CONSUMPTION INSURANCE IN NETWORKS WITH ASYMMETRIC INFORMATION

Orazio Attanasio
Sonya Krutikova

Consumption Insurance in Networks with Asymmetric Information
Orazio Attanasio and Sonya Krutikova
NBER Working Paper No. 27290
May 2020
JEL No. O10,O15

ABSTRACT

This paper uses a dataset from Tanzania that contains information on consumption, income and income shocks within and across family networks. A unique feature of this data is that it contains data on the degree of information existing between each pair of households within family networks. We use these data to construct a novel measure of the quality of information both at the level of household pairs and at the level of the network. We study risk sharing within these networks and explore whether the rejection of perfect risk sharing that we observe can be related to imperfections in the information network members have about each. We show that households within family networks with better information are less vulnerable to idiosyncratic shocks. Next, using the same information, we characterise the position of households within the family net- works constructing measures of network centrality. We show that more central households within networks are less vulnerable to idiosyncratic shocks.

Orazio Attanasio
Department of Economics
Yale University
37 Hillhouse Avenue
New Haven, CT 06511
and Institute for Fiscal Studies, FAIR, BREAD and CEPR
and also NBER
orazio.attanasio@yale.edu

Sonya Krutikova
Institute for Fiscal Studies
7 Ridgmount Street
London
UK
sonya_k@ifs.org.uk

1 Introduction

In many developing countries, risk is very pervasive. The lack of economic development implies not only that individuals have access to a much lower level of resources but also that life is much riskier. Substantial shocks to resources, when one's living standards are close to subsistence levels, can have important and dramatic consequences. The presence of risk is particularly salient in rural contexts, where individuals are exposed to large and frequent shocks to their livelihoods. Many of these shocks are idiosyncratic, implying that insurance can be very valuable and very important. Formal insurance markets, however, rarely exist. Informal insurance arrangements are common and, according to the available evidence, do provide some level of insurance(see, for instance Townsend (1994) and, more recently, Kinnan and Townsend (2012)). The observed allocations, however, are very different from those that would prevail under perfect risk sharing and complete markets.

Informal insurance in village economies has received considerable attention, both because it can enable individuals to smooth out idiosyncratic shocks and because its presence interacts with (and sometimes limits) formal insurance and its development (see, for instance, Attanasio and Rios-Rull (2000) andMunshi and Rosenzweig (2016)).The presence of perfect risk-sharing arrangements has been rejected in many contexts (see, for instance, Rosenzweig (1988), Udry (1994) and the recent paper by Kinnan (2020) and the reference therein). An important and interesting research question is, therefore, the identification of the imperfections and frictions that prevent perfect risk sharing. A better understanding of these imperfections and identification of specific frictions as being particularly salient is important not only from a research point of view, but also for informing policy reforms and different policy interventions.

The literature has looked at different types of frictions and imperfections, including the lack of enforceability of informal insurance contracts and imperfect information. Studies that focus on the lack of contract enforceability include Foster and Rosenzweig (2001), Kocherlakota (1996), Ligon et al. (2002), Attanasio and Rios-Rull (2000), Albarran and Attanasio (2003), and, more recently, Ambrus et al. (2014) and Bold and Broer (2018). Other studies have considered information frictions, including Udry (1994), Ligon (1998), Cole and Kocherlakota (2001) and Attanasio and Pavoni (2011).[1] Information frictions could prevent full risk sharing because of the difficulty in contracting on specific income shocks and/or to moral hazard behaviour.

---

[1] A few papers have considered models with both asymmetric information and limited commitment: Atkinson (1991), and more recently, Broer et al. (2017), and Denderski and Stoltenberg (2019).

Informal insurance arrangements are based on individual transactions and transfers. Data on these transfers are difficult to collect and in most of the available data sets, the information on them is very limited and imprecise. The apporach proposed by Townsend (1994) to test for perfect risk sharing is particularly attractive because it only requires information about consumption allocations (the object that is being insured) and income or income shocks. While the approch is silent about the specific mechanisms that are used in practice to implement the observed resource allocations, it can be useful to characterise the extent to which such allocations differ from perfect insurance. When such a benchmark model is rejected one can model these deviations from the perfect risk sharing model and gain information on the nature of the imperfections and frictions that generate the available data.

In this paper, we use a unique data set which contains longitudinal data, following family networks in Tanzania for nearly 20 years.One of the exceptional features of this data-set is that within each family network, all households are asked about their own wealth (including about ownership of a variety of different assets) and about the wealth of all other households in their family network. These rich data provide a unique opportunity to construct measures of the quality of the information within each family.

The paper makes two original contributions. First, we propose a new[2] methodology to convert the data on the wealth of each household as perceived by all the other households in the family network into a measure about the quality of information in the family network. As we discuss in section 4, we construct both measures about the quality of information about each household in the network and about the overall quality of information in a network. As for the former we obtain measures both of the quality of the information that the rest of the network has about the wealth of every household and of the quality of information that each household has about the wealth of the rest of the network. These concepts are related, as we discuss, to measures of network centrality.

Second, we use the measures of the quality of information to check whether they are related to deviations from perfect risk sharing. Starting from a regression similar to those estimated by Townsend (1994), which relates changes in consumption to idiosycratic shocks after controlling for network level shocks, we check whether this relationship is affected by the quality of information.

Information frictions are the focus of this paper. In particular, we will relate mea-

---

[2]De Weerdt et al. (2019) use the same data to obtain measures of information imperfections. Our approach, however, is substantially different.

sures of risk sharing in family networks to measures of the quality of information in these networks. The type of informational frictions we consider most closely resemble the local information constraints in Ambrus et al. (2019). Although we do not map explicitly the imperfections they consider on specific transfers and consumption data, our paper is probably the first to consider an empirical measure of imperfect information and relate it to risk sharing in a way that is consistent with the model considered by Ambrus et al. (2019).

We find that our measures of the quality of information exhibit a substantial amount of variation across the network and, when considering dyads of households, the measures covaries with a number of observables (such as the geodesic distance or the frequency of contact between two households in the famliy network) in the expected fashion. Given the novelty of the type of measures we use, this result is important and reassuring.

As for the level of risk sharing, we first document that, as in other contexts, perfect risk sharing within a network is rejected by our data. When we then look at the way in which information quality interacts with risk sharing, by and large, we find that the better the quality of information in a family network (whether measured at the network or at the individual household level), the closer the allocations in that network are to those that would occur under perfect risk sharing. The results are more precise and more convincing when we consider the quality of information at the household rather than family network level and, in particular, when we consider the quality of the information that the network has about a household's wealth. We find that households whose wealth seems to be better known to the other members of households in the network are less sensitive to idiosyncratic shocks.

To the best of our knowledge, this is the first result relating so directly risk sharing to the quality of information. [3] We do not map directly the association we find between the sensitivity of individual consumption changes to idiosyncratic shocks to a theoretical model of constrained efficient allocations. This evidence, however, represents an important step in that direction.

The rest of the paper is organised as follows. In Section 2, we present a conceptual framework to analyse risk sharing and relate it to information frictions. In Section

---

[3]As mentioned above, De Weerdt et al. (2019) use these same data on information within family networks, but apply a different methodology for constructing a measure of misperceptions in order to study a different set of questions - testing alternative motives for transfers within family networks. Indeed, as their measure of information imperfections is constructed using consumption data, it cannot be related to risk sharing.

3, we present our data set and in Section 4, we describe our measures of information quality and report some descriptive statistics on them. We report the results of our empirical analysis in section 5. Section 6 concludes the paper.

2 Risk Sharing with Imperfect Information: a conceptual framework

The main purpose of this paper is to analyse deviations from perfect risk sharing and relate them to the severity of information problems. In this section, we provide a conceptual framework to inform the empirical analysis that we perform. We take the risk sharing group as a given and characterise risk sharing within that group. Our definition of risk sharing within a pre-defined group does not preclude the existence of other risk sharing arrangements or alternative mechanisms to absorb individual (or group level) shocks. However, the focus is on risk sharing within the pre-defined group.

The groups we consider in the empirical exercise are family networks. The members of an original family in 1991 were followed until 2010 as they move out of the initial nucleus. The survey team was able to achieve a remarkably high response rate. It is, therefore, plausible to assume that membership of a 'risk sharing group' as we define it, can be considered as a given.

We consider the ability of group members to absorb indiosyncratic shocks, given the group level shocks. That is, we consider the ability to share risk within the group considered and relate that ability to the quality of information within that group and, additionally, to some of the properties of the network in question considered. While it is likely (as we show) that the quality of information within a group is partly a function of choices made taken by individual members, this consideration does not affect the nature of our exercise. We take the quality of information (or the nature of the network) within a risk sharing group as a given at a point in time and document the extent to which such a variable affects the ability of individuals to smooth out income shocks. Our exercise is silent, in the a first instance, about the mechanisms that lead to a specific relationship between information quality and risk sharing.

2.1 Perfect risk sharing

We start by considering risk sharing within a group of agents. Group membership is given exogenously and therefore risk sharing groups are defined exogenously. They could be villages, family networks or other groups.

4

The approach we take to characterise full risk sharing within a group goes back to Wilson (1968) and Townsend (1994). We assume that individual j belonging to group g receives a stochastic endowment $y_t^{j,g}$.

$$y_t^{j,g} = \bar{y}^g_{\ t} + \epsilon_t^j.$$

We assume that y is perishable. For expositional simplicity we do not consider saving here, but it would be straightforward to add that or any other mechanism to absorb shocks, including risk sharing arrangements with different groups and/or institutions.

Individual j in group g receives utility from consumption $c_t^{j,g}$, which is given by their endowment plus a transfer $\tau_t^{j,g}$. The implications for allocations of perfect risk sharing can be derived considering a social planner problem, as in Townsend (1994), where the planner maximises a weighted sum of of the expected utility of the risk sharing group.

$$\max_{\{\tau_t^{i,g}\}^{i=1,..K_g}} \sum_{j=1}^{K_g} \lambda_{j,g} \sum_{t=0}^{\infty} \beta^t \int_Y u(y_t^{j,g} + \tau_t^{j,g}) d\mu^t(y_t^{j,g}) \qquad s.t. \qquad (1)$$

$$\sum_{j=1}^{K_g} y_t^{j,g} = \sum_{j=1}^{K_g} c_t^{j,g} \qquad\qquad \forall t \qquad (2)$$

$$c_t^{j,g} = y_t^{j,g} + \tau_t^{j,g} \qquad\qquad \forall t, j \qquad (3)$$

where $\mu^t()$ is a probability measure of the stochastic endowment $y_t^{j,g}$, which reflects the available (and public) information. $y_t^{j,g}$ is completely observable *(ex-post)* and can be contracted upon. $\lambda_{j,g}$ is the Pareto weight given to individual j in group g, which allows for inequality and asymmetries within the group and is assumed to be constant. Different λ's might reflect different status within the risk sharing group, or access to different amount of resources by different individuals. Again, by modifying the aggregate resource constraints (2) and/or the individual budget constraint (3), it is possible to take into account additional insurance mechanisms either through saving or through risk sharing arrangements *outside the group.*

Considering the Lagrangian for this problem, in the absence of frictions (information, enforceability), the first order condition w.r.t. c_t^i is given by:

$$\lambda_{i,g} \beta^t u'(c_t^{i,g}) \mu_t^g = \nu_t^g \quad \forall i, t. \qquad (4)$$

where μ_t^g is the probability of the state of the world at time t and ν_t^g is the multiplier associated with group g aggregate resource constraint. From equation (4) we note that

the right hand side, ν_t^g, does not depend on the individual household index i. Second, the Pareto weight $\lambda_{i,g}$ is a constant that does not depend on t. Finally, the f.o.c. is not averaged across different states of the world, but holds, appropriately weighted by $\mu_t^g()$, the probability distribution of the income shocks, for every possible state. This last fact is key to the perfect risk sharing structure. As income shocks are fully observable and contractable upon, insurance contracts can diversity idiosyncratic income shocks, regardless of the properties of the stochastic process that generates them. Considering the ratio of these conditions for agents i and 1 we get :

$$\frac{u'(c_t^{i,g})}{u'(c_t^{1,g})} = \frac{\lambda_{1,g}}{\lambda_{i,g}}; \quad \forall i, t \tag{5}$$

Taking logs of equation (4) one gets :

$$ln(\lambda_{i,g}) + ln(u'(c_t^{i,g})) + tln(\beta) = ln(\nu_t^g) \tag{6}$$

This equation can be taken to data in several ways. One can introduce and estimate individual fixed effects to capture the unobservable Pareto weights, or one can can take time differences to difference them out. Having many risk sharing groups, one has to consider timeXgroup fixed effects.

Taking differences across time periods:

$$\Delta_s ln(u'(c_t^{i,g})) + sln(\beta) = \Delta_s ln(\nu_t^g) \tag{7}$$

Equations (6) and (7) capture the essence of perfect risk sharing. We notice the absence from the right hand side of any time varying idiosyncratic variable, once one controls for a group aggregate. In equation (6), which is expressed in levels, one has to control for individual fixed effects that capture the Pereto weights of the planner problem. These effects drop out in the specification in first differences. These restrictions are the key implications of perfect risk sharing. To test them, Townsend (1994) and others have augmented it with idiosyncratic variables, such as levels or changes of individual income.

$$ln(u'(c_t^{i,g})) = \tilde{\kappa}^{i,g} + ln(\nu_t^g) + \tilde{\gamma}ln(y_t^{i,g}) + \tilde{\epsilon}_t^{i,g} \tag{8}$$

$$\Delta_s ln(u'(c_t^{i,g})) = \kappa^g + \Delta_s ln(\nu_t^g) + \gamma\Delta_s ln(y_t^{i,g}) + \epsilon_t^{i,g} \tag{9}$$

where $\tilde{\epsilon}_t^{i,g}$ and $\epsilon_t^{i,g}$ reflect measurement error and other unobservables. The coefficients $\tilde{\gamma}$ and γ, which measure the *vulnerability* of a single individual to idiosyncratic shocks, should be 0 under perfect risk sharing.

These equations are the essence of Townsend (1994) approach to testing perfect risk sharing. An attractive feature of this approach is that it is based exclusively on on consumption and idiosyncratic resource information. It does not require information about the decentralization mechanisms (such as transfers) that a given realisation of income shocks would require to achieve first best.

2.2 Deviations from perfect risk sharing: imperfect information

Empirically, perfect risk sharing is often rejected. Researchers have found that, while some smoothing of idiosyncratic shocks is observed, observed allocations are different from first best ones, in that consumption is affected by idiosyncratic shocks. While individuals in a variety of contexts can achieve some insurance, the empirical evidence seems to reject perfect risk sharing (see Rosenzweig (1988), Udry (1994) and Kinnan (2020)). De Weerdt and Hirvonen (2016) find similar evidence in the same data we use below, which were collected in Tanzania.

As discussed in the introduction, in light of the rejection of perfect risk sharing, a profitable approach is to look at the relevance of specific frictions. In this paper, we analyse a very simple and specific information problem, which is partly driven by the type of data we have. In particular, we relate the level of risk sharing (or deviations from perfect risk sharing) to the quality of information in a given network. We start by assuming that individual household i in risk sharing group g (a family network), receives an exogenous income $y^{i,g}$. Each member of the risk sharing group receives some signals about the income received by everybody else.

$$x_t^{ij,g} = y_t^{i,g} + e^{ij,g} \tag{10}$$

where $x_t^{ij,g}$ is the signal received by household j about household i's income, where both households belong to group g. $e^{ij,g}$ is a zero-mean random variable, which represents the noise that somewhat masks household i's income from household j. The variance of this noise represents the quality of information that household j has about household i. We will allow the precision of the signal to be different across different members of the network and, as we discuss below, construct estimates of the quality of the information each individual household in any given pair in a network has about the other. The quality of the information that different network members have about the

economic status of any given individual household effectively defines the position of that household in the network.

In our empirical application, we follow two different approaches, which we discuss in the next sub-section. First, we assume that the extent of risk sharing within a pre-defined group (in our case a family network) is determined by the quality of information available in a network. Family networks where the information is of high quality, will be closer to the first best allocations that would be observed under perfect risk sharing. Individuals living in households that are part of family networks where information flows are of inferior quality, on the other hand, will be more vulnerable to idiosyncratic shocks that could be diversified.

Second, we relate household vulnerability to the position of the household in a family network, which in turn is determined by the quality of information in the network. This approach can be interpreted in the light of a model of bilateral transfers among all the households in the networks. In such a model, discussed by Attanasio et al. (2020, in progress), the net tranfers between households i and j are determined by contracts based on the information common to these households, which is $x_t^{ij,g}$ and $x_t^{ji,g}$. An attractive feature of this approach is that one can avoid consideration of explicit incentive compatibility constraints induced by truth telling contraints or moral hazard problems. Instead, it uses a static framework where transfers between pairs of households can only be conditional on the information available to both members. We now turn to the discussion of these two approaches.

2.3 Empirical strategy to measure the effect of asymmetric information

One of the advantages of the perfect risk sharing model is that one can be agnostic about the specific decentralization of the efficient allocation. One, therefore, does not need to keep track of all bilateral transfers and can consider only net transfers, and therefore consumption allocations, of each individual. This is not necessarily the case in the asymmetric information case with several sources of asymmetries. It is possible, however, to implement tests that identify violation of perfect risk sharing by looking at the vulnerability of individual consumption to idiosyncratic shocks and relating such vulnerability to the quality of information in a given risk sharing group.

To capture deviations from perfect risk sharing due to imperfect information empirically we extend equation (9) in several ways. First, we estimate the following equation:[4]

---

[4]We also estimate similar versions of equation (8), which is in levels. These equations have a very

$$\Delta_s ln(u'(c_t^{j,g})) = \kappa + \mu_t^g + \phi_1 Iq^g + (\gamma_0 + \gamma_1 Iq^g)\Delta_s ln(y_t^{j,g}) + \epsilon_t^{j,g} \qquad (11)$$

where Iq^g is a measure of the quality of information in family network g and $\mu_t^g = \Delta_s ln(\nu_t^g)$. Since Iq^g varies only across networks, the parameter ϕ_1 cannot be identified as the variation in Iq^g would be absorbed by the group dummies μ_t^g. Under perfect risk sharing, as mentioned above, idiosyncratic shocks and changes to household income should not be related to changed in household consumption, after controlling for group level shocks, which is what the group-time dummies μ_t^g do.[5]

The coefficient on the income (or idiosynratic) shocks in equation (11) can be interpreted as measuring the degree of vulnerability of individual members to such shocks. In this equation, vulnerability can depend on the quality of information in a given network. If γ_1 has the opposite sign of γ_0, vulnerability in high quality information networks is lower. Allocations in such networks are closer to those that would prevail under perfect risk sharing.

The estimation of equation (11) requires a measure of the quality of information within a network. We discuss such a measure in Section 4 below. The main idea here is that, family networks where information is very good are close to achieving perfect risk sharing, while family networks where information is very imperfect, deviate considerably from perfect risk sharing, so that individual households are vulnerable to idiosyncratic shocks.

In equation (11), the quality of information varies across family networks but it is constant for the households *within a family network*. In the model presented by Ambrus et al. (2019) and developed by Attanasio et al. (2020, in progress), under a certain set of assumptions, vulnerability of an individual does not depend exclusively on the average quality of the information flows in that network, but could depend on the position of a given individual in the network and an appropriate weighted average of the shocks received in the network. To bring this intuition to the data, in addition to equation (11), we also estimate the following equation:

$$\Delta_s ln(u'(c_t^{j,g})) = \kappa^g + \mu_t^g + \phi_2 NP^{i,g} + (\gamma_0 + \gamma_2 NP^{i,g})\Delta_s ln(y_t^{j,g}) + \epsilon_t^{j,g} \qquad (12)$$

---

large number of parameters, as we need individual fixed effect and we have a small number of periods. The results are similar and are available upon request.

[5]In our data, where we have 2 periods, and therefore, 1 period over which changes in consumption are observed, we only have group dummies. It is important, however, to remember that if additional time periods were available we would have fully interacted time and group dummies, to reflect the resource constraints multipliers of each group.

where $NP^{i,g}$ is a measure of household centrality within the family network, reflecting either average quality of the information about household i's income held by the other members of the family network, or quality of information that household i has about incomes of the other households in the network. We discuss the definition and interpretation of these alternative variables in Section 4, where we discuss the construction of information quality measures. As before, in addition to interacting $NP^{i,g}$ with individual shocks, we also enter that variable on its own.

Implementing the estimation of equations (11) and (12) requires the availability of variables that can capture the quality of information and the position in the risk sharing network of an individual. We discuss these issues below. Before moving to the description of the data and to the methods we use to obtain these measures, it is worth mentioning a few additional issues. First, we notice that both the Iq^g and $NP^{i,g}$ variables do not have a time subscript as we assume they are constant over time. This assumption is forced on us in part by the fact that we have measurements on the quality of information only in the last wave of the survey. We can justify this assumption, however, by pointing to the fact that we only consider two time periods. We can therefore argue that our measure of information quality captures the household vulnerability to idiosyncratic shocks that occur between those two periods.

A more worrying possible criticism is that the quality of information and the structure of the networks does not change randomly, but might be linked to individual incentives and economic opportunities. As we will see below, much of the variation in household information quality is driven by distance between geographic locations, which is in turn driven by migration decisions. While migration decisions are very likely to be driven by economic opportunities, we stress that our characterization of risk sharing, based on measuring how household vulnerability is affected by the quality of information (or the position in the network) takes these variables as predetermined.

Considering, for instance, equation (11), we notice that the left-hand side represents changes in the log of the marginal utility for household i belonging to group g. The residual term of that equation, $\epsilon_t^{i,g}$ represents measurement error and other unobservable shifts to the marginal utility of consumption. One possible source of bias in the results we obtain would be a correlation between the unobserved determinants of log marginal utility across networks and differences in the quality of information. Reassuringly, however, our results do not change when we add variables, such as changes in family composition, that might capture part of the variability of $\epsilon_t^{i,g}$, suggesting that they are unlikely to be affected by this type of omitted variable bias. Of course, it could

be that idiosyncratic changes in the marginal utility of consumption could induce migration of some network members and, as a consequence, be correlated with the quality of information. We are not particularly concerned with this possiblity because we take the information structure in an extended family network (and its drivers) as given and predetermined at the start of the period over which we consider consumption changes.

3   The data from Tanzania

The Kagera Health and Development Survey (KHDS) in Tanzania is one of the longest running African panel surveys designed to study long-run and inter-generational trends in and mechanisms of poverty persistence and economic growth in rural households. Kagera region lies at the shores of Lake Victoria and shares a border with Uganda, Rwanda and Burundi. The 2012 census estimated a population of just under 2.5 million. More than 80% of the households rely on agricultural production as their main source of income (NBS, 2013). The first round of KHDS interviews was held in 1991-1994 with 915 households originating from 51 villages and urban areas across Kagera interviewed up to four times. The first follow-up survey was organized in 2004 with the aim of re-interviewing all individuals ever interviewed at the baseline (1991-94). This involved tracking individuals who had migrated away from the village to other parts of the region, elsewhere in Tanzania or to neighboring Uganda. More than 93% of the baseline households were re-contacted after a 10-year period (see Beegle et al. (2006))[6]. The second follow-up survey was organized in 2010. This time the tracking success rate was 92%; that is at least one original household member was interviewed for 92% of baseline households (see DeWeerdt et al. (2012)).Relative to comparable panel surveys, these household level attrition rates are exceptionally low (Alderman et al., 2001).

At the individual level, the re-interview rates among survivors were 82% in 2004 and 85% in 2010. The aim in the 2004 and 2010 survey rounds was to re-interview all of the surviving individuals who were on at least one of the 1991-1994 household rosters. In total, there were 6,353 individuals on these; by 2010, 1,275 of these had died while 85% of the surviving 4,996 individuals were re-interviewed.

At each round of the survey a complete multi-topic household questionnaire was administered to all split-off households containing individuals who had resided in the original baseline sample of households. Topics covered ranged from education, health,

---

[6]This excludes 17 households in which all previous household members were deceased.

employment and migration of individual household members, to household asset ownership, consumption expenditure, formal and informal networks, remittances, history of economic shocks, and more (see DeWeerdt et al. (2012) for a detailed description).

We now present some basic sample descriptive statistics and discuss the key features which make these data uniquely suitable for our analysis.

3.1 Sample Characteristics

Table 1 some descriptive statistics for the sample. We focus here on the sample we will use in the main analysis which excludes some households (we discuss which ones below). In total we focus on 2,780 households formed from 709 original baseline households. By 2010 these were still predominantly located in rural areas with nearly 70% in well connected or remore villages. About half relied on agriculture as the main source of household income. The great majority of households (79%) have a male head who, on average has around 6 years of schooling (equivalent to completing primary school). Even focusing on individuals with the highest level of education in the household, we see that this is equivalent to incomplete lower secondary schooling. The main outcome in our analysis is consumption growth between 2004 and 2010; the table shows that this was quite substantial between 2004 and 2010, with 2010 per capita consumption being nearly 60% higher than that in 2004. Previous work shows that consumption growth has been significantly higher among those who migrated out of the original villages than those who stayed, increasing the most for those who migrated out of the baseline region (nearly tripling between 1991 and 2010 - see Beegle et al. (2006) and De Weerdt and Hirvonen (2016)).

3.2 Household Consumption Expenditure and Shocks

We now turn to the key variables in the analysis set out in Section 2.3, including consumption expenditure and household income. Detailed consumption expenditure data were collected in each round of the survey. We utilise consumption expenditure data collected in the 2004 and 2010 rounds. The questionnaires included modules capturing food and non-food consumption, with differing recall periods to reflect seasonality of consumption of certain items and several checks built in to accurately capture consumption from home-production. The surveys also included price questionnaires which were used to generate temporal and spatial deflators for the consumption expenditure aggregates (DeWeerdt et al., 2012).

Unfortunately, our data do not include detailed information on household income

Table 1: Sample Characteristics

	Mean	SD
Location of household		
Regional capital/city	0.19	0.39
District capital/peri-urban town	0.12	0.33
Well-connected village	0.54	0.50
Remote village	0.15	0.36
Main source of hh income		
Non-agricultural wage employment	0.16	0.37
Non-agricultural self-employment	0.10	0.30
Agriculture	0.53	0.50
Other	0.21	0.41
Household consumption expenditure		
2004 consumption per capita (in 2010 TZS)	444,680.83	390770.05
2010 consumption per capita (in 2010 TZS)	702,459.58	745049.09
Household composition		
Head is male	0.79	0.41
Head age	41.56	15.48
Head education	6.18	3.22
Highest level of education in hh	7.42	2.72
Household size	4.64	2.39
Male age 0 to 15	0.45	0.50
Female age 0 to 15	0.48	0.50
Male age 6 to 60	1.06	0.74
Female age 6 to 60	1.12	0.75
Male age 61+	0.08	0.27
Female age 61+	0.12	0.34
Observations	2,780	

for the last (2010) survey wave. It is, therefore, not possible to run equations (11) to (12) as specified. Instead of income, however, we have information about the shocks, positive and negative, these households recall experiencing in each year between 2004 and 2010, collected in 2010. Specifically, for each year between 2004 and 2010 all individuals who had been members of the original households in 1991-94 were asked to report whether it was a very good, good, average, bad, or very bad year. For very good and very bad years they were further asked to give reasons and for very bad years ways in which they coped.

The top panel of Table 3 shows that only a fifth of households did not report experiencing a bad or very bad shock in any of the years between 2004 and 2010. Households reported experiencing a bad shock in an average of 2.5 years over this period. Good shocks and very good shocks are less prevalent, though still nearly two-thirds of the households report experiencing at least one during the recall period.

To give some idea of what the key risks for families in this context are, the middle panel shows the reasons given by the 1,627 individuals in the sample who report having had a very bad year at least once during the recall period. Most commonly cited reasons include family illness and death, as well as poor harvest. The bottom panel of the table suggests that family and friends are an important source of support during these hard times; two out of five individuals who report experiencing a very bad shock during the recall period said that they relied on family and friends as the main coping strategy. Other most commonly cited strategies include reducing consumption and taking on more work. In this paper we focus on shocks reported for 2010, the year immediately preceding the survey. The table shows that 35% of households reported experiencing a bad or very bad shock in that period, and nearly the same proportion (33%) a good or very good one.

3.3 Family Networks

A key feature of the data that makes our analysis possible is availability of information on a relatively large number of "family networks". We define a family network as the group of households formed by individuals who were living in a single household at the beginning of the data collection and that have subsequently split-off into different households, for a variety of reasons, including marriage and migration. Some of these split-off households are located in the same village as the original nucleus, others are in near-by villages and others still live in relatively far away places. Our data include a large sample of such networks; while the original 1991-94 sample consisted of 915

households, the 2010 sample contains members from 816 of these residing across 3,313 households. That is, on average each ot the original households had split into just over 4 households. Consistently with this Table 2 shows that by 2010 only about half of the original memebers of the 1991-94 households were still living in the original village; about a third moved out of the villages but remained in the region and 15% had left the region either relocationg to another part of Tanzania or leaving the country.

Column 2 in Table 2 shows proportions of original households (or family networks) that had at least one member residing in each of the listed locations. This presents a somewhat different picture. Although about half of the original 4,287 *household members* who were re-surveyd in 2010 had moved to a different location, the great majority (86%) of the networks still had at least one member residing in the original village in 2010. Similarly, while only 15% of the individuals had moved out of the region by 2010, they came from 40% of the family networks. Maps in Appendix A show changes in the spatial distribution of the households between 1991-94 and 2010.

Table 2: KHDS 2010 sample location

	Location of original household members in 2010	Family networks in 2010 with at least one individual in:
Same village as in baseline	0.52	0.86
Nearby village	0.09	0.31
Same region as in baseline	0.23	0.60
Another region in Tanzania	0.14	0.36
Another country	0.01	0.04
Observations	4,287	816

4 Information Quality

In the 2010 (last) data-colleciton round of the KHDS a big effort was made to capture interactions within the family networks and quality of information that members possess about each other's standard of living. This is the information that makes this data-set uniquely suited for our purposes.[7]

---

[7]Ours is not the first attempt to use these data to measure the quality of information in family networks. Using the same data De Weerdt et al. (2019) construct measures aimed at quantify the quality of information. Their methodology, however, is very different from ours.

Table 3: Shocks

	Mean	SD
Shock prevalence		
Any bad or very bad shock btw 2004 and 2010	0.80	0.40
Any good or very good shock btw 2004 and 2010	0.65	0.48
Years with bad/very bad shocks btw 2004 and 2010	2.57	2.11
Years with good/very good shocks btw 2004 and 2010	2.27	2.30
Any bad or very bad shock in 2010	0.35	0.48
Any good or very good shock in 2010	0.33	0.47
N	3,313	
Shock cause (2004-2010)		
Failed harvest	0.32	0.47
Loss of employment	0.12	0.32
Family illness/death	0.38	0.49
Loss of assets	0.08	0.27
Other	0.11	0.31
N	1,627	
Coping strategies (2004-2010)		
Reduced consumption	0.29	0.46
Sold Assets	0.18	0.39
Took on more work	0.25	0.43
Diversified (business/crops)	0.11	0.31
Support from formal/informal orgs	0.03	0.16
Relied on family and friends	0.39	0.49
Migrated for work	0.03	0.17
Other	0.16	0.36
N	1,627	

4.1 Measuring information quality

Specifically, in the 2010 round each household within a family network was asked a series of questions about each of the other households in the family network. The information collected includes how often each pair interact and in what way, history of recent monetary and in-kind transfers and, critically for us, their beliefs about asset ownership among the other households in the network. The asset list includes house, land, livestock, phone, TV, and motorized vehicle. Each household was asked about whether they themselves owned these assets and whether they believed that each of the other households did. For each dyad within the network, then, these data provide information on the "truth" (self-reported asset ownership for each household in the network) and the "beliefs" regarding all other households in the network. In this set-up, the quality of information that household i has about household j can differ from the information that household j has about household i. A limitation of the data is that original household members who were still living in the original baseline village in 2010 were not asked about eachother even if they were living in diffierent households so we have to assume that they have perfect information about eachother.

In the exercise we perform, we need to construct statistics that represent the quality of information between the *nodes* of the family network. In total once we exclude households with no split offs and households for which we have no 2004 data, or are missing data for key variables we have a sample of 709 family networks, 2,780 households and 12,693 dyads (descriptive statistics for these households are presented in Table 1 above).

In order to use these data to construct a measure of information quality within a family network (and the quality of the information about a generic individual wealth), we assume that there is a latent variable $\theta_{i,g}^{j}$, which represents the wealth or well being of household i as perceived by household j; where both households i and j belong to family g. Note that $\theta_{i,g}^{i}$ represents the 'true' latent variable for household i. While we do not have direct information about either the 'true' wealth of household i or how this wealth is perceived by household j, we have answers to questions about households i owning a number of assets given by all the households in the family to which i belongs, including i themself. We denote with $A_{i,g}^{k,j}$ a variable that indicates whether household j thinks that household i of family g owns asset k. $A_{i,g}^{k,i}$ indicates whether household i actually owns asset k.

Before delving into the description of the methodology we use to synthetise the information different households have of the wealth held by other member of the family

network, we present some evidence about the correctness of the information about individual assets. In Table 4, we notice that a relative high proportion of answers about individual assets are 'correct', in the sense that they coincide with the answer given by the individual owner. For all individual assets, the proportion of correct answers is close to 80%. However, less than half the dyads report correct answer for *all* assets. Therefore, we conclude that the information in the family networks in our sample is far from perfect.

Table 4: Quality of information: descriptive

	Mean	SD
Proportion of assets reported correctly		
by other households		
House	0.80	0.40
Land	0.79	0.41
Animals	0.79	0.40
Phone	0.77	0.42
TV	0.79	0.41
Car	0.74	0.44
Proportion of all assets reported correctly	0.78	0.29
All assets reported correctly	0.46	0.50
N	12,693	

To summarize the answers that individual dyads of households give about each other asset ownerships into a single index reflecting the quality of information about that dyad, we assume that, while the latent index $\theta_{i,g}^j$ is unobservable by the researcher, the information about asset ownership is related to such a latent variable by an Item Response Theory model. In particular, we assume that the standard of living latent variable $\theta_{i,g}^j$ determines the probability that household i owns asset k according to a 1-parameter Rasch model:[8]

$$Prob\{A_{i,g}^{k,i} = 1\} = \frac{e^{\beta^k + \theta_{i,g}^i}}{1 + e^{\beta^k + \theta_{i,g}^i}} \qquad \theta_{i,g}^i \sim \mathcal{N}(0, \sigma^2) \qquad (13)$$

When information is available about at least two assets, it is possible to estimate the parameters of the measurement system in equation (13). Given estimates of the relevant paramters, it is possible to obtain, from a set of measures $A_{i,g}^{k,i}$, an estimate of the unobservable household wealth $\theta_{i,g}^i$, which we denote by $\hat{\theta}_{i,g}^i$. The parameters of this model and the answers provided by household j about household i's assets can be also

---

[8]We tried to estimate a 2-parameter Rasch model but encountered some convergence issues.
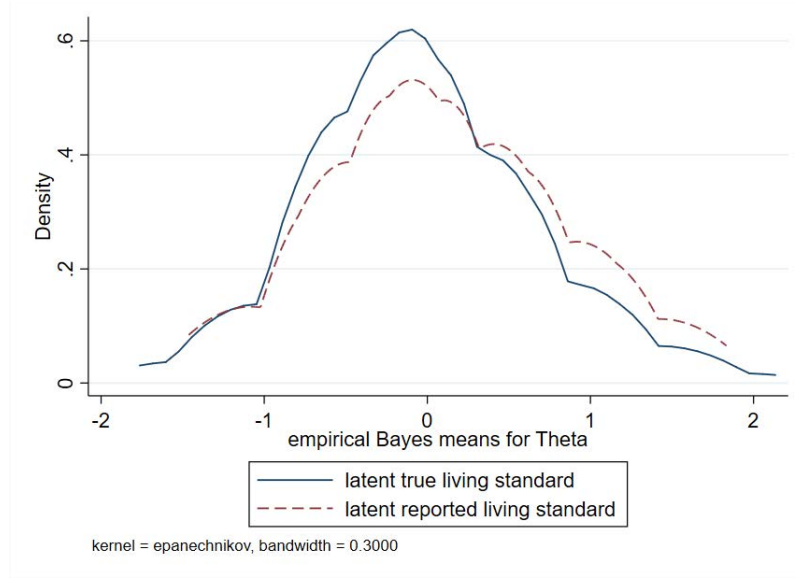
kernel = epanechnikov, bandwidth = 0.3000

Figure 1: Kdensity of "true" & "perceived" living standard θ

used to get an estimate of $\theta_{i,g}^j$, $\hat{\theta}_{i,g}^j$. The model in equation(13) effectively summarizes the information about asset ownership in a single index. Analogously, the same model can be used to summarize the answers provided by household j on household i so to obtain $\hat{\theta}_{i,g}^j$. The implicit assumption is that the latent factor representing j's perception about i's wealth can be represented by a model like the one in equation (13). Or, in other words, that household j's answers to questions about various items owned by household i are driven by the sample factor that drives the answers by their own wealth.

Figure 1 shows the distribution of the true and the perceived living standard estimates from the IRT model. The fact that the two are not perfectly overlapping suggests that households in the network are not perfectly informed about each-other's living standards.

4.2 Constructing measures of information quality

We then take assume the difference between $\hat{\theta}_{i,g}^j$ and $\hat{\theta}_{i,g}^i$ as reflecting the quality of information about income flows and shocks of the network members. We note that the model in equation (13) assumes that the latent factor $\theta_{i,g}^i$ is normally distributed. This might not be the best representation of the variability of economic well-being witin a network. We therefore consider three alternative measures that capture the different between i's 'true well-being' and the perception of the same factor as held by household

19

j. In particular, we define three different measures to capture such difference, which we label $q_{i,j}^{g,\ell}, \ell = 1, 2, 3$.

$$
\begin{aligned}
q_{i,j}^{g,1} &= |\hat{\theta}_{i,g}^i - \hat{\theta}_{i,g}^j| \\
q_{i,j}^{g,2} &= |e^{\hat{\theta}_{i,g}^i} - e^{\hat{\theta}_{i,g}^j}| \\
q_{i,j}^{g,3} &= e^{|\hat{\theta}_{i,g}^i - \hat{\theta}_{i,g}^j|}
\end{aligned}
\tag{14}
$$

The first measure takes the estimates of the latent factor $\theta_{i,g}$ straight from the Rasch model in equation (13). The second measure considers the exponents of the latent factor estimated with the Rasch model, implicitly assuming that the factor being modeled in equation (13) is the log of the factor of interest. In this case, therefore, we express the distance between the actual and perceived status in term of the levels of this variable, as specified in the second line of equation 14. The third measure is somewhat in the middle of the first two. Given these definitions of the distance between the true and perceived well-being status, we construct an index that varies between 0 and 1. In particular, for each of the measures in (14) we define:[9]

$$
\alpha_{ij}^{g,\ell} = \frac{1}{1 + q_{i,j}^{g,\ell}}, \quad \ell = 1, 2; \quad \alpha_{ij}^{g,3} = \frac{2}{1 + q_{i,j}^{g,3}}.
\tag{15}
$$

When $q_{i,j}^{g,\ell}$ is zero, that is the information j has about i is perfectly accurate, $\alpha_{i,j}^{g,\ell} = 1$. Vice-versa, when the same information is very inaccurate, $\alpha_{i,j}^{g,\ell}$ tends to zero.

Taking the simplest of these measures $(\alpha_{ij}^{g,1})$, we now analyse how our measure of the quality of information between two households are related to a series of characteristics that we would expect it to be related to. Most obviously, households that are located closer together and those that interact more should have better information about eachother. Table 5 shows that this is indeed the case - the degree of misperception is highly positively correlated with physical distance as well as with frequency of contact. For households living close to eachother (bottom quintile of geodesic distance from each other in which households are on average 400 meters apart), our measure is closed to perfect at 0.97. [10] However, already for households at the median distance from each other (17.2kms), the average index is 0.75. The same index declines to 0.67 for the households living in top quintile of distance from each other.

---

[9]We multiply the third measure by 2 so that it varies between zero and 1.

[10]This is mostly mechanical as we assume perfect information for all households within a family network still located in the baseline village, since data on how much they know about eachother were not collected for these.

A similar pattern is observed if we consider how our index varies with the frequency of contact. For households that have spoken to each other less than a month before the interview, the average index is 0.86. This declines with frequency: for households that have not spoken for more than 5 years the index is 0.64.

Table 5: Quality of information and observables

	Mean	SD
Quality of information by distance btw households		
Q1 (0.4km)	0.97	0.11
Q2 (4.2km)	0.81	0.23
Q3 (17.2km)	0.75	0.24
Q4 (81.6km)	0.70	0.25
Q5 (613.5km)	0.67	0.24
Quality of information by last time households spoke		
Less Than A Month Ago	0.86	0.21
Less Than A Year Ago	0.74	0.23
Less Than 2 Years Ago	0.71	0.25
Less Than 5 Years Ago	0.66	0.25
More Than 5 Years Ago	0.64	0.26
Don't Remember	0.59	0.26
N	12,693	

Having constructed the indexes $\{\alpha_{i,j}^{g,\ell}, \ell = 1, 2, 3\}$ for each dyad in the network, we can organise them into a matrix $A^{g,\ell} = [\alpha_{i,j}^{g,\ell}]$, for each family network in our data, where we set $\alpha_{i,i}^{g,\ell} = 0$, $\forall i$. The matrices $A^{g,\ell}$ so constructed are weighted adjacency matrices. We note that, unlike many of these matrices used in the literature, their elements are not binary, as they can take continuous values between 0 and 1. Furthermore, these matrices can be asymmetric, as the quality of information that household i has about household j might be different from the quality of information that j has about i.

Given an adjacency matrix $A^{g,\ell}$, for each household in family g, we can now construct measures of their position in the network. In our empirical application, we will use measures of degree centrality, which can be obtained summing or averagiing, for each household, the elements of the row or the columns of the adjacency matrix correspoinding to that household. As the matrices we are considering is not symmetric, the measures obtained averaging the rows or the columns are different. Averaging over the rows of the adjaciency matrices we get:
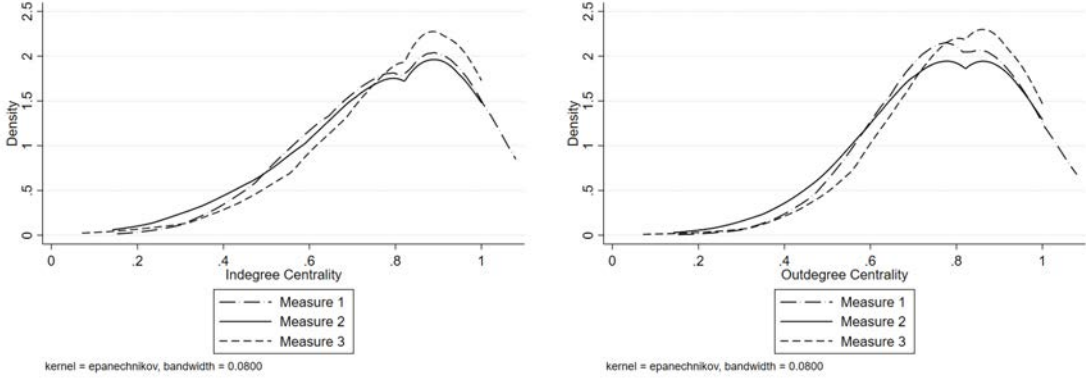
(a) Indegree Centrality $(InQ_i^{h,\ell})$ (b) Outdegree Centrality $(OutQ_i^{h,\ell})$

Figure 2: Kdensity of $InQ_i^{h,\ell}$ and $OutQ_i^{h,\ell}$

$$InQ_i^{g,\ell} = \frac{1}{K_g - 1} \sum_{k \neq i} \alpha_{ik}^{g,\ell} \tag{16}$$

where K_g is the number of households in family g. The expression in equation (16) is the in-degree centrality derived from the adjacency matrix $A^{g,\ell}$ and represents the average quality of the information the network has about the wealth of household i. Analogously, we can construct the out-degree centrality measure for household i averaging the elements of the matrix $A^{g,\ell}$ corresponding to column i. This measure represents the average quality of the information household i has about the other network members.

$$OutQ_i^{g,\ell} = \frac{1}{K_g - 1} \sum_{k \neq i} \alpha_{ki}^{g,\ell} \tag{17}$$

Finally, we can also define the quality of information in family g averaging the individual measures as:

$$IQ^{g,\ell} = \frac{1}{K_g} \sum_j^{K_g} InQ_j^{g,\ell} \tag{18}$$

We notice that the expression in equation (18) can be constructed either from equation 16) or equation (17).

In Figures 2, for each of our information quality measures, we plot the density distribution of the in-degree and out-degree centrality in our sample of individual households. The three measures, with the possible exception of the third one, are distributed in a reasonably similar fashion, which spans a large set of values. The mode of the three distributions is above 0.80. The three distributions seem to be left-skewed.

Table 6 shows the means and standard deviations of the various degree centrality measures considered, both at the family network and the household levels; for the household level measures we also include the matrix of correlation coefficients. At the family network level, mean degree centrality across the three measures is around 0.8; it is lowest for the second measure at 0.7. A similar picture emerges, not surprisingly, from the household level measures of degree centrallity. For each of the three measures, in-degree and out-degree centrality vary very similarly in our sample with a correlation coefficient which is above 0.9. We notice, however, that, for each of the three measures, the correlation between in- and out- degree centrality is much lower, at around 0.3. In our sample there seems to a considerable level of asymmetry in the adjacency matrices we construct with our measures of information quality.

Table 6: Summary statistics and correlation matrix: network centrality measures

	Mean	SD	$InQ_i^{h,1}$	$InQ_i^{h,2}$	$InQ_i^{h,3}$	$OutQ_i^{h,1}$	$OutQ_i^{h,2}$	$OutQ_i^{h,3}$
Household level								
$InQ_i^{h,1}$	0.79	0.17	1.000					
$InQ_i^{h,2}$	0.77	0.20	0.944	1.000				
$InQ_i^{h,3}$	0.81	0.18	0.983	0.929	1.000			
$OutQ_i^{h,1}$	0.78	0.16	0.318	0.313	0.303	1.000		
$OutQ_i^{h,2}$	0.77	0.18	0.316	0.322	0.303	0.956	1.000	
$OutQ_i^{h,3}$	0.80	0.16	0.305	0.301	0.294	0.983	0.943	1.000
N	2,780							
Family network								
$IQ^{h,1}$	0.79	0.13						
$IQ^{h,2}$	0.77	0.15						
$IQ^{h,3}$	0.80	0.13						
N	709							

5   Risk sharing and the quality of information within family networks.

Following the empirical strategy set out in Section 2.3, to measure the extent of risk sharing and deviations from first best allocations, similarly to Townsend (1994), we relate changes in individual consumption to idiosyncratic shocks; under perfect risk sharing, after controlling for risk-sharing group x time effects, these shocks should be diversified. In particular, we estimate the following regression:

$$\Delta_s ln(u'(c_t^{jg})) = \nu_t^g + \gamma_1 BS_t^{jg} + \gamma_2 GS_t^{jg} + + \epsilon_t^{jg}$$

where $\Delta_s ln(u'(c_t^{jg}))$ is the change in consumption for individual household j in family network g, GS_t^{jg} and BS_t^{jg} are indicators for good and bad shocks received by that household and ν_t^g reflect family network level resources and shocks at time t. These shocks might be in part attenuated by risk sharing mechanisms (such as saving or interactions with other groups) that we do not consider explicitly. The ν_t^g's are estimated as coefficients on group x time dummy variables. As discussed above, the definition of the group g is taken as given in our framework. g can represent villages or family networks or another predetermined group. In our exercise we consider risk sharing within family networks.

Unlike Townsend (1994), given the available information in our data, we do not use individual income to test for perfect risk sharing. Instead, we use the information on individual shocks of various nature, which we have discussed in Section 3.2. In any case, in the presence of perfect risk-sharing we would expect the coefficients γ_1 and γ_2, after controlling for group level shocks, represented by the group x time ν_t^h dummies, to be zero.

This is not what we find: the results in the first column of Table 7 show that the coefficients on the bad shock indicators (γ_1) is statistically significant and negative, so that experiencing a bad or very bad shock in 2010 is related to a decrease in individual household consumption between 2004 and 2010. On the other hand, the coefficient on the 'good shock' is very small and, while positive, not significantly different from zero.

This evidence represents a rejection of perfect risk sharing in that idiosyncratic shocks are not fully diversified within the family network and is consistent with findings in other work (including work on this specific study context). The main purpose of this paper is to relate these deviations from perfect risk sharing to imperfect information. As a first step in that direction, we interact the income shock variables with measures that refelct the quality of information in the network. Following the approach suggested in equation (11), we estimate the following regression:

$$\Delta_s ln(u'(c_t^{j,g})) = \nu_t^g + (\gamma_{01} + \gamma_{11}IQ^{g,\ell})BS_t^{j,g} + (\gamma_{02} + \gamma_{12}IQ^{g,\ell})GS_t^{j,g} + \epsilon_t^{j,g} \qquad (19)$$

where we interact good and bad idiosyncratic shocks with the information quality in the family network g, using the three different versions of $IQ^{g,\ell}$ derived in equation (18).As we discuss in Section 4, $IQ^{g,\ell}$ is close to zero in family network with information of very poor quality and is 1 when informationn about asset ownership is perfect. We report the results in columns (2) to (4) of Table 7.

24

Table 7: Sensitivity of risk-sharing to quality of information within family network

Inf. quality measure	(1) none	(2) $IQ^{h,1}$	(3) $IQ^{h,2}$	(4) $IQ^{h,3}$
Bad shock in 2010	-0.134***	-0.420	-0.324	-0.324
	(0.035)	(0.256)	(0.223)	(0.223)
Good shock in 2010	0.00289	0.679***	0.561**	0.561**
	(0.036)	(0.262)	(0.227)	(0.227)
Good shock X mean degree cent $IQ^{h,\ell}$		-0.863***	-0.729**	-0.729**
		(0.332)	(0.293)	(0.293)
Bad shock X mean degree cent $IQ^{h,\ell}$		0.365	0.248	0.248
		(0.324)	(0.287)	(0.287)
Constant	0.434***	0.434***	0.434***	0.434***
	(0.0125)	(0.0240)	(0.0241)	(0.0241)
Observations	2,780	2,780	2,780	2,780

Standard errors in parentheses

Dep var = change in lnpcconsumption btw 2004-2010 (2010 prices); Family network FE

Shock = 1 if reported by anyone in household

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Given the specification in equation (19), the coefficients on the good and bad shocks now represent how household consumption is affected by those idiosyncratic shocks in networks with very low levels of inofrmation quality. On the other hand, the impact of bad and good shocks in networks with perfect information quality is obtained summing by summing the coefficients in rows (2) and (3) for the good shocks and in rows (1) and (4) for the bad shocks.

Starting with the coefficient on the good shocks, which in column (1) is effectively 0, we notice that it becomes positive and strongly significant in columns (3) an (4) for networks with very poor information quality. On the other hand, in networks with very good information quality, the coefficient is close to be zero. These results hold for all three definitions of information quality that we use.

Moving now to the negative shocks, for which the coefficient in column (1) is -0.13 and significant, we notice that for networks with very poor quality of information, the coefficient is larger in absolute value. Moreover, the coefficient on the interactions is of the opposite sign (row (4)) so that the sum of these cofficients and those in the first row is close to zero. These estimates, however, are not very precise, so that all these coefficients are not significantly different from zero. These results are also very similar across the three different measures of information quality.

We conclude that the evidence in Table 7 constitutes suggestive evidence that the quality of information affects the amount of risk sharing that we observe in family networks. The evidence is particularly convincing for positive shocks, although the point estimates for negative shocks offer a similar story, albeit with low precision.

Having considered the quality of information in a network we now move to considering how the sensitivity of household consumption to idiosyncratic shocks is affected by their position in the family network, as measured by the network centrality meaures we have considered. In particular, we estimate the following equation:

$$\Delta_s ln(u'(c_t^{j,g})) = \nu_t^g + (\gamma_{01} + \gamma_{21} IP_i^{g,\ell}) BS_t^{j,g} + (\gamma_{02} + \gamma_{22} IP_i^{g,\ell}) GS_t^{j,g} + \epsilon_t^{j,g} \qquad (20)$$

where $IP_i^{g,\ell}$ is either the in-degree centrality $InQ_i^{h,\ell}$ as computed in equation (16) or the out-degree centrality measure as constructed in equation (17). As above, we compute these statistics for each of the adjacency matrices we derived. The first set of results, which we report in Table 8, measures how the quality of information about the situation of household i among other households in the family network affects the sensitivity of its consumption to shocks, while the second, reported in Table 9, measures how the quality of information household i has about the situations of other households in the network affects its sensitivity to its own shocks.

The first column of Table 8 is reproduced from the same column in Table 7. When in Columns 2 to 4 we add the measures of in-degree centrality, the results change substantially, especially for the second measure (column(3)). First, we notice that the in-degree centrality measures themselves are significant. Households that are more central experience, on average, lower levels of consumption growth. More importantly, the coefficients on both the negative and positive shocks increase in size and are both statistically significantly different from zero. These coefficients are relevant for households that have very low levels of in-degree centrality, that is households for which the information that other households in the network have is very poor. This result is particularly evident for the second and third measures that we use.

When we look at the interactions of the in-degree centrality with the shocks, we notice that both interactions are significantly different from zero and attract a coefficient which is opposite in sign to the coefficients on the shocks. For instance, in column (3), for a household which experiences a bad shock and whose economic situation is well known to other households in the network so that it has an in-degree centrality of 1, the effect of a bad shock is given by -0.047=0.379-0.426), which is not statistically

26

Table 8: Sensitivity of risk-sharing to quality of information other households in a family network have about household

In-degree cent. measure	(1) none	(2) $InQ_i^{h,1}$	(3) $InQ_i^{h,2}$	(4) $InQ_i^{h,3}$
Bad shock in 2010	-0.134***	-0.457***	-0.426***	-0.419**
	(0.0350)	(0.164)	(0.146)	(0.164)
Good shock in 2010	0.00289	0.428**	0.279*	0.417**
	(0.0363)	(0.166)	(0.144)	(0.166)
HH indegree cent $InQ_i^{h,\ell}$		-0.337**	-0.252*	-0.327**
		(0.152)	(0.139)	(0.148)
Good shock X HH indegree cent $InQ_i^{h,\ell}$		-0.535***	-0.360**	-0.511**
		(0.206)	(0.182)	(0.202)
Bad shock X HH indegree cent $InQ_i^{h,\ell}$		0.413**	0.379**	0.357*
		(0.203)	(0.184)	(0.199)
Constant	0.434***	0.697***	0.627***	0.695***
	(0.0241)	(0.121)	(0.109)	(0.121)
Observations	2,780	2,780	2,780	2,780

Standard errors in parentheses

Dep var = change in lnpcconsumption btw 2004-2010 (2010 prices); Family network FE

Shock = 1 if reported by anyone in the household

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

different from zero. Analogously, for the same households a positive shock attracts a coefficient of -0.081=0.279-0.36. Similar results hold for the third measure.

Table 9: Sensitivity of risk-sharing to quality of information household has about other households in the family network

Out-degree centr. measure	(1) none	(2) $OutQ_i^{h,1}$	(3) $OutQ_i^{h,2}$	(4) $OutQ_i^{h,3}$
Bad shock in 2010	-0.134***	-0.139	-0.126	-0.0997
	(0.0350)	(0.185)	(0.160)	(0.186)
Good shock in 2010	0.00289	0.116	0.0852	0.162
	(0.0363)	(0.191)	(0.165)	(0.195)
HH outdegree cent $OutQ_i^{h,\ell}$		0.124	0.0523	0.195
		(0.183)	(0.161)	(0.180)
Good shock X HH outdegree cent $OutQ_i^{h,\ell}$		-0.144	-0.107	-0.200
		(0.239)	(0.209)	(0.238)
Bad shock X HH outdegree cent $OutQ_i^{h,\ell}$		0.00741	-0.0104	-0.0424
		(0.233)	(0.204)	(0.230)
Constant	0.434***	0.336**	0.393***	0.278*
	(0.0241)	(0.146)	(0.126)	(0.146)
Observations	2,780	2,780	2,780	2,780

Standard errors in parentheses

Dep var = change in lnpcconsumption btw 2004-2010 (2010 prices); Family network FE

Shock = 1 if reported by anyone in the household

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

After the in-degree centrality, we also look at whether out-degree centrality play a role. That is we investigate whether the quality of information that each household i has about the economic status of the rest of the family network affects the relationship between consumption changes and idiosyncratic shocks. With this objective, we re-estimate equation (20) using as the $IP_i^{g,\ell}$ variable the out-degree centrality measures constructed in equation (17). We report the results in Table 9.

In this case, the information quality variable does not seem to play any role. None of the terms involving such a variable are stastically significant and, as the results become much nosier, no clear patterns emerge. We conclude that out-degree centrality does not play any role in the amount of risk sharing we observe in our sample.

Overall, these results suggest that the level of information within a network matters. It also matters, however, *what* information we consider. The results in Table 7 provide suggestive evidence that in family networks with better information consumption

allocations are closer to what would be observed under perfect risk sharing.

The results based on the quality of information about household income is even stronger and more precise. It indicates that households with high in-degree centrality are less sensitive to indiosyncratic shocks. It also indicates that they experience, on avergage, slower consumption growth. On the other hand, the sensitivity of individual household to idiosyncratic shocks does not seem to be affected by out-degree centrality. The next step to this analysis is to consider models that can justify these patterns.

6   Conclusions

In this paper, we have studied the relationship between risk sharing within family networks and the quality of the information within these networks. To this end we have used a unique data set from Tanzania that has followed more than 700 family networks over a period of nearly 20 years, even when some of their members migrated outside of the original villages that they we living when the data collection was started.

In our data, we can construct measures of the quality of the information flows between any two member households of the family network. We show that our method of constructing measures of information quality yields estimates that vary in a way that is consistent with what one would expect: households that are geographically closer to each other or that talk to each other often have better information about each other.

These estimates allow us to construct some average measures of the quality of information within a family network. In addition, we can also construct measures of quality of information for individual households in the network: for each household we consider the average quality of the information that the other network members have about the household's wealth and the average quality of the information that the household has about the other members of the network.

We show that the household level measures are equivalent to in-degree and out-degree network centrality measures derived from a weighted and asymmetric adjacency matrix. Each element of the matrix is not forced to be 0 or 1 and can take any value in the unit interval, representing the quality of the information between a specific dyad in the network. The matrix is not necessarily symmetric because the quality of the information that household i has about household j is not necessarily the same as the quality of the information that j has about i. This fact explains why in-degree and out-degree centrality are different.

Finally, we relate the quality of information we derive to the degree of risk sharing, as measured by some standard regressions of the type proposed by Townsend (1994) re-

lating household consumption changes to household shocks, after controlling for family network level shocks. We show that in networks with better information quality, consumption allocations are closer to what one would observe under perfect risk sharing. More precisely, we show that households that more is known about in the network, are less sensitive to idiosyncratic shocks.

To the best of our knowledge, this is the first study that relates information quality to risk sharing. While the results are very intuitive, the next step, that is being taken in Attanasio et al. (2020, in progress), is to consider risk sharing arrangements with information frictions and relate the consumption allocations that arise from such arrangements to the properties of the network that can be derived from the informatin data that we have.

Finally, we conclude with a note of caution. Our study does not characterise risk sharing fully. We only consider, partly because of the nature of the data we have, risk sharing within a specific network and ignore possible other mechanisms that could involve individuals outside the family network or other arrangements. Furthermore, we do not consider other frictions, such as the imperfect enforceability of contracts. Neither do we consider the process of network formation and specific processes (such as migration) that lead to better or worse information quality. These are important areas of future research.

References

Albarran, P. and Attanasio, O. P. (2003). Limited commitment and crowding out of private transfers: evidence from a randomised experiment. *The Economic Journal* 113(486):C77–C85.

Alderman, H., Behrman, J., Kohler, H.-P., Maluccio, J. A., and Watkins, S. (2001). Attrition in Longitudinal Household Survey Data: Some Tests for Three Developing-Country Samples. *Demographic Research* 5(4):79–124.

Ambrus, A., Gao, W. Y., and Milan, P. (2019). Informal risk sharing with local information.

Ambrus, A., Mobius, M., and Szeidl, A. (2014). Consumption risk-sharing in social networks. *American Economic Review* 104:149–182.

Atkinson, A. (1991). International lending with moral hazard and risk of repudiation .

Attanasio, O., Krutikova, S., Milan, P., and de Paula, A. (2020, in progress). Constrained optimal risk sharing in networks with imperfect information.

Attanasio, O. and Rios-Rull, J.-V. (2000). Consumption smoothing in island economies: Can public insurance reduce welfare? *European economic review* 44(7):1225–1258.

Attanasio, O. P. and Pavoni, N. (2011). Risk sharing in private information models with asset accumulation: Explaining the excess smoothness of consumption. *Econometrica* 79(4):1027–1068.

Beegle, K., Weerdt, J. D., Dercon, S., Ksoll, C., Kessy, F., Koda, G., Lwihula, G., Kwesigabo, G., Mujinja, P., and Semali, I. (2006). Kagera health and development survey 2004. World Bank Microdata library.

Bold, T. and Broer, T. (2018). Risk sharing in village economies revisited. Unpublished.

Broer, T., Kapicka, M., and Klein, P. (2017). Consumption risk sharing with private information and limited enforcement. *Review of Economic Dynamics* 23:170–190.

Cole, H. L. and Kocherlakota, N. R. (2001). Efficient allocations with hidden income and hidden storage. *The Review of Economic Studies* 68(3):523–542.

De Weerdt, J., Genicot, G., and Mesnard, A. (2019). Asymmetry of information within family networks. *Journal of Human Resources* 54(1):225–25.

De Weerdt, J. and Hirvonen, K. (2016). Risk sharing and internal migration. *Economic Development and Cultural Change* 65.

Denderski, P. and Stoltenberg, C. A. (2019). Risk sharing with private and public information.

DeWeerdt, J., Beegle, K., Lilleor, H. B., Dercon, S., Hirvonen, K., Kirchberger, M., and Krutikova, S. (2012). Kagera health and development survey 2010. *Working Paper Research Immigration and Integration.*

Foster, A. D. and Rosenzweig, M. R. (2001). Imperfect commitment, altruism, and the family: Evidence from transfer behavior in low-income rural areas. *Review of Economics and Statistics* 83(3):389–407.

Kinnan, C. (2020). Distinguishing barriers to insurance in thai villages. *Journal of Human Resources* Forthcoming.

Kinnan, C. and Townsend, R. (2012). Kinship and financial networks, formal financial access, and risk reduction. *The American Economic Review* 102(3):289–293.

Kocherlakota, N. (1996). Implications of Efficient Risk Sharing without Commitment. *The Review of Economic Studies* 63(4):595–609.

Ligon, E. (1998). Risk sharing and information in village economies. *The Review of Economic Studies* 65(4):847–864.

Ligon, E., Thomas, J., and Worrall, T. (2002). Informal Insurance Arrangements with Limited Commitment: Theory and Evidence from Village Economies. *The Review of Economic Studies* 69(1):209–244.

Munshi, K. and Rosenzweig, M. (2016). Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *The American Economic Review* 106(3):46–98.

NBS (2013). Population and housing census: Population distribution by administrative units; key findings. URT (United Republic of Tanzania).

Rosenzweig, M. (1988). Risk, implicit contracts and the family in rural areas of low-income countries. *The Economic Journal* 98:1148–1170.

Townsend, R. (1994). Risk and insurance in village india. *Econometrica* 62(3):539–591.

Udry, C. (1994). Risk and insurance in a rural credit market: An empirical investigation in northern nigeria. *The Review of Economic Studies* 61(3):495–526.

Wilson, R. (1968). The theory of syndicates. *Econometrica* 36:119–132.
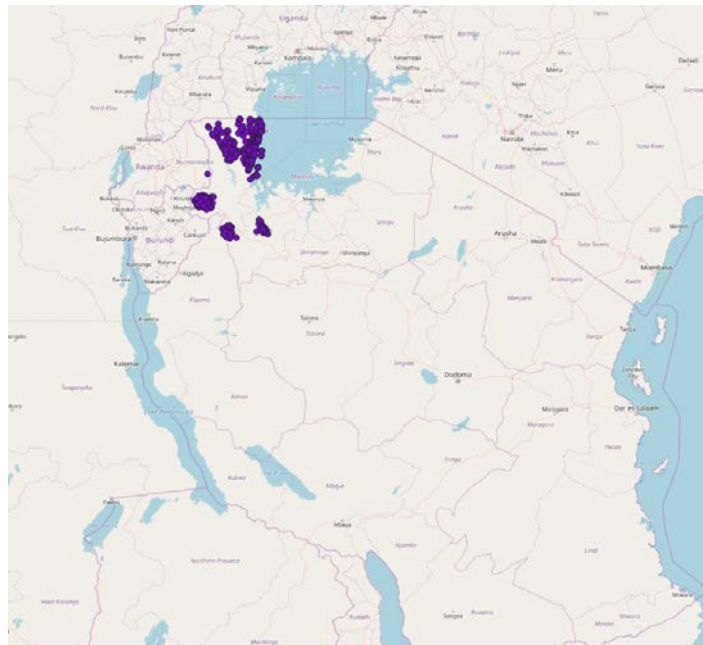
A   Location of the KHDS sample in 1991 and 2010
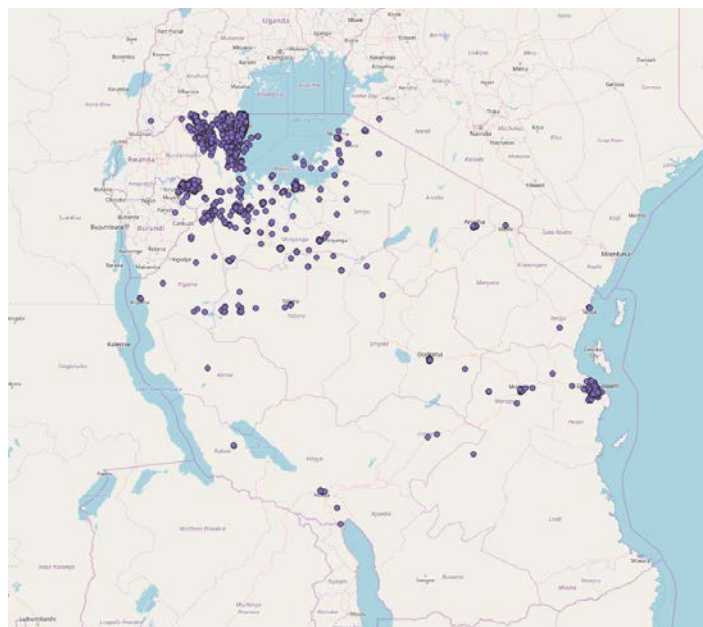

Figure 4: 1991

Figure 5: 2010