Sharing Covariate Risk in Networks: Theory and Evidence from Ghana

Daniel S. Putman*

This Version: May 31, 2024 Click Here For Most Current Version

Abstract

When risk preferences are heterogeneous, welfare can be improved by shifting covariate shocks from risk averse to risk tolerant people in exchange for a premium. However, this type of risk pooling depends on whether people prefer to share risk with others who have similar risk preferences. To investigate this question, I use detailed data from Ghana to construct village risk sharing networks and community detection to construct detected insurance groups—which bound the scope of risk pooling. With econometric models of network formation, I estimate a preference to match on risk preferences in risk sharing networks. Within detected insurance groups, the magnitude of assortative matching falls considerably. I build a theoretical model of risk pooling with heterogeneous risk preferences where a planner allocates individuals to optimal risk pooling groups. I compare this allocation of types to three benchmarks, including an optimal and worst-case scenario, finding that the observed networks deviate only slightly from optimal networks for this form of risk pooling.

Keywords: Risk Sharing, Network Formation, Assortative Matching, Risk Preferences, Community Detection

JEL Codes: D85, G52, L14, O12, O17, Z13

^{*}Postdoctoral Research Fellow, University of Pennsylvania Center for Social Norms and Behavioral Dynamics (putmands@gmail.com). This paper has benefited immensely from conversations with Travis Lybbert, Ashish Shenoy, Laura Meinzen-Dick, Thomas Walker, Michael R. Carter, Tor Tolhurst, Arman Rezaee, and Dalia Ghanem as well as comments from Steve Boucher, Jamie Hansen-Lewis, attendees at PacDev, The Conference on Network Science in Economics, several seminars in the UC Davis Development Workshop, Development Tea, and Econometrics Reading Group, and excellent research assistance from Luke Gooding. All errors herein are my own.

1 Introduction

The economic position of the rural poor is precarious, vulnerable to losses from both idiosyncratic and covariate shocks (Ligon and Schechter, 2003; Günther and Harttgen, 2009; Collins et al., 2010). Idiosyncratic risks include shocks like illness, loss of employment, and theft, or the loss of a family member, that are uncorrelated between individuals or households in localities. In contrast, covariate risks like output price and weather shocks are correlated among these individuals or households. Despite the recent adoption of digital financial services in some markets, risk management tools to manage such risks are still missing for many (Demirguc-Kunt et al., 2018). This fact may prevent risk taking which would result in higher incomes over the long term (Elbers et al., 2007; Karlan et al., 2014). In the absence of formal financial markets, informal risk sharing, mediated through social networks, is a common and important method of managing risk (Fafchamps and Lund, 2003; Comola and Fafchamps, 2017).

The classic story of informal insurance is as follows: two people are seeking to insure their consumption against idiosyncratic risks. If you lose your job, I pay you; If I lose my job, you pay me. Evidence is often consistent with a high degree of idiosyncratic risk being insured, even in light of information asymmetries (Kinnan, 2021). In contrast, insurance of covariate risks is less well explored. When risk preferences are heterogeneous, informal insurance of covariate risk can lead to welfare improvements by shifting risks from more risk averse to less risk averse agents in exchange for a premium (Chiappori et al., 2014).¹ In this story of informal insurance, the less risk averse agent takes the hit in a bad year; 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. In essence, less risk averse agents become local insurance companies for their peers.²

This story of covariate risk sharing, however, depends critically on the proximity of less and more risk averse agents in risk sharing networks. In contrast, there is a tendency to connect to those similar to oneself in social and economic networks (McPherson et al., 2001). This pattern of positive *assortative matching* on risk preferences—or the tendency of those with similar risk preferences to be connected in networks—arises as a barrier to covariate risk sharing (Attanasio et al., 2012). Given the degree of assortative matching on risk preferences found in real world

¹With the adoption of digital payment systems in recent years, it is important to delineate this story of covariate risk sharing from digitally mediated inter-village risk sharing, which might also help cope with locally covariate risk (Jack and Suri, 2014). For those who have adopted mobile money, what we think of as covariate shocks (droughts, flooding, earthquakes) may become idiosyncratic (Blumenstock et al., 2016; Riley, 2018).

²While this paper abstracts away from the specific transactions that might allow for covariate risk sharing, concrete notions of suitable arrangements can be found in the literature. For example, the literature on sharecropping places sharecropping as a way for a more risk averse renter to pass risk to their less risk averse landlord (Stiglitz, 1974; Braverman and Stiglitz, 1986). Similarly, renters would need to be more risk averse than landlords (Allen and Lueck, 1995). Sharecropping is relatively common in the context at hand, accounting for about 50% of rental contracts (Goldstein and Udry, 2008).

risk sharing networks, what quality of insurance can covariate risk sharing deliver? Empirically, I study this question by asking if individuals form connections with others who have similar or different risk preferences. In studying this question, I seek to understand how both the formation of risk-sharing network connections and local network structure (more broadly) constrain covariate risk sharing.

Using econometric models of network formation, I estimate the degree to which agents match with those who have similar risk preferences. To measure the degree of assortative matching on risk preferences, I use rich microdata featuring income shocks, network ties, and risk preferences from a survey of four villages in rural southern Ghana (Barrett, 2009). This setting features prominent correlated risk and the data includes a detailed social networks module and a set of hypothetical gambles. I measure the risk sharing network in two ways. First, I construct a risk sharing network using network trust and past gifts.³ Second, to capture local risk sharing that might extend beyond direct network connections, I use *community detection*—clustering methods which are sensitive to the details of network structure (Pons and Latapy, 2005; Newman, 2012)—to extract tightly knit groups of individuals. In previous work, co-membership in such detected insurance groups predict membership in experimental insurance groups (Putman, 2020). Likewise, previous work in this context estimates complete risk sharing among socially visible members of the network, suggesting that broader network position may matter (Vanderpuye-Orgle and Barrett, 2009). I back out risk preferences using hypothetical gambles, focusing on the coefficient of absolute risk aversion as this allows for comparisons of variance in income to average losses.

I use estimates from several econometric models of network formation to characterize assortative matching. These models use differences in risk preferences to explain connections in the risk sharing and detected insurance groups. Dyadic regression, which treats dyads of individuals as the unit of study, serves as a reduced form approach to estimating assortative matching in risk preferences (Graham, 2020). Using this model, 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.⁴ Additionally, I find they are robust to alternative network formation models and modeling choices, including Subgraph Generation Models (SUGMs) (Chandrasekhar and Jackson, 2021). In addition to estimating SUGMs, I estimate tetrad logit, a model designed to account for degree heterogeneity (Graham, 2017), where I find consistent results. The results are also robust to alternative specification choices particularly

³Despite not using a direct measure of risk sharing links, I find evidence consistent with the idea that these networks are used for risk sharing as opposed to more prosaic favor exchange. This analysis uses risk sharing transfers (gifts and loans) during the period of the study, using researcher run lotteries as shocks.

⁴My preferred results are conditional on controlling for the sum of risk aversion, though differences are small. Controlling for the sum of risk aversion so solves a subtle omitted variable problem by controlling for a correlation between popularity and risk aversion.

estimation as a logistic regression and controlling for a large set of dyadic characteristics.⁵

As I move from risk sharing network to co-membership in detected insurance groups, assortative matching on risk preferences falls. In particular, using dyadic regression, I fail to find evidence for assortative matching in detected insurance groups.⁶ Likewise, when estimating SUGMs, I find that the magnitude of assortative matching is attenuated in the detected insurance groups *vis a vis* the risk sharing network. In other words, detected insurance groups feature more diverse matching of risk preferences than risk sharing networks.

What are the welfare impacts of this degree of assortative matching? To translate my estimates of assortative matching into concrete welfare estimates, I construct a theoretical model of optimal risk sharing in a village setting. In this model, I abstract away from the question of matching to focus on how a planner allocates individuals subvillage groups.⁷ While idiosyncratic risk is assumed to be fully pooled at the group level, covariate risk is not. The social planner assigns individuals to two risk pooling groups according to their risk aversion in order to optimally share covariate risk. According to this model, optimal risk pooling happens when the composition of the groups reflects the composition of the village with respect to risk aversion. 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. Consistent with the intuition of the empirical work, this result implies that optimal risk pooling should feature no assortative matching on risk preferences.

I divide individuals into more and less risk averse types and quantify the welfare implications of their allocation of types in insurance groups. To do this, I simulate four scenarios which vary by their degree of assortative matching (and therefore optimality): (a) an optimal scenario (i.e., with no assorative matching) (b) detected insurance groups (i.e., with assortative matching implied by the detected insurance groups), (c) risk sharing networks (i.e., assorative matching as implied by risk sharing networks), and (d) a worst case scenario (i.e, with complete assortative matching). (a) and (d) are determined by the theoretical model derived earlier, while (b) and (c) derive from SUGM estimates. I structure the SUGMs to estimate assortative matching between types so that I can translate estimates of assortative matching to estimates of composition, as used within the theoretical model. I find substantial differences between the optimal and worst case scenario, with the insurance group and risk-sharing network scenario both falling close to optimal. First, despite the observed assortative matching, I find that the observed networks tend to be close to optimal networks already. I.e., if 0% is the worst case scenario, and 100% is the optimum,

⁵These controls include demographics, occupation, education, and (family) network centrality.

⁶As I do with the risk sharing network, I replicate these results using alternative specifications including LPM with controls, using dyadic logistic regression, and using tetrad logit.

⁷While of course a planner would want a single group, I constrain them to two groups as this mimics local risk sharing in villages.

observed assortative matching in networks places us 75% of the way to the optimum. As one might expect, detected insurance groups function better for covariate risk sharing than the risk sharing networks. However, if I use full covariate insurance as a benchmark, even the optimal scenario has losses equal to 16.5% of per capita consumption.

This work contributes to the present understanding of covariate risk sharing by situating it within the context of local network structure. Recent work has suggested the potential for covariate risk sharing. For example, Chiappori et al. (2014) find considerable heterogeneity in risk preferences under the assumption that risk sharing arrangements are complete within villages. An implication of their model is that less risk averse agents might take on more of the covariate risk in exchange for some increase in consumption over the long term. By relaxing the assumption of risk sharing at the village level, I am able to examine the relationship between network structure and the welfare dervied from covariate risk sharing.⁸

This work also contributes to the empirical study of assortative matching on risk preferences in social networks, and to my knowledge is the first evidence of assortative matching on risk preferences in village risk sharing networks.⁹ This reflects estimates from Attanasio et al. (2012) which find assortative matching in a risk pooling experiment done in the lab. Beyond replicating these results, the current work provides evidence of assortative matching on risk preferences in both real world risk sharing relationships and in a new country context, strengthening the external validity of this empirical result.¹⁰

Finally, these results contribute to the greater policy discussion on economic development and globalization. First, growing adoption of financial services in lower and middle income countries may have unintended consequences for risk sharing networks (e.g., Dizon et al., 2019; Dupas et al., 2019; Banerjee et al., 2022). By quantifying the importance of network structure, I reveal an important facet of the the net welfare effects of access to financial services. Second, climate change and growing interconnections in trade and financial systems may increase the scale of crises (Stiglitz, 2003; Zscheischler et al., 2018; Elliott and Golub, 2022). A greater scale of crises, exemplified by the COVID-19 pandemic, makes such covariate risk sharing all the more dear.

⁸More broadly, this work contributes to the study the risks insured by informal risk sharing networks and the constraints faced due to assortative matching: Gao and Moon (2016) and Jaramillo et al. (2015) study heterogeneity in risky endowments, while Xing (2020) studies heterogeneity in autocorrelation of (idiosyncratic) risk.

⁹In contrast to other dimensions, such as geography, wealth, religious affiliation, clan membership, and kinship (De Weerdt, 2002; Fafchamps and Gubert, 2007).

¹⁰Interestingly, these estimates are consistent with models of assortative matching on risk preferences in the presence of idiosyncratic risk sharing (Attanasio et al., 2012; Wang, 2015).

2 Data and Context

2.1 Risk Coping and Social Networks in Rural Ghana

I study risk sharing networks in rural southern Ghana. The data comes from four villages in southern rural Ghana. I utilize the 2009 household survey, which includes 633 individual respondents across the four villages. This data was collected over five rounds spaced at two month intervals over the year. The survey was designed as a husband and wife survey (Walker, 2011b). Households were randomly selected in each village, with a target of 70 husband-wife households. However, single-headed households are retained in the sample, so the number of households exceeds 70. This accounts for between 12-40% of households in each these villages. In addition to social and risk sharing networks, the survey also asked about income shocks, and elicited risk preferences (Barrett, 2009; Walker, 2011a). Further technical details, including the survey instrument itself can be found in Walker (2011b).

These villages face significant covariate risk, in particular from the pineapple export market (Conley and Udry, 2010; Walker, 2011b). Risk management is a key feature of these markets, where farmers use many strategies to manage risk. For example, Suzuki et al. (2011) documents partial vertical integration in pineapple markets in Ghana, explaining it is a strategy for small-holders to equip themselves to manage this risk through the use of local secondary markets. More relevant to this study, risk management within the villages includes substantial usage of informal networks (Udry and Conley, 2005; Walker, 2011a). A number of other empirical studies document features of the networks in this setting: Vanderpuye-Orgle and Barrett (2009) studies socially invisible members of the villages, and finds that risk pooling does not insure them as well as their richer, more socially visible counterparts. Within households, Castilla and Walker (2013) finds spouses behave non-cooperatively, hiding income through gifts to their networks.

2.2 Variable Construction

2.2.1 Risk Sharing Network

I will draw on graph theory to define 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 subscript 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 $ij \in g$, hence ij is in the edgelist. An adjacency matrix represents these nodes and edges in an $n \times n$ matrix $\mathbf{A} = \mathbf{A}(g)$. For the scope of this paper, I work with unweighted graphs. Thus $a_{ij} = 1$ if $ij \in g$ and 0 if not for all i, j. Additionally, for the majority of analyses, I work with undirected graphs, where the

	More Risk Averse	Less Risk Averse	Risk Loving
Panel A: Network Sta	tistics		
Average Degree	4.62	6.61	4.79
	(5.14)	(7.20)	(5.19)
Prop. Isolates	0.09	0.09	0.159
	(0.29)	(0.28)	(0.37)
Average Clustering	0.25	0.23	0.23
	(0.30)	(0.24)	(0.29)
Average Betweenness	86.0	119.5	99.9
	(150.4)	(212.1)	(210.1)
Panel B: Income Shoo	ks		
Average Net Losses	10.7	33.1	93.2
	(404.7)	(278.9)	(404.6)
Prop. Net Gain	0.34	0.35	0.27
	(0.48)	(0.48)	(0.45)
Prop. Net Loss	0.19	0.28	0.29
-	(0.39)	(0.45)	(0.46)
\overline{N}	236	217	82

Table 1: Summary of Risk Sharing Network and Shocks by Risk Preferences

For averages, standard errors are reported in parentheses below. 98 individuals who did not participate in the hypothetical gambles are excluded here. Risk loving have $\hat{\eta}_i \leq 0$, less risk averse (type 1) have $0 < \hat{\eta}_i < \eta_{\text{split}} \approx 0.003$, and more risk averse (type 2) have $\hat{\eta}_i \geq \eta_{\text{split}}$. Panel A: Degree is the number of other nodes directly connected to a node, $d_j = \sum_{j=1}^{N} a_{ij}$. Isolates are nodes with degree zero. Clustering is the average local clustering coefficient, which answers the question: for individual *i* connected to *j* and *k*, what proportion of the time are *j* and *k* also connected? Formally, clustering_i = $\frac{1}{d_i(d_i-1)} \sum_{j=1}^{N} \sum_{k=1}^{N} a_{ij}a_{jk}a_{ik}$. Betweenness centrality is the sum of shortest paths between other nodes in the network on which that node lies. Panel B: Shocks are unexpected losses or gains to income reported by the respondents summed up over individuals. I omit one outlying value for the tabulation of mean and variance, a net loss of about 48,000 Ghanaian Cedis reported by a type 2 (more risk averse) household.

adjacency matrix is symmetric: $a_{ij} = a_{ji}$. The diagonal $a_{ii} = 0$ by construction.¹¹

To construct the risk sharing network, I use questions related to gifts and trust. Two nodes are connected when they both report reciprocal gift exchange in the past and also would trust the other to take care of a valuable item for them. This definition includes these two network

¹¹Nodes 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.

questions as they are closely related to risk sharing, and emphasizes the importance of reciprocal ties in risk sharing (Fafchamps and Lund, 2003; Blumenstock et al., 2016). Formally, $a_{ij} =$ trust_{ij} × gift_{ij} where trust_{ij} = 1(*i* trusts *j*) × 1(*j* trusts *i*) and gift_{ij} = 1(*i* reports giving to *j*) × 1(*i* reports receiving from *j*) × 1(*j* reports giving to *i*) × 1(*j* reports receiving from *i*). For more detail on network elicitation and construction, see Appendix A.1 and Appendix A.2, respectively.

To check that these networks are related to risk sharing transfers (as opposed to more prosaic favor exchange), I test the response of transfers (i.e., gifts and informal loan disbursements) during the study period to to researcher run lotteries, finding that transfers seem to be responsive to these along the channels of the risk sharing network. While the effect is noisy due to data issues, I find evidence consistent with the notion that these networks are capturing risk sharing. A full description of this exercise and results can be found in Appendix A.3

Table 1 presents summary statistics about the risk sharing networks. When comparing less and more risk averse individuals there are differences in both degree and betweenness centrality. In particular, less risk averse individuals have higher degree—more risk sharing connections. Likewise, they have higher betweenness centrality—holding positions which bridge between other nodes—suggesting their importance in the routing of gifts and transactions through the network. This is both an unexpected and important feature of the data. Theory might predict that those with higher risk aversion would be central in risk sharing networks, reflecting a demand for insurance and issues of moral hazard (Jaramillo et al., 2015). Furthermore, heterogeneity in degree by underlying type can confound estimates of assortative matching (Graham, 2017). Despite these differences, it's interesting to note that the difference in clustering between less and more risk averse individuals would appear to be economically small.¹²

I also include summary statistics about income shocks faced by individuals in the sample in Table 1. While I caution against a strictly behavioral explanation for these shocks, some interesting patterns emerge. First, the risk averse (both type 1 and type 2) have limited downside exposure relative to their upside exposure. Second, the variation in shocks among the less risk averse and risk loving is larger than those for the more risk averse. Despite this, those who are more risk averse face more downside risk than and have greater net losses those who are less risk averse, an important reminder that exposure to shocks depends on circumstances outside of risk preferences.

¹²Though it is beyond the scope of the current work, one might interpret this as a difference in linking social capital without an accompanying difference in bonding social capital. In terms of detected insurance groups discussed later, this might also suggest that less risk averse individuals might be more likely to serve as liaisons between detected insurance groups.



(a) A stylized risk sharing network with (latent to the econometrician) insurance groups denoted by red and blue.

(b) Insurance groups: co-membership links occur within detected insurance group.



(c) The difference in network representations: green links are present in the insurance group but not the network, while orange are present in the network but not insurance groups. There is one additional connection within the red group and one less between the red and blue groups.

Figure 1: A stylized example of risk sharing networks, insurance groups and the differences in their network representations.

2.2.2 Community Detection and Insurance Groups

While canonical work on informal insurance modeled risk sharing at the village level (e.g., Townsend, 1994), empirical work has shown that risk sharing is mediated by interpersonal relationships (Fafchamps and Lund, 2003). However, both theoretical and empirical work suggest that broader network position may matter. First, Being central overall in networks may improve risk sharing overall (Vanderpuye-Orgle and Barrett, 2009). Second, transfers flowing through networks may contribute to one's insurance pool (De Weerdt and Dercon, 2006; Henderson and Alam, 2022). Conversely, where networks are sparse information asymetries may lead to barriers in such flows of transfers (Bloch et al., 2008; Ambrus et al., 2014). Third, broader network structure may drive new risk sharing connections (Putman, 2020).

I use community detection to detect insurance groups using network data (Newman, 2012) in order to characterize agents broader network position. Community detection aims to assign nodes to modular communities, which we call detected insurance groups. In our case, a good community assignment is one where most risk sharing connections fall within the community with only a few of the connections fall between communities. Such detected communities have shown to be predictive of risk sharing group formation in experimental settings (Putman, 2020).

My approach to uncovering communities is based on random walks through the network: A

random walker moves from node to node in the network by way of edges, randomly selecting the next node it visits among those in the network neighborhood. In particular, I use the *Walktrap* algorithm, which uses these random walks to determine the similarity between nodes by the destinations of random walkers originating at that node (Pons and Latapy, 2005). The intuition is that these random walks will become trapped in tightly knit sections of the local network, meaning the algorithm will see nodes in tightly knit sections of the network as interchangeable and therefore group them together. Further discussion of this algorithm for risk sharing networks can be found in Appendix A.4.

I assign individuals to insurance groups using Walktrap community detection on the risk sharing network with walks of four steps.¹³ For a visualization of the resulting community detection, see Figure A1. After I have assigned nodes to communities, I construct a network which indicates co-membership in an insurance group using these community assignments. If insurance takes place within these groups and not between them, this would represent the relevant risk sharing relationships.¹⁴ I represent co-membership in detected insurance groups using an adjacency matrix C where c_{ij} is an indicator variable for if i and j are in the same detected insurance group. Like the adjacency matrix, C is symmetric. The difference in construction of the bilateral risk sharing network and detected insurance groups is depicted in Figure 1.

2.2.3 Risk Preferences

I use four hypothetical gambles to measure individuals' risk aversion, which ask respondents to choose between a sure payment Y_A and a risky gamble Y_B . These gambles are presented in both the gains and losses domains, and with variation in the sure and variable payments. The first two menus presented are in the gains domain. In the first menu, the risky gamble Y_B is held fixed while the sure payment Y_A is increased. In the second menu, the sure payment is held fixed while the upside of the risky gamble is reduced. The third and fourth menus reflect the first and second set onto the losses domain.

To translate these hypothetical gambles into coefficients of risk aversion, I match assumptions to the theoretical model presented in Section 5. First, I assume Y_B is normally distributed and second that individuals exhibit Constant Absolute Risk Aversion (CARA, or exponential preferences). These assumptions allow for a mean-variance representation of expected utility, which is crucial for the later welfare results (Sargent, 1987).¹⁵ I compute $\hat{\eta}_i$ for each menu and individual

¹³Longer walks tend to result in larger communities relative to shorter walks. I opt for the default of four steps.

¹⁴Relatedly, the theoretical results in Ambrus et al. (2014) suggest we would expect non-zero but small amounts of risk sharing between islands—which might tend to form *ex post* within *ex ante* communities. Likewise, the empirical results in Putman (2020) suggest very little risk sharing across detected insurance groups.

¹⁵In particular, it allows for the comparison of average incomes to the variance in income.



Figure 2: Risk sharing networks in the village of Pokrom with constant absolute risk aversion coefficients indicated by color (transparent nodes are those who did not participate in the risk aversion module). 'Split' refers to $\eta_{\text{split}} (\approx 0.003)$, the coefficient value that distinguishes type 1 (more risk averse) agents from type 2 (less risk averse) agents. See Figures A4, A5, and A6 for other villages. For the distribution of risk preferences, see additionally Figure A2 which features a matching color coding.

using the sample analogue of the expression:

$$\eta_i = \frac{2(E(Y_B) - Y_A)}{V(Y_B)} \tag{1}$$

Finally, to combine these into measures of risk aversion, I average over menus. To check the assumption of asymptotic normality, I also compute CARA coefficients without making this distributional assumption and with an alternative order of computation. The assumption of normality has almost no effect on the computed CARA coefficients, while changing the order of computation results in nearly co-linear but larger coefficients of risk aversion. Precise details of how each coefficient is computed and comparisons are available in Appendix A.5.

Coefficients of Absolute Risk Aversion are plotted over the risk sharing network for one example village in Figure 2. Additionally, the distribution of coefficients and definition of types is plotted in Figure A2. Of those in the network who answered the elicitation module, I split these individuals into three groups: risk loving, less risk averse, and more risk averse. Risk loving are those with $\eta_i < 0$. This accounts for about 20% of the individuals with preferences. I split the remaining risk averse individuals into evenly sized groups of approximately 40% each, with more risk averse individuals being above a cut-point, $\eta_{split} \approx 0.003$.¹⁶

3 Empirical Strategy

In this section I describe the two main econometric models of network formation I use to estimate assortative matching on risk preferences. Each serves a different purpose within the paper. First, dyadic regression serves as a reduced form approach to describe assortative matching in the data. Its inclusion is beneficial as it allows for familiar exposition and interpretation as well as lending itself more easily to tests of robustness and comparison with past literature.¹⁷ Second, SUGMs serve to estimate assortative matching on risk preferences in a way that can be translated into the composition of risk sharing groups as is specified in the theoretical model.

3.1 Dyadic Regression

3.1.1 Risk Sharing Networks

To establish the degree of assortative matching on risk preferences, I will start by estimating dyadic regressions, an econometric model of network formation. In these regressions, each pair of nodes is treated as an observation. The most parsimonious model regresses risk sharing connections on differences in measured risk aversion,

$$a_{ij} = \beta_0 + \beta_1 |\eta_i - \eta_j| + \varepsilon_{ij} \tag{2}$$

where a_{ij} is an indicator for if *i* and *j* are connected in the risk sharing network, η_i is the risk aversion of individual *i*, and ε_{ij} is the error term. Note that all variables enter symmetrically: $a_{ij} = a_{ji}$ (as the adjacency matrix *A* is symmetric) and explanatory variables are computed as to enter symmetrically (Fafchamps and Gubert, 2007). A negative estimate of 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.

A second specification includes the sum of risk aversion η_i and η_j to control for the correlation between risk aversion and popularity, a difficult feature of the current cross-sectional setting.

¹⁶It is difficult to split the remaining risk averse individuals into exactly even groups, and the less risk averse group tends to be slightly larger in practice.

¹⁷In particular, I note similarities and differences between these results and those found in Attanasio et al. (2012).

Where one or both have low risk aversion, I would expect these agents to be more popular and hence have a higher probability of forming a link.¹⁸ Whereas, in a panel setting, I might use a fixed effects approach to account for degree heterogeneity, in this setting I rely on selection-on-observables.¹⁹ Specifically, I control for the sum of risk aversion:

$$a_{ij} = \beta_0 + \beta_1 |\eta_i - \eta_j| + \beta_2 (\eta_i + \eta_j) + \varepsilon_{ij}$$
(3)

A positive estimate of β_2 suggests that individuals who are more risk averse are less likely to link to each other. Furthermore, I take estimates of β_1 using this strategy as my preferred estimate of assortative matching from the dyadic regressions.

3.1.2 Heterogeneity by Family Ties

I examine how assortative matching varies by family ties. Family is interesting because it serves as a longstanding relationship. This means there may be a higher propensity to link overall, but also more information about potential risk sharing partners. We document the former in our third specification:

$$a_{ij} = \beta_0 + \beta_1 |\eta_i - \eta_j| + \beta_3 \text{Family}_{ij} + \varepsilon_{ij}$$
(4)

where family is an indicator variable equal to one if *i* and *j* report being kin. A positive estimate of β_3 suggests that family are more likely to be connected within the risk sharing network. A fourth specification combines specifications (2) and (3) to add the *ad hoc* control.

$$a_{ij} = \beta_0 + \beta_1 |\eta_i - \eta_j| + \beta_2 (\eta_i + \eta_j) + \beta_3 \text{Family}_{ij} + \varepsilon_{ij}$$
(5)

Finally, a fifth specification introduces interactions between the difference in coefficients of risk aversion and family ties to understand this heterogeneity.

$$a_{ij} = \beta_0 + \beta_1 |\eta_i - \eta_j| + \beta_2 (\eta_i + \eta_j) + \beta_3 \operatorname{Family}_{ij} + \beta_4 \operatorname{Family}_{ij} \times |\eta_i - \eta_j| + \varepsilon_{ij}$$
(6)

¹⁹Notably, Graham (2017) introduces an approach to control for degree heterogeneity in cross sectional settings which relies on combinations of data where fixed effect terms "net out" of the estimation, which I include as a robustness check.

¹⁸There are three basic stories about what might cause risk preferences to be correlated with popularity. First, risk preferences could be correlated with unobservable personality traits. For example, it could be that less risk averse agents differ in personality traits not directly related to risk preferences. Second, economic decision-making specifically involving risk might alter someone's fortunes and thus their social standing. If those with lower risk aversion make riskier, higher reward decisions, this may be parlayed into income growth and higher SES in the long term (Elbers et al., 2007; Karlan et al., 2014). Third, though I have assumed constant absolute risk aversion, it is plausible that having better social standing could make a person less risk averse e.g., in a model of decreasing absolute risk aversion. A fourth issue is also at play: even when risk aversion is not correlated with popularity, as outlined in Graham (2017), a person well connected to all types might be measured as not harboring a preference for similar risk-preferenced others when in fact they do.

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. This might suggest greater information about others' preferences driving matching, as in Attanasio et al. (2012).

3.1.3 Detected Insurance Groups

It is also interesting to see how assortative matching changes in detected insurance groups. I re-estimate the above dyadic regressions with the network representation of detected insurance groups as the outcome: In all of the above specifications, I replace a_{ij} with c_{ij} , the ijth entry of C, the matrix denoting co-membership in a detected insurance group. $c_{ij} = 1$ if $i \neq j$ are in the same detected insurance group, and 0 otherwise. If we treat detected insurance groups as actual groups (legible to participants, but not to the econometrician), such a specification might accord with a coalition formation game with simultaneous announcement like those in Hart and Kurz (1983).²⁰ Within this framework, The dyadic regression coefficients on the difference in risk aversion can be interpreted within this framework as measures of assortative matching. Outside of such a framework, we can think of these estimates as capturing local network structure, as opposed to choice of partners.

3.1.4 Estimation and Standard Errors

I estimate these dyadic regressions as linear probability models without controls. However, I include a number of robustness checks which are detailed in Appendix A.6. Importantly, errors are non-independent in dyadic regressions. In particular, the residuals of dyads involving a particular node might be arbitrarily correlated.²¹ To correct standard errors for this type of non-independence, I use dyadic robust standard errors (Fafchamps and Gubert, 2007; Cameron and Miller, 2014; Tabord-Meehan, 2019).

3.2 Subgraph Generation Models

3.2.1 Intuition

A useful tool for understanding risk sharing networks and insurance groups is called a Subgraph Generation Model (SUGM). SUGMs treat networks as emergent properties of their constituent

²⁰These models are not unlike those of pairwise stability found in Jackson and Wolinsky (1996). For example, in one game, to form a coalition, all members of the coalition must announce the same list of names, meaning they can exclude players by not including them in their list.

²¹Formally, it may be the case that $Cov(\varepsilon_{ij}, \varepsilon_{lk}) \neq 0$ if i = l, i = k, j = l, or j = k.

subgraphs.²² A *subgraph* (sometimes called an 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, for a subset of two nodes in a graph, the subgraph will be either a link or two unconnected nodes. For three nodes, the subgraph might be a triangle (a trio of nodes all connected by edges), a line (one central node connected to the two others), a pair and an isolate (two nodes connected and one unconnected), or an empty subgraph (three unconnected nodes). I focus on connected subgraphs for the SUGM. In a three node example, this means I leave aside pairs, isolates, and the empty subgraph, focusing on the triangle and the line. Likewise, while a link is a subgraph of interest, two unconnected nodes is not.

3.2.2 Links and Isolates Subgraph Generation Model with Risk Preference Types

Like dyadic regression, SUGMs are estimated to understand how individuals of different risk preferences connect to each other. However, for these estimates I build the SUGMs to estimate the affinity within and between risk preference types. This allows me to recover the composition of detected insurance groups in terms of risk preferences in order to assess the welfare implications of assortative matching.²³ I estimate SUGMs with both links and isolates, differentiated by types, which I base on risk preferences. There are two models of interest: a baseline model and a preference model. I start with the baseline model. For various reasons, a small subset of individuals in the network did not participate in the survey module I use to recover risk preferences.²⁴ Additionally, in the model, I study risk sharing among only those who are risk averse. I term both those who were not surveyed and those who have risk loving preferences as nuisance nodes. Therefore, to understand the baseline rate of subgraph generation among the risk averse, I estimate a model with two types. I estimate coefficients for five features: isolates of risk averse nodes, isolates of nuisance nodes, links within nuisance nodes, and links between risk averse and nuisance nodes. I refer to the second model to as the preference model. I estimate the full model with less risk averse, more risk averse, risk loving, and non-surveyed types for a total of four types. This includes isolates of each type, links within each type, and links between each pair of types for a total of 14 features.

²²While Exponential Random Graph Models have a similar motivation, they do not succeed at reconstructing graphs with any success. They depend on an assumption of independence of links. If this independence does not hold they are not consistent (Chandrasekhar and Jackson, 2021). To the contrary, many studies of risk sharing would expect links are dependent on each other. See for example Jackson et al. (2012).

²³In the case of the detected insurance groups, I am actually recovering my estimate of assortative matching on risk preferences *from* the composition of detected insurance groups. In contract, in the risk sharing network it comes from the assortative matching measure itself. This is because of the assumptions used in building detected insurance groups.

²⁴Some of these individuals were not surveyed at all, but appear in the network. Others may be part of the sample who were not interviewed in that particular round or module. I leave them in the network, consistent with recommendations from Smith et al. (2022).

I directly estimate the parameters using an algorithm given by Chandrasekhar and Lewis (2016) and Chandrasekhar and Jackson (2021). Estimating a SUGM directly is essentially estimating the relative frequency of various subgraphs in a network. However, I 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 between ij, jk, and ik would incidentally generate a triangle. To estimate the true rate of subgraph generation, I order subgraphs by the 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 (that is, one composed of a greater number of nodes). For subgraphs of the same size, order is arbitrary, but must exclude occurrences of this subgraph incidentally generated by other subgraphs which are further along in the order. For example, for a SUGM featuring links and triangles, I order links first, triangles second, etc. While counting links and potential links, I neglect pairs of nodes ij if jk and ik are in the graph.²⁵ More estimation details can be found in Appendix A.7.1.

3.2.3 Pooled Subgraph Generation Models

As my data has four unrelated networks, I need to make choices as to how to handle these multiple networks in the SUGM. One approach would be to estimate a subgraph generation model for each village and average the coefficients of these. A different strategy, and one that relies on the same asymptotics as the single network case from Chandrasekhar and Jackson (2021), is to pool the counts and potential counts from the villages to estimate a single coefficient across the villages. This leads to an adjusted class of SUGMs I term Pooled SUGMs. To do so, I cannot 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, I need to collect counts of features and potential counts of features in all four villages before combining. Details of this modification can be found in Appendix A.7.2.

3.2.4 Differences in Assortative Matching

These SUGM estimates give me a way to test for assortative matching between risk sharing networks and detected insurance groups. However, the risk sharing network and network of comembership in detected insurance groups have different degrees of attachment. To make an apples to apples comparison, I normalize my results by taking the ratio of Preferences SUGM coefficients to Baseline SUGM coefficients. I focus on the coefficients for within links for type 1 agents, within links for type 2 agents, and links between type 1 and 2 agents. For all three,

²⁵If I added lines of three 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.

I divide by the coefficient on links within any risk averse agents from the baseline model. This yields an excess affinity for connections among these dyads. Doing this for both the detected insurance group coefficients and the risk sharing network coefficients, I can compare between the models. See appendix A.7.3 for details, including a conservative analytic approximation of standard errors.

	Match Between i and j in Risk Sharing Network				
	(1)	(2)	(3)	(4)	(5)
$ \eta_i - \eta_j $	-0.00454*	-0.00564*	-0.00337	-0.00510*	-0.00427*
	(0.00218)	(0.00250)	(0.00179)	(0.00214)	(0.00193)
$\eta_i + \eta_j$		-0.00124		-0.00196	-0.00195
,		(0.00181)		(0.00164)	(0.00164)
Family _{ij}			0.285***	0.285***	0.298***
- 5			(0.0156)	(0.0156)	(0.0196)
Family _{<i>ij</i>} × $ \eta_i - \eta_j $					-0.0133
					(0.0107)
Village FE	Yes	Yes	Yes	Yes	Yes
Other Controls	No	No	No	No	No
N dyads	71052	71052	71052	71052	71052
B^2					

Table 2: Dyadic Regression: Risk Sharing Network

Dyadic robust standard errors reported in parentheses (Fafchamps and Gubert, 2007). All specifications are dyadic linear probability models with matching in the risk sharing network as the dependent variable. η_i is risk aversion of individual *i*, so $|\eta_i - \eta_j|$ is the absolute difference of risk aversion while $\eta_i + \eta_j$ is the sum. Both absolute differences and sums of risk aversion are transformed into *z*-scores. * p < 0.05, ** p < 0.01, *** p < 0.001

4 Results

4.1 Dyadic Regression

4.1.1 Risk Sharing Network

Table 2 reports the results from estimating the dyadic regression specifications. I include village level fixed effects in all dyadic regression specifications, though this does not affect the magnitudes estimated. Reported *t*-statistics are computed using dyadic robust standard errors (Fafchamps and Gubert, 2007). To make results more interpretable, I transform risk preferences into *z*-scores before computing regressors, so β_1 estimates the effect of a one-standard deviation absolute difference in risk aversion.

Columns (2) and (5) present my preferred specifications. Across all specifications I see negative estimates for the effect of difference in absolute risk aversion on the likelihood of linking in the risk sharing network. In column (1) when the sum of risk aversion is omitted from the model, the estimates is somewhat smaller than when risk aversion is controlled for, and in column 3, the estimate is insignificant. In contrast, proxying for degree with the sum of risk aversion in columns (2), (4) and (5) yields a negative and significant estimates (at the 5% level), which I interpret as evidence of assortative matching on risk preferences. In particular, focusing on column (2) I estimate a one standard deviation difference in risk aversion leads a 0.56 percentage point reduction in the probability of connection.

As in other contexts, family connections are a strong determinant of co-participation in the risk sharing network. Across specifications (3), (4), and (5), having a family connection is positively associated with connection in the risk sharing network (statistically significant at the 0.1% level). In column 5, family member dyads are 29.8 percentage points more likely to form a risk sharing relationship than non-family members.

In columns (4) and (5), when I control for family connection and risk aversion, the estimate of β_1 falls. However, this may speak more to the mechanism of assortative matching. Similar to Attanasio et al. (2012), I 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, I have $\hat{\beta}_1 + \hat{\beta}_4 = -0.0176$, however this estimate is not statistically significant as the interaction term, while negative, tends to be quite noisy ($\chi^2(1) = 2.46$, p = 0.12). Interpreting the coefficient, a one standard deviation difference in risk aversion is associated with a reduction in the probability of linkage by 1.33 percentage points between family members. This could suggest stronger assortative matching when more information about risk preferences in available, but the results are not conclusive.

4.1.2 Detected Insurance Groups

Table 3 reports the results from re-estimating equations with co-membership in a detected insurance group as the outcome of interest. The estimates of β_1 are similar in magnitude to those in the risk sharing network. However, none are statistically significantly different from 0 at standard significance levels. Hence, I fail to find evidence for assortative matching in the detected insurance groups. Heterogeneity by family connections may provide some clues as to the differences. In particular, in column (5), $\beta_1 + \beta_2 = -.0049$ is not significantly different than zero ($\chi^2(1) = 0.1$). These results suggest individuals' ability to select risk sharing partners falls to almost nothing as they place themselves into broader networks, and reflects only the shadow of the choices made in their personal connections.

	i and j are Co-members in Detected Insurance Group				
	(1)	(2)	(3)	(4)	(5)
$\overline{ \eta_i - \eta_j }$	-0.00668	-0.00651	-0.00557	-0.00600	-0.00607
	(0.00485)	(0.00543)	(0.00453)	(0.00517)	(0.00511)
$\eta_i + \eta_j$		0.000189		-0.000489	-0.000490
		(0.00426)		(0.00407)	(0.00407)
Family _{ij}			0.269***	0.269***	0.268***
			(0.0205)	(0.0206)	(0.0257)
Family _{<i>ij</i>} × $ \eta_i - \eta_j $					0.00113
					(0.0151)
Village FE	Yes	Yes	Yes	Yes	Yes
Other Controls	No	No	No	No	No
N dyads R^2	71052	71052	71052	71052	71052

Table 3: Dyadic Regression: Detected Insurance Groups

Dyadic robust standard errors reported in parentheses (Fafchamps and Gubert, 2007). All specifications are dyadic linear probability models with matching in the risk sharing network as the dependent variable. η_i is risk aversion of individual *i*, so $|\eta_i - \eta_j|$ is the absolute difference of risk aversion while $\eta_i + \eta_j$ is the sum. Both absolute differences and sums of risk aversion are transformed into *z*-scores. * p < 0.05, ** p < 0.01, *** p < 0.001

4.1.3 Addressing Threats to Validity

Before moving on to the results of the Subgraph Generation Models—which echo the results presented above—it is useful to address threats to validity for the dyadic regression results presented here. Quantitatively, the coefficients in Tables 2 and 3 tend to be robust to controlling for demographic factors and network centrality. That is, I find a similar degree assortative matching on risk preferences when controlling for the sum of risk aversions, and this assortative matching attenuates in the detected insurance groups. However, this finding comes with the caveat that these regressions have smaller and noisier effects, so that significance fades to some degree. Nevertheless, the magnitude of assortative matching tends to be consistent, suggesting a real—if subtle—relationship. See Appendix A.6.1 for detailed results using this selection-on-observables approach for the linear probability model. Likewise, results are robust to choice of specification. Appendix A.6.2 presents results from dyadic logistic regression, which echo those from the linear probability models.

To be sure I can justify using sum of risk aversion as an ad hoc control for popularity, I utilize a network formation model termed tetrad logit, which is designed to account for correlations between heterogeneity in degree and type when estimating assortative matching (Graham, 2017). Intuitively, this method selects tetrads of nodes (sets of four nodes and their connections) which

	Risk Sharing Network		Detected In	Detected Insurance Groups	
Model, Subgraph	Stat.	Std. Err.	Stat.	Std. Err.	
Baseline SUGM Coef.					
Within: All Risk Averse	0.0404	0.0009	0.0928	0.0013	
Preferences SUGM Coef.					
Within: Less risk averse	0.0561	0.0010	0.1189	0.0015	
Within: More risk averse	0.0299	0.0008	0.0713	0.0012	
Between: More, less risk averse	0.0360	0.0008	0.0876	0.0013	
Ratio of Coefs: Pref./Baseline					
Within: Less risk averse	1.389	0.040	1.281	0.023	
Within: More risk averse	0.740	0.026	0.768	0.017	
Between: Less, more risk averse	0.891	0.028	0.944	0.019	

Table 4: Links and Isolates Pooled Subgraph Generation Model: Coefficients of Interest from Baseline and Preferences Models and Coefficient Ratios.

Sample size for features of interest is 49536 dyads. Models are abridged, focusing on coefficients and ratios of interest. For full results, Baseline SUGM coefficients are presented in Tables A12 and A13 and preference SUGM Coefficients are presented in Tables A14 and A15 Coefficient ratios are used to compare the two models, since higher average degree (as is present the detected insurance groups will result in higher coefficient estimates. SEs for coefficients are computed as shown in Appendix A.7.2 and SEs for ratios are computed as shown in Appendix A.7.3.

contribute to the estimate only if the node fixed effects for degree drop out within that tetrad, thereby netting out heterogeneity in popularity. This allows for estimates of assortative matching unconfounded by popularity. I estimate models for each village network which lead to three insights. First, results unconditional on the sum of risk aversion from tetrad logit are similar to those from comparable methods (i.e., logit) when conditioning on the sum of risk aversion. Second, after accounting for popularity with tetrad logit, controlling for the sum of risk aversion does not substantially change estimates. Third, assortative matching also attenuates in detected insurance groups using this estimator. These results give me confidence that the sum of risk aversion is controlling for a nuisance correlation between risk aversion and popularity. I describe the tetrad logit estimator in greater detail and present results in Appendix A.6.3.

4.2 Subgraph Generation Models with Types

The SUGM results for the coefficients of interest are presented in Table 4. While these are abridged for clarity, full results of all SUGM models are available in Appendix A.7.4. Using the baseline model, I estimate that individuals who are risk averse tend to form links with each other at a rate of 4.04%. The network of co-membership in detected insurance groups tends to be denser than the risk sharing network: I estimate that individuals who are surveyed about preferences tend to

form links with each other at a rate of 9.28%, more than twice the rate in the risk sharing network.

Considering the coefficients of interest from the preferences model, I derive two main findings. First, I 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, I do not see the same kind of assortative matching when looking at more risk averse 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 form links between type (3.60%). In this way, less risk averse 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 detected insurance groups mirrors the pattern in the risk sharing network (see Table 4). 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%). The degree of assortative matching falls in the detected insurance groups *vis a vis* the risk sharing network when we correct for the average number of links between individuals in the network. Results measuring the degree of assortative matching as the ratio of the rate of between links to the rate of links between all risk averse individuals are also presented in Table 4. The ratio of within types for less risk averse individuals is higher in the risk sharing network, whereas the ratio of between types is lower. Essentially, this indicates a reduced degree of assortative matching in detected insurance groups relative to networks.

5 Welfare Implications of Assortative Matching

What are the welfare implications of assortative matching? In this section, I build a model that considers a risk-neutral planner seeking to construct two risk pooling groups in a village in order to maximize expected utility within risk averse members of the village. Based on this model, I translate coefficients of assorative matching from the SUGMs into estimates of welfare.

To model covariate risk sharing in groups, I leave aside insurance group size and focus on optimal group composition itself, which I define as the mix of types which make-up a group. I consider group composition with regard to risk aversion, with relatively less and more risk averse individuals.²⁶ I set up this problem in two steps. First, I characterize how risk is pooled in a group according to its composition. Second, using the solutions and value functions from the first optimization problem, I write a planner's problem maximizing aggregate expected utility of

²⁶While, empirically, I also observe some risk loving individuals, I opt not to include them within the model. I explain my reasoning for this choice in Appendix B.1.

consumption in a village with groups, conditional on the composition of those groups.

5.1 Risk Sharing in Groups

I start from a baseline of perfect idiosyncratic risk sharing. In my model, this takes the form of all idiosyncratic shocks being smoothed to zero (I will assume mean incomes are zero for the purposes of this problem). However, the *average* of these shocks (which in general is not zero) joins the covariate risk in the sharing problem. After this set of transfers takes place, a round of risk shifting takes place. Less risk averse individuals may take on more of the covariate risk. This covariate risk derives from both the average idiosyncratic shock—which in general is not zero—and a perfectly correlated covariate shock.²⁷ More risk averse agents are able to take on less of the covariate risk, shifting them onto 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.

5.1.1 Setup

Suppose a group of fixed size N that sits within a village. Group member i has exponential utility functions with coefficient of absolute risk aversion η_i :

$$u_i(c_i) = \frac{1 - e^{-\eta_i c_i}}{\eta_i}.$$

Now, suppose there are low and high risk aversion households, where type is indexed by $\ell = 1, 2$. That is, $\eta_2 > \eta_1 > 0$. N_ℓ is the number of individuals of type ℓ , and $p = N_1/N$ characterizes the composition of the group in terms of these types. All households face a shock perfectly correlated at the village level, \tilde{y}_v and an idiosyncratic shock \tilde{y}_i . Risk is symmetric between households and between types: Household level shocks, $\tilde{y}_i \sim^{\text{iid}} N(0, \sigma^2)$ and village level shocks $\tilde{y}_v \sim^{\text{iid}} N(0, \nu^2)$. Income for agent i and type ℓ is computed $y_{\ell i} = \tilde{y}_i + \tilde{y}_v$. Taking account of the risk sharing process, I write the consumption of household i of type ℓ as a weighted sum of the idiosyncratic and covariate shocks in the group. For type $\ell = 1, 2$,

$$N_1 c_{1i} \le \theta \left(\sum_{i=1}^N \tilde{y}_i + N \tilde{y}_v \right) - N_1 \lambda_{1i} \text{ and } N_2 c_{2i} \le \theta \left(\sum_{i=2}^N \tilde{y}_i + N \tilde{y}_v \right) - N_2 \lambda_{2i}$$
(7)

²⁷This shock is perfectly correlated because I want to explore the role of assortative matching on risk preferences in the presence of covariate risk, as opposed to heterogeneity in income correlation between individuals.

which can be re-written,

$$c_{1i} = \left(\frac{\theta}{p}\right) \left(\frac{1}{N} \sum_{i=1}^{N} \tilde{y}_i + \tilde{y}_v\right) - \lambda_{1i} \text{ and } c_{2i} = \left(\frac{1-\theta}{1-p}\right) \left(\frac{1}{N} \sum_{i=1}^{N} \tilde{y}_i + \tilde{y}_v\right) - \lambda_{2i}.$$
(8)

Two assumptions are made in constructing this risk sharing function. First, the proportion $\theta \in [0, 1]$ of the covariate risk is borne by the less risk averse individuals in the group. When $\theta = 1$, all covariate risk is taken on by less risk averse individuals, when $\theta = p$, covariate risk is shared equally among all members of the group (i.e., only idiosyncratic risk is pooled) and when $\theta = 0$, all risk is taken on by more risk averse households. Second, there is a second, recurring transfer (of fixed value) from the more risk averse to the less risk averse, which is embodied by the parameters λ_{1i} and λ_{2i} . As we should expect, the total recurring transfers from type 2 must (weakly) exceed the transfers out to type 1, though this constraint will bind in practice: $-N_1\lambda_{1i} \leq N_2\lambda_{2i}$. With these assumptions, Eq. 7 can be read that the total consumption of type 1 agents is less than or equal to the share of the covariate shock they take on plus their total recurring transfers. Finally, due to the exponential utility function and normal distribution of shocks, I represent expected utility as a mean-variance decomposition (for details, see Appendix B.2):

$$E(U_{\ell}(c_{\ell i})) = E(c_{\ell i}) - \frac{\eta_{\ell i}}{2} Var(c_{\ell i}).$$

5.1.2 Optimization Problem

The planner maximizes expected utility of less risk averse agents subject to several constraints.

$$\max_{\lambda_1,\lambda_2,\theta} E(U_1(c_{1i})) \tag{9}$$

subject to

$$E(U_2(c_{2i}|\theta = p))) \le E(U_2(c_{2i}))$$
(10)

$$c_{1i} = \left(\frac{\theta}{p}\right) \left(\frac{1}{N} \sum_{i=1}^{N} \tilde{y}_i + \tilde{y}_v\right) - \lambda_{1i}$$
(11)

$$c_{2i} = \left(\frac{1-\theta}{1-p}\right) \left(\frac{1}{N}\sum_{i=1}^{N}\tilde{y}_{i} + \tilde{y}_{v}\right) - \lambda_{2i} \qquad (12)$$

$$0 \le p\lambda_1 + (1-p)\lambda_2 \tag{13}$$

Constraint (10) is an incentive compatibility constraint: more risk averse agents cannot be worse off, *ex ante* than in the case where they only perfectly pool idiosyncratic risk.²⁸ Constraints (11) and (12) serve as individual budget constraints for each type, and finally, constraint (13) serves

²⁸As I abstract away from commitment issues, this constraint is *ex ante*.

to ensure the feasibility of the recurring transfers (for details, see Appendix B.3).

5.1.3 Solutions

How much covariate risk is shifted to the less risk averse agents? I solve the model, and present this process in Appendix B.4. The proportion of covariate risk shared will depend on the risk aversion and proportion of each type:

$$\theta^*(p,\eta_1,\eta_2) = \frac{p\eta_2}{(1-p)\eta_1 + p\eta_2}.$$
(14)

Recall, if $\theta = 1$, all covariate risk shifts to less risk averse individuals, and if $\theta = p$, the baseline of perfect idiosyncratic risk sharing is maintained. Since $\eta_2 > \eta_1$, $\theta^*(.) > p$ (for proof, see Appendix B.5). This means some degree of covariate risk is shifted to less risk averse individuals. Likewise, unless $\eta_1 = 0$ (I assume it does not) or p = 1, some risk is still taken on by the more risk averse. Furthermore, group composition matters for the degree of covariate risk sharing.

What are more risk averse agents willing to pay to shift risk away? Since λ_2^* is paid into the group pot, type 2's willingness to pay depends on their own risk aversion, type 2's risk aversion, and group composition:

$$\lambda_2^*(p,\eta_1,\eta_2) = -\frac{\eta_2}{2} \left(1 - \frac{\eta_1^2}{((1-p)\eta_1 + p\eta_2)^2} \right).$$
(15)

where the expression in parentheses lies between 0 and 1. Because risk is symmetric in this model (i.e., risk averse and risk loving types face the same covariate risk), the transfer does not depend on covariate risk. Finally, type 1 will maximize their utility and hence the payments they receive from type 2. I can write λ_1^* by converting type 2's willingness to pay into type 1's average payment:

$$\lambda_1^*(p,\eta_1,\eta_2) = -\left(\frac{1-p}{p}\right)\lambda_2^*(p).$$
(16)

5.2 The Planner's Problem

The risk neutral planner seeks to maximize aggregate expected utility of consumption, conditional on the composition of groups. For ease of exposition, the planner allocates individuals to two groups, g = A, B. I will update the notation from the first stage slightly. For a given group g, N_g is the group size and $N_A + N_B = N$. Then $N_{g\ell}$ is the number of individuals of type ℓ in group g and $p_{g\ell} = \frac{N_{g\ell}}{N_g}$.



Figure 3: Optimal Allocation of Types Between Unequally Sized Groups

I state the planner's problem as follows:

$$\max_{N_{1A}} N_{A1}V_1(p_{A1}) + N_{A2}V_2(p_{A1}) + N_{B1}V_1(p_{B1}) + N_{B1}V_2(p_{B1})$$
(17)

subject to

$$N_{\ell} = N_{A\ell} + N_{B\ell}, \ \ell = 1, 2$$
 (18)

$$N_g = N_{g1} + N_{g2}, \ g = A, B$$
 (19)

To simplify this problem, I consider the simple case where there is an equal number of more and less risk averse types. That is, $N_1 = N_2$. This implies that I can encompass the entire problem just by looking at one choice parameter, p_{1A} , and conditioning it on the size of the smaller group, N_A . $p_{A1} = \frac{N_{1A}}{N_A}$, and I can express $p_{A2} = 1 - p_{A1}$, $p_{B1} = \frac{2N_{B1}}{N} = \frac{2(N_1 - N_{A1})}{N}$ and $p_{B2} = 1 - p_{B2}$. Setting $N_1 = N_2$ reduces the set of constraints to three, and simple computations take account of these three constraints:

$$\max_{N_{A1}} N_{A1}V_1\left(\frac{N_{A1}}{N_A}\right) + N_{A2}V_2\left(\frac{N_{A1}}{N_A}\right) + N_{B1}V_1\left(\frac{2(N_1 - N_{A1})}{N}\right) + N_{B1}V_2\left(\frac{2(N_1 - N_{A1})}{N}\right).$$
(20)

Solving this planner's problem for an analytic solution is relatively difficult. However, it is easy to characterize the optimal allocation of types numerically. In Figure 3, I plot the objective in Problem 20 against p_{A1} , the new choice variable. To construct this example, I use parameter values from the data. I set $\sigma_c^2 = 292.88^2$ (the square of the standard error of net losses in the data),



(a) No Assortative Matching: optimal composition of insurance groups. $\tilde{\beta}_{L,1,2}^C=0.5$ and $p^U=0.5$.



(c) Some assortative matching: a suboptimal composition of insurance groups. $\tilde{\beta}_{L,1,2}^C = 0.375$ and $p^U = 0.75$.



(b) Some assortative matching in a risk sharing network. $\tilde{\beta}_{L-1,2}^C = 0.3125$ and $p^U = 0.8061$.



(d) Complete Assortative Matching: a worst case composition of insurance groups. $\beta_{L,1,2}^C = 0$ and $p^U = 1$.

Figure 4: Stylized scenarios. Yellow is more risk averse, teal is less risk averse.

 $N = 100, N_1 = N_2 = 50$, and set $\eta_1 \approx 0.0016, \eta_2 \approx 0.0037$, the average coefficients of absolute risk aversion in my data (see Section 2.2.3 for coefficients of risk aversion and types). Inspecting Figure 3, welfare is maximized when $p_{A1} = 0.5$, i.e., when diversity of types is maximized. Likewise, welfare is minimized as p_{A1} approaches 0 or 1, when diversity of types is minimized. For a demonstration that this is not an artifact of equal numbers of type 1 and type 2 agents, see Appendix B.6.

5.3 Simulating the Welfare Implications of Assortative Matching

What are the welfare implications of the degree of assortative matching? To quantify this effect, I compare among four scenarios. I list these scenarios, which are visualized in Figure 4, from high to low in terms of aggregate welfare:

- (a) Optimal scenario: The planner's optimum with equal numbers of types. This scenario features no assortative matching.
- (b) Detected insurance group scenario: takes the degree of assortative matching implied by detected insurance group SUGM estimates. Features some assortative matching.
- (c) Risk sharing network scenario: takes the degree of assortative matching implied by risk sharing network SUGM estimates. This features slightly more assortative matching than in the detected insurance group scenario.

(d) Worst case scenario: complete assortative matching.

Using the results from our SUGMs I am able to construct implied membership of groups. In the special case where all group members form a clique, I am able to directly estimate the ratio of SUGM coefficients using only the number of each type in the group. This is useful because it can give us an analytic expression of the average proportion of the majority type in each group as a function of the SUGM coefficients. By construction, the majority type will be type 1 in about half of the groups, and type 2 in the other half. Using simplifying assumptions (covered in detail in Appendix C), I am able to express the average proportion of the majority type, p^U ("p upper"):

$$p^{U} = 0.5 + 0.5 \times \sqrt{1 - \left(\frac{N-G}{N-1}\right) \left(\frac{\tilde{\beta}_{L,1,2}}{\tilde{\beta}_{ra}}\right)}$$
(21)

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

To examine these counterfactual scenarios, I use a simulation approach. Each simulation proceeds as follows: first, I sort detected insurance groups into two lists with equal total membership. If a group is assigned to the first list they will be majority type 1 and if they are assigned to the second they will be majority type 2. Second, I randomly assign individuals to detected insurance groups using a binomial process, varying the probability of assignment by scenario. Specifically, the membership of group g is N_g draws from a binomial distribution with \bar{p}^U probability of success—success being defined as a type 1 agent or a type 2 agent, depending on which should be the majority type. Third, I compute the value functions for these random assignments using the derived value functions. For details of the simulations, see Appendix C.

I simulate insurance group membership 50,000 times, compute the value functions, and plot the results in Figure 5. The results are as follows:

- (a) With no assortative matching, the optimal scenario has type 1 and type 2 agents each chosen at 0.5. The average loss is -\$136.57 (PPP).
- (b) The detected insurance group scenario has some assortative matching, as $R_{1,2} = 0.944$. I compute $p^U = 0.754$. The average loss due to risk is -\$141.37 (PPP).
- (c) The risk sharing network scenario has slightly more assortative matching, as $R_{1,2} = 0.891$. I compute $p^U = 0.774$. The average loss due to risk is -\$142.13 (PPP) in this scenario.
- (d) Finally, in the worst case scenario, there is complete assortative matching, so groups chosen as type 1 majority are completely type 1 and groups chosen as type 2 are completely type 2. The average loss due to risk is -\$156.43 (PPP).

		less Scenari	0
Scenario	(b) Group	(c) Network	(d) Worst Case
(a) Optimal	4.79	5.56	19.86
(b) Group		0.77	15.06
(c) Network			14.29

Table 5: Differences in Average Losses from Risk per Capita.

Results are averages from 50,000 simulation draws. Each entry in the table is the average per capita welfare from the scenario in the column less the average per capita welfare in the scenario in the row. Differences are in PPP Dollars.

The average differences in scenarios are presented in Table 5. Due to relatively similar degrees of assortative matching in the network and the group scenario as estimated by the SUGM, I see relatively similar degrees of welfare. However, given larger differences in the degree of assortative matching, there could be potentially be large reductions in welfare. These are bounded, holding community size and risk aversion constant, by the worst case scenario. These results are also influenced by the size of the measured coefficients of risk aversion, for which the upper bound binds for a number of respondents. Appendix B.7 discusses the impact of varying measured risk aversion on the welfare impact of assortative matching in theory.

6 Conclusion

In this paper, I explore assortative matching on risk preferences as a barrier to covariate risk sharing. Using data on risk sharing, I estimate that individuals tend to match with those people who have similar degree of risk aversion. When looking at detected insurance groups, I see a reduction in this assortative matching. Taking seriously a theoretical model of covariate risk sharing, I simulate welfare outcomes and find that the magnitude of assortative matching is small from the perspective of *ex ante* economic welfare. While I find large reductions in *ex ante* welfare due to covariate risk, the losses due to assortative matching are small when compared to the losses due to the relatively small size of risk pooling groups.

How can we square the empirical results on assortative matching with the theory above? Does the failure to achieve no assortative matching suggest that individuals are failing to maximize utility? I would not go so far. In particular, the model presented here abstracts away from issues of asymmetric information that tend to plague idiosyncratic risk sharing. Models where agents can take risky actions might provide an incentive for this type of assortative matching. Indeed, this logic is reflected in theoretical models where asymmetric information over risky actions drives assortative matching when there is heterogeneity in preferences (Attanasio et al., 2012; Wang,



Figure 5: Histogram plots of welfare losses due to risk from 50,000 simulations. Scenario means are denoted by vertical black lines.

2015). Similarly, where risk endowments differ, these serve as a driver of assortative matching (Jaramillo et al., 2015; Gao and Moon, 2016). Finally, where shocks are autocorrelated, we may find assortative matching on this dimension (Xing, 2020). This suggests that future avenues may need to balance the apparent substitution between idiosyncratic and covariate risk sharing.

Beyond exploring substitution between forms of risk sharing, the results here may also reflect a multiplexity trap, where risk sharing networks are influenced by other, seemingly unrelated networks (Cheng et al., 2021). For example, risk sharing networks might formed in dyads among co-workers. Such a story would lead to similarly preferenced individuals joining the same risk pools as we seen in our empirical exercise. In such a setting, while endogenous choice is exercised in forming relationships, this choice is both path dependent and may lead to lower utility than if each network were formed independently.

A final point, and one avenue for future exploration arises from a problem of the empirical setting: less risk averse agents tend to be more popular in risk sharing networks than their more risk averse peers. While this issue has not been rigorously modeled, intuition might suggest the opposite.²⁹ For example, if we consider risk sharing as a coping strategy for those excluded from

²⁹Some related work has been done. For example, the theoretical model in Jaramillo et al. (2015), which focuses on heterogeneity in risky endowments, relates demand to network structure. They find that those who face the least

formal risk management tools, we would expect (and possibly hope) that those who are more risk averse would demand more insurance and thus find themselves more deeply embedded in these risk sharing networks. To the contrary, more risk averse agents tend to find themselves distant from the center of networks, with fewer connections. This feature of network formation yields a puzzle and a problem for future research.

risk will be accepted by any risk sharing group.

References

- B. D. W. Allen and D. Lueck. Risk Preferences and the Economics of Contracts. *American Economic Review*, 85(2), 1995.
- A. Ambrus, M. Mobius, and A. Szeidl. Consumption Risk-sharing in Social Networks. *American Economic Review*, 104(1):149–182, 2014. doi: 10.1257/aer.104.1.149.
- 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. doi: 10.1257/app.4.2.134.
- A. Banerjee, E. Breza, A. G. Chandrasekhar, E. Duflo, M. O. Jackson, and C. Kinnan. Changes in Social Network Structure in Response to Exposure to Formal Credit Markets, 2022.
- C. B. Barrett. AMA-CRSP Survey of Risk Coping and Social Networks in Rural Ghana, 2009.
- F. Bloch, G. Genicot, and D. Ray. Informal insurance in social networks. *Journal of Economic Theory*, 143:36–58, 2008. doi: 10.1016/j.jet.2008.01.008.
- J. E. 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. doi: 10.1016/j.jdeveco.2016.01.003.
- A. Braverman and J. E. Stiglitz. Cost-Sharing Arrangements under Sharecropping: Moral Hazard, Incentive Flexibility, and Risk. *American Journal of Agricultural Economics*, 68(3):642–652, 1986. ISSN 0002-9092. doi: 10.2307/1241548.
- A. C. Cameron and D. L. Miller. Robust Inference for Dyadic Data, 2014.
- C. Castilla and T. Walker. Is Ignorance Bliss? The Effect of Asymmetric Information between Spouses on Intra-Household Allocations. *American Economic Review*, 103(3):263–268, 2013. ISSN 0002-8282. doi: 10.1257/aer.103.3.263.
- A. G. Chandrasekhar and M. O. Jackson. A Network Formation Model Based on Subgraphs, 2021.
- A. G. Chandrasekhar and R. Lewis. Econometrics of Sampled Networks, 2016.
- C. Cheng, W. Huang, and Y. Xing. A Theory of Multiplexity: Sustaining Cooperation with Multiple Relations. *SSRN Electronic Journal*, 2021. ISSN 1556-5068. doi: 10.2139/ssrn.3811181.
- 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. doi: 10.3982/QE131.
- D. Collins, J. Morduch, S. Rutherford, and O. Ruthven. *Portfolios of the Poor: How the World's Poor Live on \$2 a Day.* Princeton University Press, 2010.
- M. Comola and M. Fafchamps. Testing unilateral and bilateral link formation. *The Economic Journal*, 124(579):954–976, 2014. doi: 10.1111/ecoj.12071.
- M. Comola and M. Fafchamps. The missing transfers: Estimating misreporting in dyadic data. *Economic Development and Cultural Change*, 65(3):549–582, 2017. doi: 10.1086/690810.
- T. G. Conley and C. R. Udry. Learning about a New Technology: Pineapple in Ghana. *American Economic Review*, 100(1):35–69, 2010. ISSN 0002-8282. doi: 10.1257/aer.100.1.35.
- J. De Weerdt. Risk-sharing and endogenous network formation, 2002.
- J. De 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.
- A. Demirguc-Kunt, L. Klapper, D. Singer, S. Ansar, and J. Hess. The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution. Technical report, World Bank Group, 2018.

- F. Dizon, E. Gong, and K. Jones. The Effect of Promoting Savings on Informal Risk-Sharing: Experimental Evidence from Vulnerable Women in Kenya. *Journal of Human Resources*, 55(3): 963–998, 2019. doi: 10.3368/jhr.55.3.0917-9077R2.
- P. Dupas, A. Keats, and J. Robinson. The effect of savings accounts on interpersonal financial relationships: Evidence from a field experiment in rural Kenya. *The Economic Journal*, 129 (617):273–310, 2019. doi: 10.1111/ecoj.12553.
- C. Elbers, J. W. Gunning, and B. Kinsey. Growth and risk: Methodology and micro evidence. *World Bank Economic Review*, 21(1):1–20, 2007. doi: 10.1093/wber/lhl008.
- M. Elliott and B. Golub. Networks and Economic Fragility. *Annual Review of Economics*, 14(1):665–696, Aug. 2022. ISSN 1941-1383, 1941-1391. doi: 10.1146/annurev-economics-051520-021647.
- M. Fafchamps. Rural poverty, risk and development. (October):262, 2003.
- 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.
- M. Fafchamps and S. Lund. Risk-sharing networks in rural Philippines. *Journal of Development Economics*, 71:261–287, 2003. doi: 10.1016/S0304-3878(03)00029-4.
- 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. *Physical Review E*, 69, 2004. doi: 10.1103/PhysRevE.69.026113.
- M. Goldstein and C. Udry. The profits of power: Land rights and agricultural investment in Ghana. *Journal of Political Economy*, 116(6):981–1022, 2008. ISSN 00223808. doi: 10.1086/595561.
- B. S. Graham. An Econometric Model of Network Formation With Degree Heterogeneity. *Econometrica*, 85(4):1033–1063, 2017. ISSN 0012-9682. doi: 10.3982/ecta12679.
- B. S. Graham. Dyadic regression. In *The Econometric Analysis of Network Data*, pages 23–40. 2020. doi: 10.1016/b978-0-12-811771-2.00008-0.
- I. Günther and K. Harttgen. Estimating Households Vulnerability to Idiosyncratic and Covariate Shocks: A Novel Method Applied in Madagascar. *World Development*, 37(7):1222–1234, July 2009. ISSN 0305750X. doi: 10.1016/j.worlddev.2008.11.006.
- S. Hart and M. Kurz. Endogenous Formation of Coalitions. *Econometrica*, 51(4):1047–1064, 1983.
- H. Henderson and A. Alam. The structure of risk-sharing networks. *Empirical Economics*, 62(2): 853–886, Feb. 2022. ISSN 0377-7332, 1435-8921. doi: 10.1007/s00181-021-02037-z.
- W. C. Horrace and R. L. Oaxaca. Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters*, 90(3):321–327, Mar. 2006. ISSN 01651765. doi: 10.1016/j.econlet.2005.08.024.
- 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. doi: 10.1257/aer.104.1.183.
- M. O. Jackson and A. Wolinsky. A Strategic Model of Social and Economic Networks. *Journal of Economic Theory*, 71(1):44–74, 1996. ISSN 00220531. doi: 10.1006/jeth.1996.0108.
- M. O. Jackson, T. Rodriguez-Barraquer, and X. Tan. Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*, 102(5):1–45, 2012. doi: 10.1257/aer.102. 5.1857.
- F. Jaramillo, H. Kempf, and F. Moizeau. Heterogeneity and the formation of risk-sharing coalitions. *Journal of Development Economics*, 114:79–96, 2015. ISSN 03043878. doi: 10.1016/j. jdeveco.2014.09.003.

- D. Kahneman and A. Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–292, 1979. ISSN 0012-9682.
- D. Karlan, R. Osei, I. Osei-Akoto, and C. Udry. Agricultural Decisions After Relaxing Credit and Risk Constraints. *The Quarterly Journal of Economics*, 129(2):597–652, 2014. doi: 10.1093/qje/qju002.
- C. Kinnan. Distinguishing barriers to insurance in Thai villages. *Journal of Human Resources*, 57 (1):44–78, 2021. doi: 10.3368/jhr.57.1.0219-10067R.
- E. Ligon and L. Schechter. Measuring Vulnerability. *The Economic Journal*, 113(486):95–102, 2003.
- M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27:415–444, 2001. doi: 10.1146/annurev.soc.27.1.415.
- M. E. J. Newman. Communities, modules and large-scale structure in networks. *Nature Physics*, 8(1):25–31, 2012. doi: 10.1038/nphys2162.
- P. Pons and M. Latapy. Computing communities in large networks using random walks. In *Lecture Notes in Computer Science*, volume 3733 of *International Symposium on Computer and Information Sciences*, pages 284–293. 2005. doi: 10.1007/11569596_31.
- D. Putman. *Social Network Structure and Risk in the Developing World*. PhD thesis, University of California, Davis, 2020.
- 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.
- T. Sargent. Macroeconomic Theory. 1987.
- J. A. Smith, J. H. Morgan, and J. Moody. Network sampling coverage III: Imputation of missing network data under different network and missing data conditions. *Social Networks*, 68:148–178, Jan. 2022. ISSN 03788733. doi: 10.1016/j.socnet.2021.05.002.
- J. E. Stiglitz. Incentives and Risk Sharing in Sharecropping. *Review of Economic Studies*, 41(2):219, 1974. ISSN 00346527. doi: 10.2307/2296714.
- J. E. Stiglitz. Globalization and Its Discontents. 2003.
- A. Suzuki, L. S. Jarvis, and R. J. Sexton. Partial Vertical Integration, Risk Shifting, and Product Rejection in the High-Value Export Supply Chain: The Ghana Pineapple Sector. *World Development*, 39(9):1611–1623, 2011. ISSN 0305750X. doi: 10.1016/j.worlddev.2011.02.007.
- M. Tabord-Meehan. Inference With Dyadic Data: Asymptotic Behavior of the Dyadic-Robust t-Statistic. *Journal of Business & Economic Statistics*, 37(4):671–680, 2019. doi: 10.1080/07350015. 2017.1409630.
- R. M. Townsend. Risk and Insurance In Village India. *Econometrica*, 62(May):539–591, 1994. doi: 10.2307/2951659.
- C. R. Udry and T. G. Conley. Social networks in Ghana. In C. B. Barrett, editor, *The Social Economics* of *Poverty*. 2005. ISBN 0203799658.
- J. Vanderpuye-Orgle and C. B. Barrett. Risk management and social visibility in Ghana. *African Development Review*, 21(1):5–35, 2009. ISSN 10176772. doi: 10.1111/j.1467-8268.2009.00201.x.
- 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.
- Y. Xing. Who Shares Risk with Whom and How? Endogenous Matching Selects Risk Sharing Equilibria, 2020.

J. Zscheischler, S. Westra, B. J. J. M. van den Hurk, S. I. Seneviratne, P. J. Ward, A. Pitman, A. AghaKouchak, D. N. Bresch, M. Leonard, T. Wahl, and X. Zhang. Future climate risk from compound events. *Nature Climate Change*, 8(6):469–477, June 2018. ISSN 1758-678X, 1758-6798. doi: 10.1038/s41558-018-0156-3.

A Empirical Appendix

A.1 Details of Network Elicitation

The network elicitation is based on the set of names of everyone else in the sample, which were pre-printed on the enumeration form. For each pre-printed name, respondents are asked if they know who this person is and if they know them personally. If they answer to either of these questions is "no," the enumerator moves on to the next name. If they answer "yes" to both, they are asked how the person is related to them, how long they've known them, if they consider them a friend, frequency of communication, if they would trust the person to look after a valuable item for them, and if they've received or given gifts. Two of these questions appear most useful in measuring risk sharing networks. First, gift networks may serve as useful networks as they involve the transfer of cash, in kind goods, or services, such transfers being an integral part of risk sharing. The questions posed to the respondents are: "Have you ever received a gift (of money, goods, or services) from this person?" and "Have you ever given a gift (of money, goods, or services) to this person?" Second, as risk sharing trust networks often provide assistance as credit, trust networks may provide useful in measurement. The survey asks "Would you trust this person to look after a valuable item for you?"

Respondents tend to report reciprocal gift giving. Respondent i who reports giving a gift to j also tends to report receiving a gift from j. Yet, answers often differ between those asked. For example, i reports giving gifts to person j, but j does not report receiving. While people tend to forget the gifts they've received more readily than those they give (true here as it is elsewhere, e.g., Comola and Fafchamps, 2014), this does not account for the majority of the reduction. Instead, they report receiving gifts from others, who did not claim to give. Likewise, in the trust network, we find that often trust is not reciprocated. This puzzle might be explained by the elicitation method. In particular, a prerequisite to both the gift and trust questions is the question "Do you know them personally?" If i regards j as a personal connection they are asked about j and may report gift giving. However, if j does not regard i as a personal connection, then they will not be queried about gift giving.

Comola and Fafchamps (2014) outlines a procedure to determine how to construct networks when there is such discordance. While I do not currently implement their procedure, it is possible to do so with this data, and this could represent an expansion of the analysis. However, I present it here as their approach empirically differentiates between three stories of link formation, which are useful mental models in constructing this network: First, *desire to link*, where reported links in the survey elicitation represent possible links that might be made if the need for help arises. Second, *bilateral link formation*, where both must agree to form a link (i.e., links are pairwise stable a la Jackson and Wolinsky, 1996). Third, *unilateral link formation*, where the action of one party is sufficient to form a link. For unilateral and bilateral link formation, discordant answers may occur due to misreporting (perhaps forgetting to report a connection).

When considering our network, it is important to note that misreporting is less likely than in other network surveys because the network elicitation differs considerably from other work. For example, in the Nyakatoke data (used in Comola and Fafchamps, 2014, 2017), asks the question "Can you give a list of people from inside or outside of Nyakatoke, who you can personally rely on for help and/or that can rely on you for help in cash, kind or labour?" Comola and Fafchamps (2014) finds that these measure desire to link risk sharing networks. In contrast, and as discussed above, respondents in this survey are queried about all others in the sample as opposed to asked to produce a list, suggesting misreporting to be a considerably smaller concern. This means that we can more readily think of discordant links as differences of opinion between respondents (as opposed to misreporting). Additionally, as one of our questions ask about gifts given in the past, as opposed to potential future risk sharing transfers. This reduces the likelihood of finding desire to link networks as opposed to hypothetical questions.

A.2 Construction of Networks

A.2.1 Gift and Trust Networks

I start by constructing i's gift and trust networks, constructing a bilateral gifts network and a bilateral trust network. In the bilateral gifts network a link occurs between i and j if i reports reciprocal giving (i received and gave to j), and vice versa. In the bilateral trust network, a link occurs if i reports trusting j and j reports trusting i. Then using the gifts and trust networks, I construct a possible risk sharing network, where i and j are linked in the risk sharing network if they are in both the gifts and the trust network.

A.2.2 Insurance Organizations Network

I also have access to data about organizational membership in the village. These include a short description of type of organization and the organization name. I form a co-membership network which draws on organizations related to mutual insurance. In particular, while funeral groups are easy to observe, I comb through the data to find other insurance related organizations. I select those organizations defined by any member as "assistance," "support," or "welfare" (e.g., "support group" or "welfare organisation") as mutual insurance organizations. In order to reduce measurement error, I first clean group names and types. For names, I harmonize group names with small differences (e.g., "kuw" vs. "kuo," both meaning group). Second, I clean group types to remove health, advice, trade unions, and religious groups that fall under funeral or mutual insurance labels.
	Gifts	Trust	Risk Sharing	Insurance Orgs	Combined
Panel A: Netwo	rk Densitie	es			
Density	0.0614	0.0476	0.0329	0.0137	0.0457
Panel B: Correl	ations				
Gifts		0.587	0.721	0.015	0.609
Trust			0.825	0.026	0.701
Risk Sharing				0.018	0.842
Insurance Orgs					0.538

Table A1: Comparison of risk sharing networks using trust, gifts, and insurance group membership.

Gifts is an indicator for if i and j report giving and receiving gifts (i.e., equal to one if true, zero otherwise). Trust is an indicator for if both i and j would trust the other to look after a valuable item. Risk sharing is the interaction of trust and gift. Insurance orgs is an indicator for if respondents were co-members in a funeral insurance, support, assistance, or welfare group. Finally, combined is an indicator for if there is a link in either the risk sharing or the insurance orgs network. Panel A presents network density. Panel B presents spearman correlations between networks.

I construct a co-membership network based on group names. Additionally, some reported memberships may have lapsed, as many report being in a group, but not having attended any meetings in the past year. If respondents are not attending meetings this makes these groups an unlikely place for informal transactions to take place. In this network a link is formed whenever i and j are current members group together (defined as having attended in the past year). I then restrict this network to only co-memberships in insurance organizations. To do so, again if one of the members must describe it this way.³⁰ I refer to this network as the insurance organizations network.

A.2.3 Family Network

For the family network, I use the relationship codes collected to identify close family. In this definition, family includes spouses, children, step-children, parents, grandparents, and grand-children. These relationships are lineal marriage related ties as well as collateral ties such as siblings.

³⁰This is to correct for the fact that many lines in the organization membership data do not feature descriptions, despite the same organization being described elsewhere. This is takes place after a round of cleaning of group types, to avoid groups that are spuriously identified under these labels (e.g., "herbalist society" as a welfare group).

A.2.4 Risk Sharing Network

There is no-one-size-fits-all approach to network construction. Constructing a network depends on what role the network plays and how the network is elicited, among other considerations. I construct a two risk sharing networks. First, I use the bilateral gift and trust network, which I will use in the main text. Given past documentation of the importance of reciprocal relationships in risk sharing (e.g., Fafchamps, 2003; Blumenstock et al., 2016), I prefer bilateral networks to unilateral networks. Furthermore, due to the exhaustive elicitation method, there is no reason to suspect under reporting of links and the networks generated by this approach are not particularly sparse. In fact, unilateral links may suggest desired links that may or may not provide support in a time of need.

Second, for robustness, I construct a second combined risk sharing network using the insurance organization network and the first risk sharing network (i.e., bilateral trust and gift network): i and j are linked in the risk sharing network if a link is present in either the gift and trust network or the organization network. These two networks represent different spheres of risk sharing, and are not strongly correlated (Pearson correlation, r = 0.018). In some sense, the organization network speaks for itself as groups are labeled for some kind of risk sharing. However, many village members belong to no organizations and would therefore appear to have no risk sharing connections if we used only the organizations network. Additionally, as we will see below, these organizations do not do well in predicting informal transfers.

As I am considering the formation of risk sharing networks, I opt not to include upstream relationships, i.e., those which might lead to formation of risk sharing networks. These include relationships like family ties, geography, and ethnicity (e.g., Fafchamps and Gubert, 2007). I instead reserve these for later analysis/controls.

A.3 Validation of Risk Sharing Networks

One approach to validate my choices in risk sharing networks is important to document that current transfers align not only with the network in question, but that these transfers are made in response to shocks. I estimate a directed dyadic regression, which takes the form

$$\tau_{ij} = \alpha + \beta \ a_{ij} + \gamma \tilde{y}_i + \delta \ a_{ij} \times \tilde{y}_i + \varepsilon_{ij} \tag{22}$$

where τ_{ij} is a transfer (a loan or gift) from *i* to *j*, a_{ij} is the network of interest, \tilde{y}_i is the idiosyncratic income shock to *i*. If a_{ij} is an informal risk sharing network, I expect that $\delta > 0$, or that those with positive shocks relative to others give loans and gifts within the network. I take this as evidence that these networks are representative of risk sharing networks. Additionally, I expect that $\beta > 0$, i.e., transfers happen more often within these networks in general, and $\gamma > 0$, those with positive shocks relative to others tend to give gifts or lend money.

		Any t	ransfer from	i to j	
	(1)	(2)	(3)	(4)	(5)
Trust	0.0118^{***} (0.00262)				
Trust $\times \; i$ won lotto	0.00723 (0.00456)				
Gifts		0.0121^{***}			
Gifts $\times \; i$ won lotto		(0.00223) 0.00577 (0.00382)			
Risk Sharing			0.0171***		
Risk Sharing $\times \; i$ won lotto			(0.00394) 0.0108 (0.00669)		
Insurance Orgs				0.00427	
Insurance Orgs $\times i$ won lotto				(0.00314) -0.000761 (0.00397)	
Combined					0.0131***
Combined $\times i$ won lotto					(0.00232) 0.00460 (0.00469)
i won lotto	0.000760*	0.000710*	0.000733*	0.00112**	0.000771*
Constant	(0.000311) 0.00105^{***} (0.000162)	(0.000295) 0.000884*** (0.000137)	(0.000306) 0.00105^{***} (0.000161)	(0.000369) 0.00154^{***} (0.000177)	(0.000310) 0.00105^{***} (0.000156)
\overline{N}	99704	99704	99704	99704	99704

Table A2: Validating risk sharing networks with current transfers and lottery winnings

*p<0.1; **p<0.05; ***p<0.01. Dyadic robust standard errors in paratheses (Fafchamps and Gubert, 2007). The outcome is an indicator for if any transfer, i.e., a loan or a gift, was made from i to j during the study period. i won lottery is an indicator for if respondent i won a lottery during the study period. Gifts is an indicator for if i and j report giving and receiving gifts (i.e., equal to one if true, zero otherwise). Trust is an indicator for if both i and j would trust the other to look after a valuable item. Risk sharing is the interaction of trust and gifts. Insurance orgs is an indicator for if respondents were co-members in a funeral insurance, support, assistance, or welfare group. Finally, combined is an indicator for if there is a link in either the risk sharing or the insurance orgs network.

To operationalize these variables I use records of gifts and loans within villages within the data as my measure of transfers and lotteries performed by researchers as random shocks. I restrict loans and gifts to those given between identified villagers. For loans, I restrict to those given for consumption or emergency uses. These include loans for consumption, medical bills, funerals or other ceremonies, and school fees, as well as miscellaneous labelled uses (one loan for a court case and one for an electricity bill). I restrict gifts as well, removing gifts for holidays (Christmas, New Years, Valentines, Birthdays), bartering, and remittances. Both loans and gifts include both cash and in kind transfers. Lottery and prize winnings are collected in the shocks module of the survey (these are labeled "lottery", "lotto", "prize winnings", "Lottery or prize winnings"). These lotteries were administered via drawings of bottle caps. Prizes varied between 10 cedis and 70 cedis. Some were given in cash and others in kind (livestock vouchers). With this data, τ_{ij} is replaced with an indicator equal to one if *i* gave a gift or a loan to *j* and zero otherwise, and \tilde{y}_i is an indicator variable equal to one if *i* won any lottery and zero otherwise.

Table A2 presents the results of this dyadic regression. While limited somewhat by the data used for this exercise, the results are consistent with a social network that is used for risk sharing. I find that the risk sharing network (i.e., using bilateral gifts and trust) explains transfers from those who have received lottery winnings to those in their networks. First, across networks we tested, we see that those who received prize winnings are more likely to transfer money to others. Second, across most of the networks, network connections increase the probability of transfers. This is particularly pronounced in the chosen risk sharing network, where the probability of a transfer increases 1.7 percentage points. Those who won the lottery were 1.1 percentage points more likely to give a gift or grant a loan to those in their network (Table A2, column 3), though I cannot reject the null hypothesis at standard significance levels when using dyadic robust standard errors. Nevertheless, the marginal increase probability of giving a transfer to someone in the risk sharing network member is over ten times the marginal effect on giving to a random village member. More to the point, this network had the strongest risk sharing response among the networks tested.

The large standard errors on this estimate are likely due to two reasons, First, the fact that the network of gifts and loans in this period are from recall, and therefore, consequently, the network is relatively sparse (in particular, about 0.2% of directed dyads feature a transfer) (Comola and Fafchamps, 2017). Similarly, many did not receive lottery winnings, and only 1.1% of directed risk sharing network ties feature i as a lottery winner. Therefore when accounting for correlations between ties, if there is correlation in people's decisions to make transfers, this will be reflected in the higher standard errors.³¹

 $^{^{31}}$ Indeed, when running the same regression with heterosked asticity robust standard errors, I find that this estimate is significant at the 10% level

Taken together, this suggests that the risk sharing network I have identified is capturing risk sharing behavior as well as any we could construct, and as opposed to more prosaic favor exchange. Notably, and in contrast, the insurance organization network is not as strong a predictor of transfers, and its addition to the risk sharing network (in combined) reduces the size of the average risk sharing transfers considerably. While the organizational transfers within these groups may simply be imperfectly captured here, it nevertheless indicates that this network not be the most relevant for this study. Nevertheless, we will perform robustness checks using it.

A.4 Community Detection

A.4.1 Walktrap Algorithm

At a high level, the Walktrap algorithm proceeds as follows (Pons and Latapy, 2005):

- 1. To start, each node is assigned to its own community. Compute distances for all adjacent communities. See Appendix A.4.2 for a description of the computation of distances.
- 2. Merge the two adjacent communities with the smallest distance into one community.
- 3. Recompute the distances between communities.
- 4. Repeat steps 2 and 3 until all communities have been merged into one, recording the order of merges in a dendrogram (a hierarchical diagram documenting community merges).
- 5. Using the dendrogram, compare the modularity of all possible community assignments and choose the one with the highest modularity. See Appendix A.4.3. for the computation of modularity.

A.4.2 Computing Distances using Random Walks

The Walktrap algorithm uses random walks to compute node similarity (Pons and Latapy, 2005). A random walk proceeds as follows: A random walker starts at node i and moves to an adjacent node with probability $1/d_i$ (where d_i is the degree of i). This process is repeated from the landing node, k, moving to an adjacent node with probability $1/d_k$, a total number of s times. If nodes are in the same community, random walks of length s from nodes i and j should often land on the same nodes. Of course, nodes with higher degree will more often receive these walks, so the measure of distance takes account of the degree of receivers.

$$r_{ij}(s) = \sqrt{\sum_{k=1}^{n} \frac{(P_{ik}^s - P_{jk}^s)^2}{d_k}}.$$
(23)

where P_{ik}^s is the probability that a walk starting at node *i* ends its walk on node *k*. The distance overall can be thought of as the L^2 distance between P_{ik}^s and P_{ik}^s .

Building on this definition, the authors also define the distance between communities:

$$r_{C_1,C_2}(s) = \sqrt{\sum_{k=1}^n \frac{(P_{C_1,k}^s - P_{C_2,k}^s)^2}{d_k}}.$$
(24)

where the source of the random walk is drawn randomly and uniformly from members of that community: $P_{C,k}^s = \frac{1}{|C|} \sum_{i \in C} P_{ik}^s$.

A.4.3 Modularity

Modularity measures the internal quality of the community by looking at how many links exist within the community compared to how many would be expected at random (Newman, 2012). 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. The average number of within community links from rewiring is used as a counterfactual. Having many more links within the community than the counterfactual implies a good community detection. Fewer implies a poor community structure.

To compute modularity, let d_i and d_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 $d_i d_j / (2m - 1) \approx d_i d_j / 2m$. 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 $A_{ij} - \frac{d_i d_j}{2m}$. I can interpret this as connections over expected connections in a random graph conditional on node pair degrees. Then, these values are weighted by if they reside in the same community, i.e., if $c_{ij} = 1$. Finally, I aggregate to the graph level and normalize by twice the number of links present:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] c_{ij}$$

This serves as an easily computable and straightforward measure of the internal quality of communities.

A.4.4 Edge Betweenness Community Detection

A different approach to detecting communities come from Girvan and Newman (2004). This algorithm utilizes *edge betweenness*—the edge counts of shortest paths through the network. To do so, find the shortest path in the network between each pair of nodes, the count the number of paths that pass through each edge, awarding partial credit if there are multiple shortest paths between nodes. The intuition of this method is that these serve to constrain the flow of information and transfers themselves. This makes monitoring or learning about others difficult.

The algorithm proceeds as follows:

- 1. Compute edge betweenness for all edges
- 2. Find the edge with the highest betweenness, and remove it from the network
- 3. Recalculate betweenness for all edges that remain
- 4. Repeat until all edges have been removed

This leaves us with a set of potential community assignments based on network components (connected subgraphs). Every time the network is split into multiple components, this is a potential community assignment. As in the Walktrap algorithm, these are compared using modularity, and the assignment with the highest modularity is selected.

A.4.5 Community Detection Results

Walktrap community detection results are presented in Figure A1, with adjustments to visualization to help the reader see community structure. Visual inspection of the community detection yields a number of interesting points. First, within the giant component (the largest connected portion of the network) we tend to see large central communities, medium-sized offshoots, and small marginal communities. Second, outside of this giant component, communities coincide with the small components. This pattern parallels the social visibility findings of Vanderpuye-Orgle and Barrett (2009).

The community assignments produced by Edge Betweenness are similar to those from Walktrap. Community co-membership between the two methods has a relatively strong correlation, around 0.6 (see Table A3). In contrast, the detected communities are less closely related to the original risk sharing network, with correlations around 0.4. The communities are more dense than the risk sharing network, and are composed a bit differently. The detected insurance groups discard about one third of the ties in the risk sharing network and fill in a large number more, more than doubling the number of ties (see Table A4). As is suggested by the correlation, the ties kept from the network and the ties added are similar between methods.

Drawing on the literature, the differences in the risk sharing network and the detected insurance groups are not arbitrary. If edges tend to cross communities this is indicative that these edges are less likely to be activated for the purposes of risk sharing (Putman, 2020). Conversely,

	Edge Betweenness	Walktrap
Risk Sharing	0.418	0.394
Edge Betweenness		0.590

Table A3: Correlations Between Community Detection Methods

those nodes that that are not adjacent in the original risk sharing network but are part of the same community are more likely than other similar nodes (e.g., at the same network distance) to be incorporated into informal insurance. This is either due to flows over networks (similar to results in De Weerdt and Dercon, 2006) or matching with new risk sharing partners (Putman, 2020). I interpret these detected insurance groups as the furthest extent of risk sharing in these networks.

Table A4: Comparison of Detected Insurance Groups to Risk Sharing Network

			Number of Edges in					
Method	Density	Diff.	Both	Only RSN	Only Detected	Neither		
Walktrap	0.0788	-0.0459***	2106	1110	5600	88956		
Edge Betweenness	0.0746	-0.0417***	2156	1060	5134	89422		

*p<0.1; **p<0.05; ***p<0.01. Diff. is the risk sharing network density (0.0329) less the detected insurance group co-membership network. 'Both' indicates the edge exists in both networks, 'Only RSN' are those edges that appear in the risk sharing network but not the detected insurance group co-membership network, 'Only Detected' is the opposite of Only RSN, and 'Neither' indicates the edge exists in neither network.



Figure A1: Risk sharing networks in villages with walktrap community detection assignment overlaid. Nodes are individuals and edges are links in the risk sharing network. Detected communities are represented by shaded regions and node colors.

A.5 Risk Preferences

A.5.1 Hypothetical Gambles

Amounts for each hypothetical gamble are presented in Table A5. The choice between gambles is framed around choice to purchase agricultural inputs. In the gains domain, the gambles are framed around fertilizer and in the losses domain, they are framed around insecticide. While the gambles themselves are not normally distributed, $y_B - E(y_B)$ is both distributed symmetrically around zero and relatively small compared to incomes.

			First Questionnaire									
	Prob.	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
y_A	100%	85	90	95	100	105	110	115	120	125	130	135
	50%	20	20	20	20	20	20	20	20	20	20	20
y_B	50%	200	200	200	200	200	200	200	200	200	200	200
				S	Second	Questi	onnair	e				
	Prob.	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9		
y_A	100%	80	80	80	80	80	80	80	80	80		
	50%	40	40	40	40	40	40	40	40	40		
y_B	50%	110	115	120	125	130	135	140	145	150		

Table A5: Hypothetical Gamble Questionnaires

Each set of questions was asked in the domain of gains and the domain of losses, for a total of four sets of questions. Amounts are in Ghanaian Cedis (about 0.54 GHC/\$PPP, so 200 GHC would be around \$370 PPP in 2009). The script proceeded from midpoint gamble (Q6 in the first questionnaire and Q5 in the second) to the direction implied. For example, choosing A in Questionnaire 1 Q6 would direct you to Q5, which lowers the sure value of A. The elicitation ends on the question where the respondent switches from their choice.

A.5.2 Exponential Preferences with Normally Distributed Shocks

For each gamble, let Y_A be constant and let Y_B be normally distributed. For an agent with CARA preferences, I represent expected utility as a mean variance utility function (Sargent, 1987).

$$EU_i(Y) = E(Y) - \frac{\eta_i}{2}V(Y)$$
(25)

Respondents are able to choose between two gambles y_A and y_B , and will be indifferent between the two when

$$E(y_B) - \frac{\eta_i}{2}V(y_B) = y_A.$$

If an individual reaches a point of indifference between two gambles, I assign them to the midpoint between the two gambles. Hence, if the mean differs, I take the average of the mean of the two gambles and assign this value to the point of indifference. If the variance differs, I 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. Then we can express risk aversion for agent i as a function of their indifference point,

$$\eta_i = \frac{2(E(y_B) - y_A)}{V(y_B)}$$

and recover the coefficient of absolute risk aversion.

A.5.3 Alternative Approaches to Risk Preferences

While it is useful to compute risk preferences as outlined above, the assumption of asymptotic normality may not be defensible in every case. To check the robustness of results to this assumption, I compute CARA coefficients while adjusting two assumptions.

First, to compute coefficients of risk aversion without making the assumption of normally distributed shocks, we must draw directly on the exponential utility function. At the point of indifference, the utility of the two gambles should be equal:

$$\frac{1 - e^{-\eta_i y_A}}{\eta_i} = 0.5 \left(\frac{1 - e^{-\eta_i y_B^L}}{\eta_i} \right) + 0.5 \left(\frac{1 - e^{-\eta_i y_B^H}}{\eta_i} \right)$$
(26)

where y_B^L is the low payout and y_B^H is the higher payout of choice B. If $\eta_i \neq 0$, this reduces to the following equality:

$$2e^{-\eta_i y_A} - e^{-\eta_i y_B^L} - e^{-\eta y_B^H} = 0$$
⁽²⁷⁾

Using this equality, I use a numerical solver to compute η_i at this indifference points implied by the questions. If an individual reaches a point of indifference between two gambles, I assign them to the midpoint between the two gambles. Hence, I take the average of the coefficients implied by the two questions. Figure A3(a) compares the method to this alternative method, and finds that while coefficients of risk aversion are almost perfectly co-linear, those used in the paper are smaller than those that would be computed without the normality assumption, implying lower degrees of risk aversion or risk loving.



Figure A2: Histogram and distribution of risk preferences within the four villages. The histogram fill, which depicts the measured degree of risk aversion from the hypothetical gambles, is matched with Figures 2, A4, A5, and A6. Vertical lines indicate distinctions between types, which are annotated on the x-axis.



(a) Normally distributed vs. no dis- (b) Normally distributed with paper (c) Normally distributed vs. no distributional assumption with alterna- computation vs. alternative compu- tributional assumption with same (altive computation.

Figure A3: Comparison of CARA Coefficients by Underlying Assumptions

However, the computation has a second assumption embedded in it. As opposed to averaging the mean and variance of income before computing the coefficient, it averages the coefficients after computing them at the mean and variance of the gambles. This difference, while subtle, actually accounts for the majority of the difference in the estimates of risk aversion. The second two subfigures in Figure A3 decompose these two assumptions. Figure A3(b) shows the results of changing the computation method (computing preferences first and then averaging), while Figure A3(c) shows the result of changing the distributional assumption, holding computation constant. This shows that in this case the assumption of asymptotic normality should be treated as relatively benign.



Figure A4: Darmang Risk Sharing Networks with Risk Preferences Indicated by Color



Figure A5: Oboadaka Risk Sharing Networks with Risk Preferences Indicated by Color



Figure A6: Konkonuru Risk Sharing Networks with Risk Preferences Indicated by Color

A.6 Robustness Checks

A.6.1 Selection-on-Observables Results: Addressing Popularity and Homophily

Assortative matching on risk preferences could reflect assortative matching on some other social or economic dimension. In addition to the inclusion of kinship and risk aversion, my approach for controlling for other observables related to popularity and homophily will be straightforward. Homophily is a common feature of social networks and is similarly present in the context of risk pooling (?Fafchamps and Gubert, 2007; Jaramillo et al., 2015).

Table A6 presents results from the selection-on-observables approach. I control for if the pair is married, are co-wives, have the same occupation, are the same gender, are (additionally) both men, have the same level of schooling, are both primary, secondary, or tertiary educated (no/missing education left out), and for sums and absolute differences in: age, and family network degree centrality. Additionally, all regressions feature village fixed effects. In general, the magnitude of β_1 falls when controls are included. For example, in Column (2), the estimate of β_1 fall slightly to -0.00425 and is no longer statistically significant. However, in column (5), statistical significance returns (at the 5% level) with just a marginally higher estimate. Discounting changes in statistical significance, the magnitudes of the effects when controlling for this battery of related factors is very consistent, suggesting a real—if subtle—relationship.

Table A7 presents results from the selection-on-observables approach for detected insurance groups. I control for if the pair is married, are co-wives, have the same occupation, are (additionally) both men, have the same level of schooling, are both primary, secondary, or tertiary educated (no/missing education left out), and for sums and absolute differences in: age, and family network degree centrality. as above, all regressions feature village fixed effects. One small difference in the controls included is made for practical reasons. In an idiosyncrasy, all detected insurance group specifications including same gender as a control variable result in highly singular variance matrices when estimating dyadic robust standard errors. As it is clear these errors matter for this application, I opt to remove this variable from the control function to ensure robust inference. Here have similar effect sizes to the main specification, and noisy standard errors.

A.6.2 Logistic Regression Results

I estimate assortative matching using a dyadic linear probability model because this allows me to utilize village fixed effects in my specifications. However, logistic regression is typically considered a more appropriate approach for binary dependent variables, including in network formation models, as when predictions are outside of the unit interval, coefficient estimates could suffer from inconsistency (Horrace and Oaxaca, 2006). Therefore, to ensure my choice of specification does not influence the estimates of assortative matching, I replicate Tables 2 and 3 here

	(1)	(2)	(3)	(4)	(5)
$\overline{ \eta_i - \eta_j }$	-0.00235	-0.00425	-0.00250	-0.00437	-0.00430*
	(0.00188)	(0.00236)	(0.00184)	(0.00229)	(0.00208)
$\eta_i + \eta_j$		-0.00214		-0.00211	-0.00211
		(0.00177)		(0.00175)	(0.00174)
Family _{ij}			0.194***	0.194***	0.195***
			(0.0149)	(0.0149)	(0.0191)
Family _{ij} $\times \eta_i - \eta_j $					-0.00105
					(0.0106)
Village FE	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes
N	65102	65102	65102	65102	65102
R^2					

Table A6: Dyadic Regression: Risk Sharing Network with Controls

Dyadic robust standard errors are reported in parentheses (Fafchamps and Gubert, 2007). All specifications are dyadic linear probability models with matching in the risk sharing network as the dependent variable. η_i is risk aversion of individual *i*, so $|\eta_i - \eta_j|$ is the absolute difference of risk aversion while $\eta_i + \eta_j$ is the sum. Both absolute differences and sums of risk aversion are transformed into *z*-scores. Controls include married, are cowives, have the same occupation, are the same gender, are both men, have the same level of schooling, are both primary, secondary, or tertiary educated (no/missing education left out), and for sums and absolute differences in: age and family network degree centrality. * p < 0.05, ** p < 0.01, *** p < 0.001

	(1)	(2)	(3)	(4)	(5)
$\overline{ \eta_i - \eta_j }$	-0.00327	-0.00622	-0.00338	-0.00632	-0.00703
	(0.00456)	(0.00547)	(0.00453)	(0.00544)	(0.00541)
$\eta_i + \eta_j$		-0.00333		-0.00332	-0.00333
		(0.00426)		(0.00425)	(0.00426)
Family _{ij}			0.155***	0.155***	0.145***
			(0.0190)	(0.0190)	(0.0252)
Family _{<i>ij</i>} × $ \eta_i - \eta_j $					0.0110
					(0.0162)
Village FE	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes
N	65102	65102	65102	65102	65102
R^2					

Table A7: Dyadic Regression: Detected Insurance Group with Controls

Dyadic robust standard errors are reported in parentheses (Fafchamps and Gubert, 2007). All specifications are dyadic linear probability models with matching in the risk sharing network as the dependent variable. η_i is risk aversion of individual *i*, so $|\eta_i - \eta_j|$ is the absolute difference of risk aversion while $\eta_i + \eta_j$ is the sum. Both absolute differences and sums of risk aversion are transformed into *z*-scores. Controls include married, are co-wives, have the same occupation, are both men, have the same level of schooling, are both primary, secondary, or tertiary educated (no/missing education left out), and for sums and absolute differences in: age, (family network) degree centrality, betweenness centrality, and eigenvector centrality. * p < 0.05, ** p < 0.01, *** p < 0.001

using logistic regression. The results of logistic dyadic regression are estimated for the risk sharing network in Table A8 and the detected insurance groups in A9. These replicate the pattern of results from the dyadic regressions in the main text. For the risk sharing network, we document stronger assortative matching when using the sum of risk aversion to control for popularity. Results based on heterogeneity around family also replicate. Assortative matching on risk preference in the detected insurance groups replicates its pattern of (non-)significance even when the sum of risk aversion is controlled for.

	(1)	(2)	(3)	(4)	(5)
$\overline{ \eta_i - \eta_j }$	-0.141*	-0.167*	-0.113	-0.161*	-0.206*
	(0.0707)	(0.0744)	(0.0638)	(0.0687)	(0.0927)
$\eta_i + \eta_j$		-0.0302		-0.0567	-0.0567
		(0.0466)		(0.0489)	(0.0487)
$Family_{ij}$			3.030***	3.034***	2.939***
			(0.118)	(0.118)	(0.158)
Family _{<i>ij</i>} × $ \eta_i - \eta_j $					0.103
					(0.0940)
Village Dummies	Yes	Yes	Yes	Yes	Yes
Other Controls	No	No	No	No	No
N	71052	71052	71052	71052	71052

Table A8: Dyadic Logistic Regression: Risk Sharing Network

Dyadic robust standard errors are reported in parentheses (Fafchamps and Gubert, 2007). All specifications are dyadic linear probability models with matching in the risk sharing network as the dependent variable. η_i is risk aversion of individual i, so $|\eta_i - \eta_j|$ is the absolute difference of risk aversion while $\eta_i + \eta_j$ is the sum. Both absolute differences and sums of risk aversion are transformed into z-scores. * p < 0.05, ** p < 0.01, *** p < 0.001

	(1)	(2)	(3)	(4)	(5)
$\overline{ \eta_i - \eta_j }$	-0.0942	-0.0907	-0.0797	-0.0838	-0.105
·	(0.0701)	(0.0743)	(0.0681)	(0.0733)	(0.0820)
$\eta_i + \eta_j$		0.00404		-0.00474	-0.00479
-		(0.0545)		(0.0548)	(0.0548)
Family _{ij}			1.906***	1.906***	1.806***
			(0.112)	(0.112)	(0.145)
Family _{<i>ij</i>} × $ \eta_i - \eta_j $					0.107
					(0.0913)
Village Dummies	Yes	Yes	Yes	Yes	Yes
Other Controls	No	No	No	No	No
N	71052	71052	71052	71052	71052

Table A9: Dyadic Logistic Regression: Detected Insurance Groups

Dyadic robust standard errors are reported in parentheses (Fafchamps and Gubert, 2007). All specifications are dyadic linear probability models with matching in the risk sharing network as the dependent variable. η_i is risk aversion of individual i, so $|\eta_i - \eta_j|$ is the absolute difference of risk aversion while $\eta_i + \eta_j$ is the sum. Both absolute differences and sums of risk aversion are transformed into z-scores. * p < 0.05, ** p < 0.01, *** p < 0.001

A.6.3 Tetrad Logit Results

As a robustness check on the role of degree heterogeneity, I estimate tetrad logit in each village (Graham, 2017).³² I estimate the models in Python 3.7 using the netrics package.

Turning to the results, in Table A10 Panel A I focus on the estimates unconditional on the sum of risk aversion for each village. As this method tends to be unfamiliar, some interpretation of the reported output sizes may be useful. For each network, the procedure generates all tetrads of nodes. For example, for Darmang, we have 164 nodes, so it generates $\binom{164}{4} = 29,051,001$ tetrads. 159, 341 tetrads are selected by the kernel function (i.e., such that degree heterogeneity is balanced). (The fact that most of the tetrads are not used is not totally surprising as in most real world social networks, most tetrads will be empty.) I have included the number of nodes, and the number of tetrads used, as well as the fraction of tetrads used for each village.

Two of these village coefficients are negative and of similar magnitude to the logit coefficient, while one is very close to zero, and one is 2.5-3 times as large as the logit coefficient. While three of these estimates are themselves insignificant, this is largely due to the loss in power from splitting my sample into four parts. In fact, the simple average of the village coefficients without controls is very similar to the logistic coefficient when controlling for the sum of risk aversion. Additionally, as presented in Table A10 Panel B, the change in the estimated effect of the difference in risk aversion is not as pronounced in these estimates as it was in linear probability model or the logistic regression results presented earlier. The coefficients on the sum of risk aversion also fall in the tetrad logit specifications. This gives me confidence in the validity of my preferred specification as presented in the main text.

Interestingly, when this same back of the envelope calculation is done for the detected insurance group tetrad logit results, I also find similar results to the logistic regression results. These can be found, by village, in Table A11. As before, this represents an attenuation of assortative matching in detected insurance groups relative to risk sharing networks.³³

³²While it is theoretically possible to build an estimate from multiple villages by brute force, a back-of-theenvelope calculation indicates to me that I do not have the computing resources to do so as my disposal. This might be avoided with greater understanding of the function that indexes tetrads, using this same function and adjusting the inputs to feed in the dyadic and tetrad mappings within villages.

³³This fact may be useful for future empirical work on network formation. In particular, since community detection can construct communities of varying size, walktrap communities with short path lengths might in fact serve as useful in estimating assortative matching in practice.

	Outcome: Match Between i and j in Risk Sharing Network							
		V	ïllage		Simple Average			
	Darmang	Pokrom	Oboadaka	Konkonuru	(1)+(2)+(3)+(4)			
	(1)	(2)	(3)	(4)	=4			
Panel A: Uncondition	nal Estimat	es						
$\overline{ \eta_i - \eta_j }$	-0.432	-0.144	-0.088	0.098	-0.141			
· • • • •	(0.109)	(0.117)	(0.109)	(0.082)				
Panel B: Conditional	l Estimates							
$\overline{ \eta_i - \eta_j }$	-0.451	-0.216	-0.053	0.184	-0.134			
	(0.115)	(0.124)	(0.116)	(0.097)				
$\eta_i + \eta_j$	-0.097	-0.197	0.127	0.231	0.016			
	(0.101)	(0.141)	(0.140)	(0.114)				
Nodes	164	154	150	165				
${\cal N}$ Tetrads Used	159341	20463	38532	149422				
Fraction Tetrads Used	0.54%	0.09%	0.19%	0.50%				

Table A10: Tetrad Logit with Risk Sharing Network

Standard errors are presented in parentheses below logistic regression coefficients. Fraction of tetrads used denotes the fraction which are selected via the kernel function presented in Graham (2017), and is static across the two regressions.

	Outcor	Outcome: i and j co-members in Detected Insurance Group							
		V	'illage		Simple Average				
	Darmang	(1)+(2)+(3)+(4)							
	(1)	(2)	(3)	(4)	<u>=</u> <u> </u>				
Panel A: Uncondition	nal Estimat	es							
$\overline{ \eta_i - \eta_j }$	-0.139	-0.183	0.027	0.061	-0.059				
	(0.102)	(0.073)	(0.068)	(0.064)					
Panel B: Conditional	l Estimates								
$ \eta_i - \eta_j $	-0.154	-0.201	0.002	0.067	-0.063				
	(0.103)	(0.086)	(0.073)	(0.067)					
$\eta_i + \eta_j$	-0.115	-0.054	-0.099	0.034	-0.059				
	(0.116)	(0.101)	(0.092)	(0.086)					
Nodes	164	154	150	165					
Tetrads Used	279058	72769	99476	225251					
Fraction Tetrads Used	0.96%	0.33%	0.49%	0.76%					

Table A11: Tetrad Logit with Detected Insurance Groups

Standard errors are presented in parentheses below logistic regression coefficients. Fraction of tetrads used denotes the fraction which are selected via the kernel function presented in Graham (2017), and is static across the two regressions.

A.7 Subgraph Generation Models

A.7.1 Estimation

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

$$\tilde{\beta}_{I,\ell} = \frac{\sum_{i=1}^{n} \mathbf{1}(deg(i) = 0 | l_i = \ell)}{n_{\ell}}$$
(28)

$$\tilde{\beta}_{L,\ell,\ell} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{ij} \times \mathbf{1}(l_i = \ell) \times \mathbf{1}(l_j = \ell)}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbf{1}(l_i = \ell) \times \mathbf{1}(l_j = \ell)}$$
(29)

$$\tilde{\beta}_{L,l,r} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{ij} \times (\mathbf{1}(\ell_i = l) \times \mathbf{1}(\ell_j = r) + \mathbf{1}(\ell_i = t) \times \mathbf{1}(\ell_j = l))}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbf{1}(\ell_i = l) \times \mathbf{1}(\ell_j = r) + \mathbf{1}(\ell_i = t) \times \mathbf{1}(\ell_j = l)}.$$
(30)

For simplicity I index features with s. From proposition C.2 in Chandrasekhar and Jackson (2021) under a sparsity condition³⁴, $\Sigma^{-1/2}(\tilde{\beta}_n - \beta_0^n) \rightarrow N(0, I)$ where β_0^n is the true rate of subgraph generation. For a feature ℓ , the variance of the feature is the entry on the diagonal and the standard errors are the square root:

$$\Sigma_{s,s} = \frac{\beta_{0,s}^n (1 - \beta_{0,s}^n)}{\kappa_s \binom{n}{m_s}} \text{ and } \tilde{\sigma}_{s,s} = \sqrt{\frac{\tilde{\beta}_s^n (1 - \tilde{\beta}_s^n)}{\kappa_s \binom{n}{m_s}}}.$$
(31)

where m_s is the number of nodes involved in the feature and κ_s is the number of different possible relabelings of the feature (note: for both isolates and links $\kappa_s = 1$). For the results, $\kappa_s {n \choose m_s}$ is the sample size of the feature.

A.7.2 Pooled Subgraph Generation Models

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 28, 29, and 30 above. I estimate the coefficient associated with some subgraph s

$$\tilde{\beta}_s = \frac{\sum_{v=1}^4 \text{count}_{sv}}{\sum_{v=1}^4 \text{potential}_{sv}}.$$
(32)

³⁴First, my networks are sparse by the definition of Chandrasekhar and Jackson (2021). If I assume a constant growth rate of the density of links, then density is growing at about $n^{1/3}$ or less (which is acceptable). Second, 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, noting the sparsity condition may be cracking a walnut with a sledgehammer, so to speak.

This estimate uses only the relevant potential occurrences of the feature. Similarly, when estimating the standard errors of a feature, I cannot use the same effective sample size as I would use if I combined the networks. Let n_v be the number of nodes in the village network. If I take $\kappa_s\left(\sum_{m_s}^{n_v}\right)$, I would include many combinations of nodes that in reality could not form the subgraph in question. Hence I estimate the standard errors the of pooled SUGM

$$\tilde{\sigma}_{s,s} = \sqrt{\frac{\tilde{\beta}_s (1 - \tilde{\beta}_s)}{\kappa_s \times \sum_{v=1}^4 \binom{n_v}{m_s}}}.$$
(33)

A.7.3 Approximation of Variance of Ratios

I use an approximation of the variance of ratios.³⁵ We want the ratio of the variance of two coefficients $\tilde{\beta}_{L,s}$ and $\tilde{\beta}_{L,ra}$,

$$Var\left(\frac{\tilde{\beta}_{L,s}}{\tilde{\beta}_{L,ra}}\right) = \left(\frac{\tilde{\beta}_{L,s}}{\tilde{\beta}_{L,ra}}\right)^2 \left(\frac{(\sigma_s)^2}{(\tilde{\beta}_{L,s})^2} - \frac{2Cov(\tilde{\beta}_{L,s},\tilde{\beta}_{L,ra})}{\tilde{\beta}_{L,s}\tilde{\beta}_{L,s}} + \frac{\sigma_{ra}^2}{\tilde{\beta}_{L,ra}^2}\right)$$

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

$$Var\left(\frac{\tilde{\beta}_{L,s}}{\tilde{\beta}_{L,ra}}\right) = \left(\frac{\tilde{\beta}_{L,s}}{\tilde{\beta}_{L,ra}}\right)^2 \left(\frac{(\sigma_s)^2}{(\tilde{\beta}_{L,s})^2} + \frac{\sigma_{ra}^2}{\tilde{\beta}_{L,ra}^2}\right)$$

for the variance of the ratios.

³⁵See https://www.stat.cmu.edu/ hseltman/files/ratio.pdf.

A.7.4 Full Results of SUGM Estimation

Feature	Count	Potential	Sample Size	Coef.	Std. Err.
Isolates:					
Nuisance	51	180	633	0.2833	0.0179
Risk Averse	41	453	633	0.0905	0.0114
Within Links:					
Nuisance	76	3973	49852	0.0191	0.0006
Risk Averse	1030	25472	49852	0.0404	0.0009
Between Links:					
Risk Averse, Nuisance	502	20407	49852	0.0246	0.0007

Table A12: Baseline Pooled Subgraph Generation Model with Risk Sharing Network

Baseline Pooled SUGM using the risk sharing network with features including links and isolates by whether nodes are risk averse or are nuisances. Nuisance nodes are those who either have unmeasured risk aversion (i.e., were not surveyed) or who are risk loving, who I assume would not engage in risk sharing. Count is the number of subgraphs which actually display the feature, potential is the total number that could display the feature, and sample size is that used to estimate the standard errors.

Table A13: Baseline Pooled Subgraph Generation Model with Detected Insurance Groups

	Feature	Count	Potential	Sample Size	Coef.	Std. Err.
Isolates:						
	Nuisance	58	180	633	0.3222	0.0186
F	Risk Averse	60	453	633	0.1325	0.0135
Within Links:						
	Nuisance	177	3973	49852	0.0446	0.0009
F	Risk Averse	2365	25472	49852	0.0928	0.0013
Between Links:						
Risk Averse	e, Nuisance	1311	20407	49852	0.0642	0.0011

Baseline Pooled SUGM using detected insurance groups with features including links and isolates by whether nodes are risk averse or are nuisances. Nuisance nodes are those who either have unmeasured risk aversion (i.e., were not surveyed) or who are risk loving, who I assume would not engage in risk sharing. Count is the number of subgraphs which actually display the feature, potential is the total number that could display the feature, and sample size is that used to estimate the standard errors.

Feature	Count	Potential	Sample Size	Coef.	Std. Err.
Isolates:					
Less risk averse	22	236	633	0.0932	0.0116
More risk averse	19	217	633	0.0876	0.0112
Risk loving	13	82	633	0.1585	0.0145
Not surveyed	38	98	633	0.3878	0.0194
Within links:					
Less risk averse	421	7511	49852	0.0561	0.001
More risk averse	186	6223	49852	0.0299	0.0008
Risk loving	33	814	49852	0.0405	0.0009
Not surveyed	20	1181	49852	0.0169	0.0006
Between links:					
Less risk averse, more risk averse	423	11738	49852	0.036	0.0008
Less risk averse, risk loving	163	4765	49852	0.0342	0.0008
Less risk averse, not surveyed	132	5994	49852	0.022	0.0007
More risk averse, risk loving	141	4475	49852	0.0315	0.0008
More risk averse, not surveyed	66	5173	49852	0.0128	0.0005
Risk loving, not surveyed	23	1978	49852	0.0116	0.0005

Table A14: Preferences Pooled Subgraph Generation Model with Risk Sharing Network

Preferences Pooled SUGM using the risk sharing network with features including links and isolates by whether nodes are less risk averse, more risk averse, are risk loving, or have unmeasured risk aversion (were not surveyed). Count is the number of subgraphs which actually display the feature, potential is the total number that could display the feature, and sample size is that used to estimate the standard errors.

Feature	Count	Potential	Sample Size	Coef.	Std. Err.
Isolates:					
Less risk averse	38	236	633	0.161	0.0146
More risk averse	22	217	633	0.1014	0.012
Risk loving	17	82	633	0.2073	0.0161
Not surveyed	41	98	633	0.4184	0.0196
Within links:					
Less risk averse	893	7511	49852	0.1189	0.0014
More risk averse	444	6223	49852	0.0713	0.0012
Risk loving	59	814	49852	0.0725	0.0012
Not surveyed	42	1181	49852	0.0356	0.0008
Between links:					
Less risk averse, more risk averse	1028	11738	49852	0.0876	0.0013
Less risk averse, risk loving	373	4765	49852	0.0783	0.0012
Less risk averse, not surveyed	379	5994	49852	0.0632	0.0011
More risk averse, risk loving	311	4475	49852	0.0695	0.0011
More risk averse, not surveyed	248	5173	49852	0.0479	0.001
Risk loving, not surveyed	76	1978	49852	0.0384	0.0009

Table A15: Preferences Pooled Subgraph Generation Model with Detected Insurance Groups

Preferences Pooled SUGM using the detected insurance groups with features including links and isolates by whether nodes are less risk averse, more risk averse, are risk loving, or have unmeasured risk aversion (were not surveyed). Count is the number of subgraphs which actually display the feature, potential is the total number that could display the feature, and sample size is that used to estimate the standard errors.

B Theoretical Model: Covariate Risk Sharing in Groups

B.1 Focus on Risk Averse Individuals

As presented in Section 2, about 20% of the individuals in the sample are measured as having risk loving preferences. Close readers might reason that these individuals would take on covariate risk from others and would also appreciate the premium to do so. However, I do not model them as so, instead focusing on only risk averse individuals.

I make this decision for two reasons. First, I find it likely that risk loving individuals would be less inclined and less welcomed in risk sharing arrangements. In this model idiosyncratic and covariate risk sharing are a "package deal." That is, to be in the covariate risk sharing arrangement, one must also be in the idiosyncratic risk sharing arrangement. This would serve as a disincentive for these risk loving individuals. Depending on the ratio of idiosyncratic and covariate risk, this might be enough to dissuade these individuals from participating. If risk sharing networks are pairwise stable (Jackson and Wolinsky, 1996), this may also serve as a disincentive for risk averse individuals to accept risk loving respondents as risk sharing partners. While the formal model abstracts away from heterogeneity in income variance and downside risk, risk loving individuals' preferences would suggest they would take on more risk (or different kinds of risk, e.g., downside risk). Indeed, we see that risk loving individuals tend to have high income risk (though similar in variance to others who are less risk averse), have greater average losses from risk, and a greater ratio of net losses to net gains (see Table 1). While we do not model heterogeneity in income variance, others do. They find that high risk individuals are included only by those with similar risk profiles (e.g., Jaramillo et al., 2015; Gao and Moon, 2016; Xing, 2020). Indeed, risk loving individuals are more likely to be isolates in the risk sharing networks (see Table 1).

Second, more practically, it is somewhat difficult to make sense of respondents whose average choices in the hypothetical gamble are consistent with being risk loving. Their average choice hides some nuance, that respondents are not always consistent across domains. Notably, a number of respondents give answers that tend to accord with the *s*-shaped value function of prospect theory, where individuals are risk averse in gains and risk loving in losses (Kahneman and Tversky, 1979). Specifically, only 13.9% of respondents are risk loving in gains (considering those who answered both questions in this domain) while 35.8% are risk loving in losses. It seems unlikely that a respondent with prospect theory preferences would take on risk for free. A small number of others are risk loving in gains and risk averse in losses. The full distribution of risk preferences by domain can be found in Figure B1. While less intuitive considering the social science literature, it is also not so clear that net risk loving individuals who are risk averse in losses would take on risk for free. Therefore, this choice also saves me from forcing risk loving behavior on individuals who are risk averse in some domains.



Figure B1: A visualization of CARA risk preferences by gamble domain. The upper right and lower left quadrants accord (loosely) with expected utility theory. The lower right quadrant accords with prospect theory. The upper right hand quadrant has no associated behavioral theory, and are therefore coined as 'unicorns.' The diagonal line indicates risk neutrality on average.

B.2 Expected Utility

Because shocks are normally distributed, expected utility for both types is equivalent to maximizing the mean-variance representation as seen in Sargent (1987).

$$E(U_{\ell}(c_{\ell i})) = E(c_{\ell i}) - \frac{\eta_{\ell i}}{2} Var(c_{\ell i})$$

Also note CARA utility function increases in consumption. Thus, the agent consumes all income and transfers available in all states of the world. Expected consumption for type 1 is $E(c_{1i}) = \lambda_{1i}$ and for type 2, $E(c_{2i}) = \lambda_{2i}$. Variance for the two types can be computed:

$$Var(c_{1i}) = \left(\frac{\theta}{p}\right)^2 \left(\frac{\sigma^2}{N} + \nu^2\right) \text{ and } Var(c_{2i}) = \left(\frac{1-\theta}{1-p}\right)^2 \left(\frac{\sigma^2}{N} + \nu^2\right).$$

So then I write expected utility

$$E(U_{\ell}(c_{1i})) = \lambda_{1i} - \frac{\eta_{1i}}{2} \left(\frac{\theta}{p}\right)^2 \left(\frac{\sigma^2}{N} + \nu^2\right) \text{ and } E(U_{\ell}(c_{2i})) = \lambda_{2i} - \frac{\eta_{2i}}{2} \left(\frac{1-\theta}{1-p}\right)^2 \left(\frac{\sigma^2}{N} + \nu^2\right).$$

For ease of notation, I define $\sigma_c^2 = \frac{\sigma^2}{N} + \nu^2$ and note that the utility of the more risk averse agents when only idiosyncratic risk is pooled is equal to $EU_0 = -\frac{\eta_{2i}}{2}\sigma_c^2$.

B.3 Feasibility of Risk Sharing

Due to constraints 11, 12 and 13, budget constraints bind at the group level. To see this, I sum up the two types using weights:

$$pc_{1i} + (1-p)c_{2i} \le \theta \left(\frac{1}{N}\sum_{i=1}^{N}\tilde{y}_{i} + \tilde{y}_{v}\right) + p\lambda_{1} + (1-\theta)\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{y}_{i} + \tilde{y}_{v}\right) + (1-p)\lambda_{2}$$
$$N_{1}c_{1i} + N_{2}c_{2i} \le \sum_{i=1}^{N}\tilde{y}_{i} + N\tilde{y}_{v}.$$

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

B.4 Solving the Lagrangian

I construct the Lagrangian retaining constraints 10 and 13 (with a_2 and a_3 as multipliers, respectively) and incorporate the consumption constraints into expected utility.

$$\mathcal{L} = \lambda_1 - \frac{\eta_1}{2} \frac{\theta^2}{p^2} \sigma_c^2 + a \left(\lambda_2 - \frac{\eta_2}{2} \frac{(1-\theta)^2}{(1-p)^2} \sigma_c^2 + \frac{\eta_2}{2} \sigma_c^2 \right) + b \left(p \lambda_1 + (1-p) \lambda_2 \right)$$

The first order conditions are as follows:

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = 1 + bp = 0 \tag{34}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = a + b(1-p) = 0 \tag{35}$$

$$\frac{\partial \mathcal{L}}{\partial \theta} = \frac{-\eta_1 \theta \sigma_c^2}{p^2} + a_2 \left(\frac{\eta_2 (1-\theta) \sigma_c^2}{(1-p)^2} \right)$$
(36)

$$\frac{\partial \mathcal{L}}{\partial a} = \lambda_2 - \frac{\eta_2}{2} \left(\frac{1-\theta}{1-p}\right)^2 \sigma_c^2 + \frac{\eta_2}{2} \sigma_c^2 = 0$$
(37)

$$\frac{\partial \mathcal{L}}{\partial b} = p\lambda_1 + (1-p)\lambda_2 = 0 \tag{38}$$

Using FOC 34 I note that $b = -\frac{1}{p}$. Likewise, using FOC 35 I note that $a = \frac{1-p}{p}$. Rearranging FOC 37, FOC 38, and substituting :

$$\lambda_2 = -\frac{\eta_2}{2} \left(1 - \left(\frac{1-\theta}{1-p}\right)^2 \right) \Rightarrow \lambda_1 = -\left(\frac{1-p}{p}\right) \lambda_2 = \left(\frac{1-p}{p}\right) \frac{\eta_2}{2} \left(1 - \left(\frac{1-\theta}{1-p}\right)^2 \right)$$

Finally, I simplify FOC 36 to find θ :

$$\frac{\eta_1 \theta \sigma_c^2}{p^2} = \frac{1-p}{p} \left(\frac{\eta_1 (1-\theta) \sigma_c^2}{(1-p)^2} \right) \Rightarrow \left(\frac{\eta_1}{\eta_2} \right) \left(\frac{1-p}{p} \right) = \frac{1-\theta}{\theta}$$
$$\Rightarrow \frac{1}{\theta} = \left(\frac{\eta_1}{\eta_2} \right) \left(\frac{1-p}{p} \right) + 1 \Rightarrow \theta = \frac{p\eta_2}{(1-p)\eta_1 + p\eta_2}.$$

Covariate risk will not be taken on fully by the less risk averse agents. $\theta = 1$ only if either $\eta_1 = 0$ (type 1 is risk neutral, which we've assumed is not true) or p = 1. Note

$$(1-\theta)^2 = \left(1 - \frac{p\eta_2}{(1-p)\eta_1 + p\eta_2}\right)^2 = \left(1 - \frac{(1-p)\eta_1}{(1-p)\eta_1 + p\eta_2}\right)^2 = \frac{(1-p)^2\eta_1^2}{((1-p)\eta_1 + p\eta_2)^2}.$$

So then we can express the payment between type 1 and type 2 agents:

$$\lambda_2 = -\frac{\eta_2}{2} \left(1 - \frac{\eta_1^2}{((1-p)\eta_1 + p\eta_2)^2} \right).$$

B.5 The Rate of Risk Pooling

One result of the theoretical model is that the proportion of risk taken on by less risk averse individuals in a group in equilibrium is greater than their proportion of the group. To see this, note that since $\eta_1 < \eta_2$ by assumption $p\eta_2 + (1-p)\eta_1 < p\eta_2 + (1-p)\eta_2 = \eta_2$. Thus,

$$\theta^*(p,\eta_1,\eta_2) = \frac{p\eta_2}{p\eta_2 + (1-p)\eta_1} > \frac{p\eta_2}{\eta_2} = p.$$

B.5.1 Value Functions

These solutions lead to the value functions:

$$V_{1}(p,\eta_{1},\eta_{2}) = \frac{\eta_{2}}{2} \left(\frac{1-p}{p}\right) \left(1 - \left(\frac{\eta_{1}}{(1-p)\eta_{1}+p\eta_{2}}\right)^{2}\right)$$

$$-\frac{\eta_{1}}{2} \left(\frac{\eta_{2}}{(1-p)\eta_{1}+p\eta_{2}}\right)^{2} \left(\frac{\sigma^{2}}{n}+\nu^{2}\right)$$

$$V_{2}(p,\eta_{1},\eta_{2}) = -\frac{\eta_{2}}{2} \left(1 + \left(\frac{\eta_{1}}{(1-p)\eta_{1}+p\eta_{2}}\right)^{2} \left(\left(\frac{\sigma^{2}}{n}+\nu^{2}\right)-1\right)\right).$$
(39)
(39)

I compute the value functions for type 1 and type 2 individuals.

$$\begin{split} V_1(p,\eta_1,\eta_2) &= E(U_1(c_{1i})|\theta^*(p),\lambda_1^*(p)) = \lambda_1^*(p) - \frac{\eta_1}{2} \left(\frac{\theta^*(p)}{p}\right)^2 \sigma_c^2 \\ &= \lambda_1^*(p) - \frac{\eta_1}{2} \left(\frac{p\eta_2}{((1-p)\eta_1 + p\eta_2)p}\right)^2 \sigma_c^2 = \lambda_1^* - \frac{\eta_1}{2} \left(\frac{\eta_2}{((1-p)\eta_1 + p\eta_2)}\right) \sigma_c^2 \\ V_1(p,\eta_1,\eta_2) &= \frac{\eta_2}{2} \left(\frac{1-p}{p}\right) \left(1 - \left(\frac{\eta_1}{(1-p)\eta_1 + p\eta_2}\right)^2\right) - \frac{\eta_1}{2} \left(\frac{\eta_2}{(1-p)\eta_1 + p\eta_2}\right)^2 \sigma_c^2 \\ V_2(p,\eta_1,\eta_2) &= E(U_2(c_{2i})|\theta^*(p),\lambda_2^*(p)) \\ &= \lambda_2^*(p) - \frac{\eta_2}{2} \left(\frac{1-\theta^*(p)}{1-p}\right)^2 = \lambda_2^*(p) - \frac{\eta_2}{2} \left(\frac{1-\frac{p\eta_2}{(1-p)\eta_1 + p\eta_2}}{1-p}\right)^2 \sigma_c^2 \\ &= \lambda_2^*(p) - \frac{\eta_2}{2} \left(\frac{(1-p)\eta_1 + p\eta_2 - p\eta_2}{(1-p)((1-p)\eta_1 + p\eta_2)}\right)^2 \sigma_c^2 \\ &= \lambda_2^*(p) - \frac{\eta_2}{2} \left(\frac{(1-p)\eta_1 + p\eta_2 - p\eta_2}{(1-p)((1-p)\eta_1 + p\eta_2)}\right)^2 \sigma_c^2 \\ &= \lambda_2^*(p) - \frac{\eta_2}{2} \left(\frac{\eta_1}{(1-p)((1-p)\eta_1 + p\eta_2)^2}\right) - \frac{\eta_2}{2} \left(\frac{\eta_1}{(1-p)\eta_1 + p\eta_2}\right)^2 \sigma_c^2 \\ &= -\frac{\eta_2}{2} \left(1 - \frac{\eta_1^2}{((1-p)\eta_1 + p\eta_2)^2}\right) - \frac{\eta_2}{2} \left(\frac{\eta_1}{(1-p)\eta_1 + p\eta_2}\right)^2 \sigma_c^2 \\ &= -\frac{\eta_2}{2} \left(1 - \frac{\eta_1^2}{((1-p)\eta_1 + p\eta_2)^2}\right) - \frac{\eta_2}{2} \left(\frac{\eta_1}{(1-p)\eta_1 + p\eta_2}\right)^2 \sigma_c^2 \end{split}$$

B.6 Optimal Assignment and Village Composition

Optimal composition of groups occurs when the proportion of individuals within the group is equal to that in the village. As a demonstration is not an artifact of equal sized groups, I vary the composition of types in the population in Figure B2. In this figure, welfare is maximized when $p_{A1} = p_1$, the proportion of type 1 agents in the population.

In addition, it is interesting to understand what proportion of covariate risk is shared in each group as a planner sorts types into two groups. Figure B3 demonstrates how the proportion of risk sharing in larger and smaller groups varies by composition. As type 1 individuals move from the larger group to the smaller group, a greater proportion of covariate risk, encapsulated by θ is taken on by these individuals within the smaller group. This results in a risk management frontier which is bowed out. When more risk neutral agents are all in the larger or smaller group, they come close to taking on all of the covariate risk.



Figure B2: Optimal Allocation of Types Between Unequally Size Groups with varying numbers of type 1 and type 2 agents.



Figure B3: A Risk Management Frontier: Proportion of Covariate Risk Taken on by Less Risk Averse Agents in Groups. From top left to bottom right, type 1 agents move from the larger group to the smaller one.

B.7 The Welfare Implications of Risk Preferences

I measure risk aversion using hypothetical gambles. Though these gambles return those who are more and less risk averse, it is likely that the relatively low stakes of the hypothetical gamble may yield coefficients of risk aversion much lower than we might observe with a high stakes incentivized gamble. Moreover, if risk aversion is underestimated, then the welfare impact of risk sharing will also be underestimated. Even within the local range of risk aversion measured, we can see non-trivial differences in losses due to risk. For example, Figure B4 shows how losses due to observed assortative matching increase with risk aversion of more risk averse agents. Furthermore, see Figure A2 which shows a mass of top-coded coefficients of risk aversion.



Figure B4: Greater risk aversion increases the welfare impact of assortative matching. As risk aversion increases villages with greater assortative matching will suffer more than those without. However, the delta between degrees of assortative matching is subject to diminishing marginal losses. The dashed vertical line indicates the measured degree of risk aversion among type 2 agents.
C Simulation Methods

C.1 Simulation Algorithm

Before simulating, I remove all individuals who do not have preference data or who are not risk averse, and discard resulting groups with only one member.

- 1. Sort groups into two bins with roughly equal total populations. The first bin will be majority type 1 and the second will be majority type 2. To assign groups, first I sort the groups into a random order. I designate a bin of type 1 majority and one for type 2 majority, and then I construct a running membership sum for each bin. I add a group to bin 1 when $sum_1 \leq sum_2$ and to bin 2 otherwise and proceed until all groups have been added.³⁶
- 2. Assign nodes of differing types to groups using a binomial process, varying the probability of success in that process according to what is implied by that scenario (i.e., p^U). A success assigns a majority type node to that group while a failure assigns a minority type node.
- 3. Compute the value functions for type 1 and type 2 agents in each group according to the formulas found in Appendix B.5.1 and average across individuals to determine the per capita losses due to covariate risk. These are reported in units of Purchasing Power Parity (PPP).

Each of these steps is repeated for each repetition of the simulation.

C.2 Rate of Between Link Generation

How many connections are there between types in groups? The complete bipartite graphs yields simple counts. A complete bipartite 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 groups is $\sum_{g=1}^{G} N_{1g}N_{2g}$. Additionally, the total number of potential links between types in the entire village graph will be

$$\left(\sum_{g=1}^{G} N_{1g}\right) \left(\sum_{g=1}^{G} N_{2g}\right) = N_1 N_2. \implies \tilde{\beta}_{1,2} = \frac{\sum_{g=1}^{G} N_{1g} N_{2g}}{N_1 N_2}.$$

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

$$\tilde{\beta}_{1,2} = \frac{\sum_{g=1}^{G} N_{1g} N_{2g}}{\frac{N^2}{2^2}} = \frac{4 \times \sum_{g=1}^{G} N_{1g} N_{2g}}{N^2}$$

 $^{^{36}}$ Directly minimizing the difference in total membership in type 1 and type 2 majority groups is an np-hard problem. This approach serves as a workaround.

$$\tilde{\beta}_{1,2} = 4 \times \sum_{g=1}^{G} \frac{N_{1g}}{N} \frac{N_{2g}}{N} = 4 \times \sum_{g=1}^{G} \frac{N_g p_{1g}}{N} \frac{N_g p_{2g}}{N} = 4 \times \sum_{g=1}^{G} \left(\frac{N_g}{N}\right)^2 p_{1g} p_{2g}$$

For the last equality, recall that $p_{\ell g} = \frac{N_{\ell g}}{N_g}$. I make the (heroic) simplifying assumption that group sizes are the same, hence there's a fixed $\frac{N_g}{N} = \frac{1}{G}$. Additionally, I fix $p_{1g} = \bar{p}^U$ and $p_{2g} = \bar{p}^L$ when $p_{1g} \ge p_{2g}$ and vice-versa when $p_{1g} < p_{2g}$, where $\bar{p}^U = 1 - \bar{p}^L$.

$$\tilde{\beta}_{1,2} = \frac{4}{G^2} \times \sum_{g=1}^{G} p_{1g} p_{2g} = \frac{4}{G^2} \times \sum_{g=1}^{G} \bar{p}^U \bar{p}^L$$

Finally, I sum across groups and then rearrange to get the expression for $\tilde{\beta}_{1,2} = \frac{4}{G} \bar{p}^U \bar{p}^L$.

C.3 Rate of Within Risk Averse Link Generation

The total number of potential links generated is $\frac{N(N-1)}{2}$. With completely connected groups, the number of connections ends up being $\frac{\sum_{g=1}^{G} N_g(N_g-1)}{2}$. Suppose also, as above, that $N_g = \frac{N}{G}$. Then,

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

C.4 Ratio of Rates

The ratio of rates is

$$\frac{\tilde{\beta}_{1,2}}{\tilde{\beta}_L} = \frac{\left(\sum_{g=1}^G N_{1g} N_{2g}\right) / N_1 N_2}{\left(\frac{\sum_{g=1}^G N_g (N_g - 1)}{2}\right) / \left(\frac{N(N-1)}{2}\right)}.$$

Based on the simplifications above, however, I can express the ratio of the link generation coefficients as an expression relating the proportion of types in each group to the rate of generation.

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

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

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

where $p^L \leq 0.5 \leq p^U$.