b.

X. Y. Wang. Risk Sorting, Portfolio Choice, and Endogenous Informal Insurance. 2015.

J. D. Weerdt and S. Dercon. Risk-sharing networks and insurance against illness. Journal of Development Economics, 81:337{356, 2006. doi: 10.1016/j.jdeveco.2005.06.009.

Theoretical Appendix

A1. Risk Sharing in Communities

Feasibility of Risk Sharing. | Due to constraints 3, 4 and 5, budget constraints bind at the community level. To see this, we sum up the two types using weights:

$$p c_{1i} + (1 \quad p) c_{2i} \quad \left(\frac{1}{N}\sum_{i=1}^{N} \bar{y}_i + \bar{y}_v\right) + p \; {}_1$$

$$+ (1 \quad) \left(\frac{1}{N}\sum_{i=1}^{N} \bar{y}_i + \bar{y}_v\right) + (1 \quad p) \; {}_2$$

$$\frac{1}{N}\sum_{i=1}^{N} \bar{y}_i + \bar{y}_v + 0$$

$$N_1 c_{1i} + N_2 c_{2i} \quad \sum_{i=1}^{N} \bar{y}_i + N \bar{y}_V :$$

Hence total consumption shocks to types 1 and 2 are bounded by total income shocks and informal insurance is feasible.

Expected Utility. | Because shocks are normally distributed, expected utility for both types is equivalent to maximizing the mean-variance representation This can be seen in Sargent (1987).

$$E(U(c_{\ell i})) = E(c_{\ell i}) \quad \frac{\ell_i}{2} Var(c_{\ell i})$$

Also note CARA is an increasing function in consumption, so in all states of the world we consume all income and transfers available. So then we can write. Expected consumption for type 1 is $E(c_{1i}) = {}_{1i}$ and for type 2, $E(c_{2i}) = {}_{2i}$.

Variance for the two types can be computed:

$$Var(c_{1i}) = \left(\frac{\;}{p}\right)^2 \left(\frac{\;^2}{N} + \;^2\right)$$

$$Var(c_{2i}) = \left(\frac{1}{1-p}\right)^2 \left(\frac{\;^2}{N} + \;^2\right):$$

So then we write expected utility

$$E(U(c_{1i})) = \;_{1i} - \frac{\;_{1i}}{2}\left(\frac{\;}{p}\right)^2 \left(\frac{\;^2}{N} + \;^2\right)$$

and

$$E(U(c_{2i})) = \;_{2i} - \frac{\;_{2i}}{2}\left(\frac{1}{1-p}\right)^2 \left(\frac{\;^2}{N} + \;^2\right):$$

For ease of notation, we de ne

$$\;^2_c = \frac{\;^2}{N} + \;^2:$$

and note that the utility of the the more risk averse agents when only idiosyncratic risk is pooled is equal to

$$EU_0 = -\frac{\;_{2i}}{2}\left(\frac{1}{N}\sum_{i=1}^{N} y_i + y_v\right):$$

Solving the Lagrangian. | We construct the Lagrangian retaining constraints 2 and 5 (with a_2 and a_3 as multipliers, respectively) and incorporate the consumption constraints into expected utlity.

$$L = \;_1 \quad \frac{1}{2}\frac{^2}{p^2}\;{}^2_c + a \quad {}_2 \quad \frac{^2}{2}\frac{(1 \quad)^2}{(1 \quad p)^2}\;{}^2_c + \frac{^2}{2}\;{}^2_c$$

$$+ \; b(p \;_1 + (1 \quad p) \;_2)$$

The rst order conditions are as follows:

(A1) $$\frac{@L}{@_1} = 1 + bp = 0$$

(A2) $$\frac{@L}{@_2} = a + b(1 \quad p) = 0$$

(A3) $$\frac{@L}{@} = \frac{_1 \quad {}^2_c}{p^2} + a_2 \quad \frac{_2(1 \quad) \; {}^2_c}{(1 \quad p)^2}$$

(A4) $$\frac{@L}{@a} = \;_2 \quad \frac{^2}{2} \; \frac{1}{1 \quad p}^{\;2} \; {}^2_c + \frac{^2}{2} \; {}^2_c = 0$$

(A5) $$\frac{@L}{@b} = p \;_1 + (1 \quad p) \;_2 = 0$$

Using FOC A1 we note that $b = \frac{1}{p}$. Likewise, using FOC A2 we note that $a = \frac{1 \quad p}{p}$. Rearranging FOC A4,

$$_2 = \frac{^2}{2} \quad 1 \quad \frac{1}{1 \quad p}^{\;2}!$$

We rearrange FOC A5 and substitute in FOC A4:

$$_1 = \frac{1 \quad p}{p} \;_2 = \frac{1 \quad p}{p}\frac{^2}{2} \quad 1 \quad \frac{1}{1 \quad p}^{\;2}!$$

Finally, we simplify FOC A3 to nd :

$$\frac{\ _1\ \ ^2_c}{p^2} = \frac{1\ \ p}{p}\ \ \frac{\ _1(1\ \ \ )\ ^2_c}{(1\ \ p)^2}$$

$$\frac{\ _1}{\ _2}\ \ \frac{1\ \ p}{p} = \frac{1}{\ \ }$$

$$\frac{1}{\ } = \frac{\ _1}{\ _2}\ \ \frac{1\ \ p}{p} + 1$$

$$= \frac{p\ \ _2}{(1\ \ p)\ \ _1 + p\ \ _2}$$

Thus only if either $\ _1 = 0$ (type 1 is risk neutral, which we've assumed is not true) or $p = 1$, $\ = 1$. Hence, covariate risk will not taken on fully by the less risk averse agents. Note that

$$(1\ \ \ )^2 = \ \ 1\ \ \frac{p\ \ _2}{(1\ \ p)\ \ _1 + p\ \ _2}^{\ 2}$$

$$= \ \ 1\ \ \frac{(1\ \ p)\ \ _1}{(1\ \ p)\ \ _1 + p\ \ _2}^{\ 2}$$

$$= \frac{(1\ \ p)^2\ \ ^2_1}{((1\ \ p)\ \ _1 + p\ \ _2)^2}$$

So then

$$\ _2 = \ \ \frac{\ ^2}{2}\ \ 1\ \ \frac{\ ^2_1}{((1\ \ p)\ \ _1 + p\ \ _2)^2}$$

Value Functions. | We compute the value functions for type 1 and type 2 individuals.

$$V_1(p;\ \ _1;\ \ _2) = E(U(c_1 i) j\ \ (p);\ \ _1(p))$$

$$= \quad {}_1 \quad \frac{1}{2} \quad \frac{(p)}{p}^{2} \quad {}^{2}_{c}$$

$$= \quad {}_1 \quad \frac{1}{2} \quad \frac{p \;{}_2}{((1 \quad p)\;{}_1 + p\;{}_2)p} \quad {}^{2}_{c}$$

$$= \quad {}_1 \quad \frac{1}{2} \quad \frac{{}_2}{((1 \quad p)\;{}_1 + p\;{}_2)} \quad {}^{2}_{c}$$

$$= \frac{1}{2} \quad \frac{1 \quad p}{p} \quad 1 \quad \frac{{}_1}{(1 \quad p)\;{}_1 + p\;{}_2}^{2}! \quad \frac{1}{2} \quad \frac{{}_2}{(1 \quad p)\;{}_1 + p\;{}_2}^{2} \quad {}^{2}_{c}$$

$$V_2(p; \;{}_1; \;{}_2) = E(U(c_2 i)j \quad (p); \;{}_2(p))$$

$$= \quad {}_2(p) \quad \frac{{}_2}{2} \quad \frac{1 \quad (p)}{1 \quad p}^{2}$$

$$= \quad {}_2(p) \quad \frac{{}_2}{2} \quad \frac{1 \quad \frac{p\;{}_2}{(1 \quad p)\;{}_1 + p\;{}_2}}{1 \quad p}^{!\;2} \quad {}^{2}_{c}$$

$$= \quad {}_2(p) \quad \frac{{}_2}{2} \quad \frac{(1 \quad p)\;{}_1 + p\;{}_2 \quad p\;{}_2}{(1 \quad p)((1 \quad p)\;{}_1 + p\;{}_2)}^{2} \quad {}^{2}_{c}$$

$$= \quad {}_2(p) \quad \frac{{}_2}{2} \quad \frac{(1 \quad p)\;{}_1 + p\;{}_2 \quad p\;{}_2}{(1 \quad p)((1 \quad p)\;{}_1 + p\;{}_2)}^{2} \quad {}^{2}_{c}$$

$$= \quad {}_2(p) \quad \frac{{}_2}{2} \quad \frac{(1 \quad p)\;{}_1}{(1 \quad p)((1 \quad p)\;{}_1 + p\;{}_2)}^{2} \quad {}^{2}_{c}$$

$$= \quad {}_2(p) \quad \frac{{}_2}{2} \quad \frac{{}_1}{(1 \quad p)\;{}_1 + p\;{}_2}^{2} \quad {}^{2}_{c}$$

$$= \quad \frac{{}_2}{2} \quad 1 \quad \frac{{}^{2}_{1}}{((1 \quad p)\;{}_1 + p\;{}_2)^2} \quad \frac{{}_2}{2} \; ! \quad \frac{{}_1}{(1 \quad p)\;{}_1 + p\;{}_2}^{2} \quad {}^{2}_{c}$$

$$= \frac{{}_2}{2} \quad 1 + \quad \frac{{}_1}{(1 \quad p)\;{}_1 + p\;{}_2}^{2} \quad 1 + \quad {}^{2}_{c}$$

More. |

$$p\ _2 + (1 \quad p)\ _1 < p\ _2 + (1 \quad p)\ _2$$
$$= \ _2$$

so then

$$\frac{1}{p\ _2 + (1 \quad p)\ _1} > \frac{1}{_2}$$
$$\frac{p\ _2}{p\ _2 + (1 \quad p)\ _1} > \frac{p\ _2}{_2} = p$$

A2. Planner's Problem

Some useful equations:

(A6)
$$\frac{@ V_1(q)}{@ q} = \quad \frac{1 \quad q}{1} \quad {}^2_1 + \quad {}^2_c\ {}^2_2 \quad \frac{(\ _2 \quad _1)\ _2}{((1 \quad q)\ _1 + q\ _2)^3} \quad \frac{^2_2}{2q^2} \quad 1 \quad \frac{_1}{(1 \quad q)\ _1 + q\ _2} \quad {}_2 !$$

(A7)
$$\frac{@ V_2(q)}{@ q} = \quad \frac{2\ _1(\ _2 \quad _1)}{((1 \quad q)\ _1 + q\ _2)^3} \quad {}^2_c \quad 1$$

(A8)
$$\frac{@ V_2(1 \quad q)}{@ q} = \quad \frac{2\ _1(\ _2 \quad _1)}{(q\ _1 + (1 \quad q)\ _2)^3} \quad {}^2_c \quad 1$$

(A9)

$$\frac{@V_1(q)}{@q} =$$

$$\frac{2}{2}\left(\frac{1}{q^2}\quad 1 \quad \frac{1}{(1 \quad q)\ _1 + q\ _2}\right)^2 + \frac{1 \quad q}{q}\left(\frac{2\ _1^2(\ _2 \quad _1)}{((1 \quad q)\ _1 + q\ _2)^3}\right.$$

$$\left.\frac{2}{2}\ \frac{2\ _2^2(\ _2 \quad _1)}{((1 \quad q)\ _1 + q\ _2)^3}\ _c^2\right)$$

(A10)

$$= \frac{1 \quad q}{1}\ _1^2\ \frac{(\ _2 \quad _1)\ _2}{((1 \quad q)\ _1 + q\ _2)^3} + \ _c^2\ _2^2\ \frac{(\ _2 \quad _1)\ _2}{((1 \quad q)\ _1 + q\ _2)^3}$$

$$\frac{_2^2}{2q^2}\left(1 \quad \frac{1}{(1 \quad q)\ _1 + q\ _2}\right)^2$$

(A11)

$$= \frac{1 \quad q}{1}\ _1^2 + \ _c^2\ _2^2\ \frac{(\ _2 \quad _1)\ _2}{((1 \quad q)\ _1 + q\ _2)^3} \quad \frac{_2^2}{2q^2}\left(1 \quad \frac{1}{(1 \quad q)\ _1 + q\ _2}\right)^2$$

Empirical Appendix

B1. Data

B2. Community Detection

Other Approaches. | While the walktrap algorithm features good properties for community detection, it is by no means the only option. One appealing approach to assign households to communities is edge betweenness community detection. This algorithm is appealing because it takes advantage of information bottlenecks in networks, connecting naturally to the types of asymmetric information problems that constrain risk sharing (Girvan and Newman, 2003). For more on edge betweenness, see B.B2. For large networks (like mobile money transaction networks) there may be better methods from the perspective of computation speed. However, detected communities from a given algorithm may not scale well to larger networks if community size grows with network size. An iterated min cut algorithm may fall closer to the conditions that create ex post risk sharing islands Ambrus et al. (2014).

Edge Betweenness. | Another intuitive approach to finding communities hinges on information in networks. Edge betweenness measures how central an edge is in a network by counting which nodes it lies between. In particular, if edge kl lies on the shortest path from i to j edge kl is awarded a unit of edge betweeness. Summing up all of these awards from all pairs of nodes i and j, we get the edge betweeness for edge kl. We can think of edges with high betweenness as relationships in the risk sharing network where both individuals knowledge of the other individual and the individuals beyond them comes directly through that individual. There is a great deal of potential to broker information over these links but because of frictions due to adverse selection, hidden income, and various types of moral hazard, individuals will be hesitant to share risk with those who they do not have good information about.

The algorithm is named because it calculates the betweenness centrality of all edges in the network and deletes those edges with highest centrality. Edge betweenness centrality is a measure of how central an edge is in a network based on who relies on that edge connect to other portions of the network. In order to compute this measure, compute all shortest paths between nodes on a network. Then, count of the number of shortest paths passing through each edge of interest (in the case multiple paths tie for shortest path for two nodes, a partial count is awarded across the edges in these paths). Intuitively, betweenness centrality is often used as a measure the potential for brokerage. In this case, we can think of information flow being costly in the link that has high betweenness centrality because of the gatekeeper on each side. This would create a specific rationale for these communities as risk sharing units, then. One issue with edge betweenness is that weighted versions of the algorithm may not interpret weights in an intuitive manner for our application. In particular larger capacity connections are lower cost and hence have more least cost paths contributing to betweenness. This will cause these edges to be cut early. Initially I will solve this by supplying inverse weights to the algorithm. At the start of the algorithm, all nodes in a particular component are assigned to the same community. Every time a component breaks into two with the deletion of an edge, the communities membership is reassigned for the nodes in the broken component. After initially computing all edge betweenness centralities, the algorithm works in two steps:

1) Delete the edge with the highest betweenness centrality
2) Recompute the betweenness centrality of remaining edges

This two step process continues until all edges are deleted and thus all nodes reside in their own component and hence community. To choose a final community assignment, the algorithm cuts the dendrogram using a tuning statistic. For this, I have two possible solutions. First, the off the shelf choice in network science is modularity, which measures the internal quality of communities assigned (see Appendix B.B2 for details). The algorithm cuts of the dendrogram at the assignment with the maximum value of the tuning statistic (Girvan and Newman, 2003). In previous unpublished work, I used edge betweenness community detection to detect risk sharing islands using data from southern India. Alternatively, a regression coefficient measuring the degree of risk sharing in the community assignment could be used to tune the assignment. I leave discussion of this second option until tests of risk sharing have been discussed.

Modularity. — To compute modularity let k_i and k_j be the degrees of nodes i and j, respectively. Let m be the number of edges in the graph. The expected number of edges between i and j from this rewiring is equal to $k_i k_j/(2m - 1) \approx k_i k_j/2m$ ($2m$ since each link has two "stubs", so to speak). I can then compare this expected number of links between i and j to the actual connections: letting A_{ij} be the ijth entry of the matrix, I take the difference these two numbers:

$$A_{ij} - \frac{k_i k_j}{2m}$$

I can interpret this as connections over expected conditional on node pair degrees. Then, letting c_i be the community membership of node i, connections over expectation are weighted by the function δ:

$$\delta(c_i; c_j) = \begin{cases} 1 & \text{if } c_i = c_j \\ 0 & \text{otherwise} \end{cases}$$

Finally I aggregate to the graph level and normalize by twice the number of links present:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i; c_j)$$

This serves as an easily computable and straightforward measure of the internal quality of communities (Newman, 2011).

B3. Approximation of Variance of Ratios

We want the ratio of the variance of two coefficients $\tilde{\beta}_{L;s}$ and $\tilde{\beta}_{L;ra}$, (B1) $$Var\left(\frac{\tilde{}_{L;s}}{\tilde{}_{L;ra}}\right) = \left(\frac{\tilde{}_{L;s}}{\tilde{}_{L;ra}}\right)^2 \left(\frac{(\ _s)^2}{(\tilde{}_{L;s})^2} \quad \frac{2Cov(\tilde{}_{L;s};\tilde{}_{L;ra})}{\tilde{}_{L;s}\ \tilde{}_{L;s}} + \frac{{}^2_{ra}}{\tilde{}^2_{L;ra}}\right)$$

Given that the two coefficients derive from a similar data generating process and measure a similar quantity, it is intuitive that $Cov(\tilde{}_{L;s};\tilde{}_{L;ra}) > 0$. My priors are that the correlations between these two coefficients would be close to one. Maintaining this assumption, it is conservative to estimate the variance of the ratio by assuming $Cov(\tilde{}_{L;s};\tilde{}_{L;ra}) = 0$. This assumption leaves us with

(B2) $$Var\left(\frac{\tilde{}_{L;s}}{\tilde{}_{L;ra}}\right) = \left(\frac{\tilde{}_{L;s}}{\tilde{}_{L;ra}}\right)^2 \left(\frac{(\ _s)^2}{(\tilde{}_{L;s})^2} + \frac{{}^2_{ra}}{\tilde{}^2_{L;ra}}\right)$$

B4. Welfare Simulations

Rate of Between Link Generation. | When we want to look at how many connections there are between types in communities, we can look at complete bipartite graphs for counts. A complete bitartite graph with N_{1g} of type 1 and N_{2g} of type 2, will have $N_{1g}N_{g2}$ connections. Thus the total number of actual connections between types within communities is

(B3) $$\sum_{g=1}^{G} N_{1g}N_{2g}:$$

Additionally, the total number of potential links between types in the entire village graph will be

$$\text{(B4)} \qquad \left(\sum_{g=1}^{G} N_{1g}\right)\left(\sum_{g=1}^{G} N_{2g}\right) = N_1 N_2$$

So then

$$\text{(B5)} \qquad \sim_{1,2} = \frac{\sum_{g=1}^{G} N_{1g}N_{2g}}{N_1 N_2}$$

We assume equal parts type 1 and type 2 agents, which we impose empirically as well, so then $N_1 = N_2$ and $N_1 + N_2 = N$ so $N_1 = N_2 = \frac{N}{2}$

$$\text{(B6)} \qquad \sim_{1,2} = \frac{\sum_{g=1}^{G} N_{1g}N_{2g}}{\frac{N^2}{2^2}} = \frac{4\sum_{g=1}^{G} N_{1g}N_{2g}}{N^2}$$

We then distribute the $\frac{1}{N^2}$ through, so we have

$$\text{(B7)} \qquad \sim_{1,2} = 4\sum_{g=1}^{G} \frac{N_{1g}}{N}\frac{N_{2g}}{N}$$

Recall that $p_{\ell g} = \frac{N_{\ell g}}{N_g}$. (How does N_g relate to N?)

$$\text{(B8)} \qquad \sim_{1,2} = 4\sum_{g=1}^{G} \frac{N_g p_{1g}}{N}\frac{N_g p_{2g}}{N}$$

$$\text{(B9)} \qquad \sim_{1,2} = 4\sum_{g=1}^{G} \left(\frac{N_g}{N}\right)^2 p_{1g}p_{2g}$$

Here we make the (unrealistic) simplifying assumption that community sizes are the same, hence there's a fixed $\frac{N_g}{N} = \frac{1}{G}$.

$$\tilde{\rho}_{1;2} = \frac{4}{G^2} \sum_{g=1}^{G} p_{1g} p_{2g} \tag{B10}$$

Finally, fix $p_{1g} = p^U$ and $p_{2g} = p^L$ when $p_{1g} \geq p_{2g}$ and vice-versa when $p_{1g} < p_{2g}$ where $p^U = 1 - p^L$.

$$\tilde{\rho}_{1;2} = \frac{4}{G^2} \sum_{g=1}^{G} p^U p^L \tag{B11}$$

We sum across groups and then rearrange:

$$\tilde{\rho}_{1;2} = \frac{4}{G} p^U p^L \tag{B12}$$

Rate of Within Risk Averse Link Generation. — The total number of potential links generated is

$$\frac{N(N-1)}{2} \tag{B13}$$

With completely connected communities, the number of connections ends up being

$$\frac{\sum_{g=1}^{G} N_g (N_g - 1)}{2} \tag{B14}$$

So then

$$\tilde{\ }_L = \frac{\frac{\sum_{g=1}^{G} N_g(N_g-1)}{2}}{\frac{N(N-1)}{2}} = \frac{\sum_{g=1}^{G} N_g(N_g - 1)}{N(N - 1)} \tag{B15}$$

Suppose, as above, that $N_g = \frac{N}{G}$. Then

$$\tilde{\ }_L = \frac{\sum_{g=1}^{G} \frac{N}{G}(\frac{N}{G} - 1)}{N(N - 1)} \tag{B16}$$

$$= \frac{N(\frac{N}{G} - 1)}{N(N - 1)} \tag{B17}$$

$$= \frac{(\frac{N}{G} - 1)}{(N - 1)} \tag{B18}$$

$$= \frac{(N - G)}{G(N - 1)} \tag{B19}$$

(B20)

Ratio of Rates. — Based on this, we can express the ratio of the link generation coefficients as an expression relating the proportion in each community to the rate of generation.

$$\frac{\tilde{\ }_{1;2}}{\tilde{\ }_L} = \frac{\frac{4}{G}p^U p^L}{\frac{(N-G)}{G(N-1)}} \tag{B21}$$

$$= 4\frac{(N - 1)}{(N - G)}p^U p^L \tag{B22}$$

Hence we write:

(B23)
$$p^U p^L = \frac{1}{4}\left(\frac{N-G}{N-1}\right)\left(\frac{\tilde{\ }_{1,2}}{\tilde{\ }_L}\right)$$

The RHS of the equation lies between 0 and $\frac{1}{4}$. Note that as N becomes large, $\frac{N-G}{N-1} \to 1$, but that this kind of small sample correction does account for the fact that between connections make up a larger share of connections than within connections when loops are ommited. Another way to think of this is when sampling pairs, sampling without replacement only matters when sampling pairs within a type. We can solve the above by using a system of equations where $p^U + p^L = 1$, and use the quadratic formula to get an analytic solution:

(B24)
$$(p^U, p^L) = 0.5 \pm 0.5\sqrt{1 - \left(\frac{N-G}{N-1}\right)\left(\frac{\tilde{\ }_{1,2}}{\tilde{\ }_L}\right)}$$

Additional Tables

Table C1—Links and Isolates Pooled Subgraph Generation Model: Risk Sharing Network with Surveyed and Non-surveyed Network Members.

Feature	Count	Potential	Sample size	Coef.	Std. Err.
Isolates:					
Not surveyed	36	96	631	0.375	0.0193
Surveyed	54	535	631	0.1009	0.012
Within links:					
Not surveyed	20	1133	49536	0.0177	0.0006
Surveyed	1367	35526	49536	0.0385	0.0009
Between links:					
Surveyed, not	221	12877	49536	0.0172	0.0006

Table C2—Links and Isolates Pooled Subgraph Generation Model: Community Network with Surveyed and Non-surveyed Network Members.

Feature	Count	Potential	Sample size	Coef.	Std. Err.
Isolates:					
Not surveyed	39	96	631	0.4062	0.0196
Surveyed	77	535	631	0.1439	0.014
Within links:					
Not surveyed	42	1133	49536	0.0371	0.0008
Surveyed	3108	35526	49536	0.0875	0.0013
Between links:					
Surveyed, not	703	12877	49536	0.0546	0.001

Table C3—Pooled Subgraph Generation Model with Links and Isolates and Type: Surveyed. Network size = 631.

Feature	Count	Potential	Sample size	Coef.	Std. Err.
Isolates:					
Surveyed	50	535	631	0.0935	0.0116
Not surveyed	35	96	631	0.3646	0.0192
Within links:					
Surveyed	1535	35526	49536	0.0432	0.0009
Not surveyed	20	1133	49536	0.0177	0.0006
Between links:					
Surveyed, not	247	12877	49536	0.0192	0.0006

Table C4—Pooled Subgraph Generation Model using Community Graph with Links and Isolates and Type: Surveyed. Network size = 631.

Feature	Count	Potential	Sample size	Coef.	Std. Err.
Isolates:					
Surveyed	71	535	631	0.1327	0.0135
Not surveyed	39	96	631	0.4062	0.0196
Within links:					
Surveyed	3411	35526	49536	0.096	0.0013
Not surveyed	38	1133	49536	0.0335	0.0008
Between links:					
Surveyed, not	653	12877	49536	0.0507	0.001

Table C5—Pooled Subgraph Generation Model with Links and Isolates and Type: Preferences. Network size = 631.

Feature	Count	Potential	Sample size	Coef.	Std. Err.
Isolates:					
Less risk averse	21	236	631	0.089	0.0113
More risk averse	19	217	631	0.0876	0.0113
Risk loving	10	82	631	0.122	0.013
Not surveyed	35	96	631	0.3646	0.0192
Within links:					
Less risk averse	471	7511	49536	0.0627	0.0011
More risk averse	203	6223	49536	0.0326	0.0008
Risk Loving	36	814	49536	0.0442	0.0009
Not surveyed	20	1133	49536	0.0177	0.0006
Between links:					
Less, more risk averse	478	11738	49536	0.0407	0.0009
Less risk averse, risk loving	192	4765	49536	0.0403	0.0009
More risk averse, risk loving	155	4475	49536	0.0346	0.0008
Less risk averse, not surveyed	146	5879	49536	0.0248	0.0007
More risk averse, not surveyed	75	5057	49536	0.0148	0.0005
Risk loving, not surveyed	26	1941	49536	0.0134	0.0005

Table C6—Pooled Subgraph Generation Model using Community Graph with Links and Isolates and Type: Preferences. Network size = 631.

Feature	Count	Potential	Sample size	Coef.	Std. Err.
Isolates:					
Less risk averse	36	236	631	0.1525	0.0143
More risk averse	21	217	631	0.0968	0.0118
Risk loving	14	82	631	0.1707	0.015
Not surveyed	39	96	631	0.4062	0.0196
Within links:					
Less risk averse	991	7511	49536	0.1319	0.0015
More risk averse	475	6223	49536	0.0763	0.0012
Risk Loving	64	814	49536	0.0786	0.0012
Not surveyed	38	1133	49536	0.0335	0.0008
Between links:					
Less, more risk averse	1171	11738	49536	0.0998	0.0013
Less risk averse, risk loving	379	4765	49536	0.0795	0.0012
More risk averse, risk loving	331	4475	49536	0.074	0.0012
Less risk averse, not surveyed	356	5879	49536	0.0606	0.0011
More risk averse, not surveyed	229	5057	49536	0.0453	0.0009
Risk loving, not surveyed	68	1941	49536	0.035	0.0008

Table C7—Links and Isolates Pooled Subgraph Generation Model: Risk Sharing Network with Risk Preferences

Feature	Count	Poten.	S. size	Coef.	Std. Err.
Isolates:					
Less risk averse	22	236	631	0.0932	0.0116
More risk averse	19	217	631	0.0876	0.0113
Risk loving	13	82	631	0.1585	0.0145
Not surveyed	36	96	631	0.375	0.0193
Within links:					
Less risk averse	421	7511	49536	0.0561	0.001
More risk averse	186	6223	49536	0.0299	0.0008
Risk loving	33	814	49536	0.0405	0.0009
Not surveyed	20	1133	49536	0.0177	0.0006
Between links:					
Less, more risk averse	423	11738	49536	0.036	0.0008
Less risk averse, risk loving	163	4765	49536	0.0342	0.0008
More risk averse, risk loving	141	4475	49536	0.0315	0.0008
Less risk averse, not surveyed	132	5879	49536	0.0225	0.0007
More risk averse, not surveyed	66	5057	49536	0.0131	0.0005
Risk loving, not sureveyed	23	1941	49536	0.0118	0.0005

Table C8—Links and Isolates Pooled Subgraph Generation Model: Community with Risk Preferences

Feature	Count	Poten.	S. size	Coef.	Std. Err.
Isolates:					
Less risk averse	38	236	631	0.161	0.0146
More risk averse	22	217	631	0.1014	0.012
Risk loving	17	82	631	0.2073	0.0161
Not surveyed	39	96	631	0.4062	0.0196
Within links:					
Less risk averse	893	7511	49536	0.1189	0.0015
More risk averse	444	6223	49536	0.0713	0.0012
Risk Loving	59	814	49536	0.0725	0.0012
Not surveyed	42	1133	49536	0.0371	0.0008
Between links:					
Less, more risk averse	1028	11738	49536	0.0876	0.0013
Less risk averse, risk loving	373	4765	49536	0.0783	0.0012
More risk averse, risk loving	311	4475	49536	0.0695	0.0011
Less risk averse, not surveyed	379	5879	49536	0.0645	0.0011
More risk averse, not surveyed	248	5057	49536	0.049	0.001
Risk loving, not surveyed	76	1941	49536	0.0392	0.0009

Additional Figures

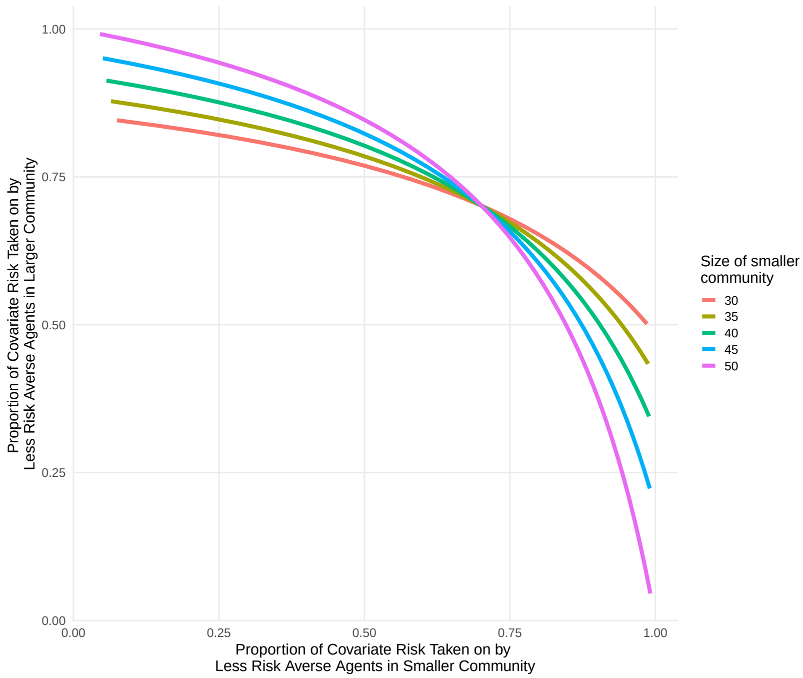

Figure D1. A Risk Management Frontier: Proportion of Covariate Risk Taken on by Less Risk Averse Agents in Communities. From top left to bottom right, type 1 agents move from larger community to smaller.