The Social Economics of Networks and Learning

Pau Milán

TESI DOCTORAL UPF / ANY 2016

DIRECTOR DE LA TESI
Jan Eeckhout
Departament D'Economia i Empresa

Universitat Pompeu Fabra
Barcelona

To Montse and Salva

Acknowledgments

This thesis would not exist without the remarkable guidance and encouragement I received from three exceptional researchers. Over the years, they have not only become my mentors, but also true friends.

From the very beginning Joan de Martí trusted in my ability to conduct serious research. It is from him that I acquired the conviction to pursue an academic career, and, after long conversations, we developed joint work and fostered a unique friendship that I am extremely proud of. Joan has always challenged me to greater heights, and I am always grateful to count on his support.

I am hugely indebted to Jan Eeckhout for his acute sense of what first-class research should look like. As my advisor, he taught me to think more profoundly about the appropriate role of research, and, as a result, he refined my academic taste. He was always elegant and never complacent, and I benefited greatly from his advice.

Finally, in the last couple years I had the great fortune to work with Vasco Carvalho. In fact, I would certainly have not gone on the academic job market had it not been for Vasco's encouragement, dedication, and immense generosity. He continuously made sure I was on track, motivated, prepared. The trips to Cambridge always remain a pleasant reminder of his easy demeanor, and good humor, that led us to become good friends.

I also owe many thanks to the faculty at UPF and CREI. The comments and discussion I received in seminars, workshops, and meetings where of tremendous help in sharpening and refining my research projects. I remember lengthy conversations with Gabor Lugosi, Fabrizio Germano, Christian Brownlees, Omrios Papaspiliopoulos, and Larbi Alaoui.

Similarly, I want to thank the faculty and students at The Univeristy of Cambridge for making me feel at home and for dedicating so much of their time to helpful discussions on my research, specifically of my job-market paper. In particular, I want to thank Sanjeev Goyal, Kaivan Munshi, Matt Elliott, David Minarsch, Axel Gottfries, Roberto Guadarrama, and Anja Prummer.

A number of fellow PhD students at UPF became close friends over the years; we shared ideas, frustrations, and even small victories. I want to thank Jorg Stahl, Eduard Talamás, Marc Goñi, Miguel Martínez, and Egemen Eren.

I always keep Michael Anreiter in my thoughts, for his unique energy and bright disposition.

I especially want to thank Jordi and Afra who have for long inhabited a particular place in my life and without whom it is impossible to imagine my development the past several years.

My family – old and new– here in Barcelona has increased these past several years. An exceptional friendship emerged with Andy, and he made sure I found myself surrounded with caring individuals, concerned about my progress and well-being. I have benefitted greatly from all the warm moments, the late-night jam sessions, the ”vermutillos”, the blind beer, and the long meals that I shared with everyone that has been with me these past years: Avis, Manel, Marc i les Nenes, Andy, Alex, Bruno, Cris, Elia, Graham, Hurón, Isa, Joe, Kenchan, Noelia, Pilar, Ramón, Tria, Sian, Santi ... and many others.

I also want to thank Nuriya, who bared the brunt of it this past year, and cared for me, shrewdly and lovingly. I am very grateful to have shared this past year together.

Then there are those old friends who, on my frequent visits to Chile, maintained full confidence in me and in my ability to realize my academic objectives. Many doubtful moments where relieved by their words: Javiera, Vero, Chofa, Vinchuka, Sebas, David, Juancho,

Finally, I know this thesis would never have succeeded had it not been for all that I was allowed to unload on my beautiful family. Although we are far away, I availed myself, repeatedly, of my mother’s patience and tenderness. I often heard my father’s sound advice, and my sister always found time to send warm letters of encouragement. As this project developed, so did another exciting project in Chile, and the enthusiasm the three of you sent my way no doubt kept me engaged in my own successes. I remain amazed, every day, at my good fortune to count you as my own.

Abstract

This thesis explores various economic environments where the structure of social interactions across individuals determines outcomes. In the first chapter, I study mutual insurance arrangements restricted on a social network. I test the theory on data from Bolivian communities. I find that the observed exchanges across households match the network-based sharing rule predicted by the theory. I argue that this framework provides a reinterpretation of the standard risk sharing results, predicting household heterogeneity in response to income shocks. In the second paper, I study individual and collective behavior in coordination games where information is dispersed through a network. I show how changes in the distribution of connectivities in the population affect the types of coordination in equilibrium as well as the probability of success. In the third chapter, I explore a framework of learning and turnover in the labor market. I show that positive assortative matching (PAM) extends beyond the stable environment of Eeckhout and Weng (2010) to a situation of residual uncertainty that exhibits periods of unlearning. I also extend this setting to allow for career concerns and I show that PAM can only be sustained under strong assumptions.

Resumen

Esta tesis explora diversos entornos económicos en los que la estructura de las interacciones sociales entre los individuos determina los distintos resultados. En el primer capítulo, se estudia acuerdos de seguro mutuo restringidos en una red social. Utilizo datos de comunidades bolivianas para medir las predicciones teóricas y encuentro que los intercambios observados entre los hogares coinciden con la regla de reparto basada en la red obtenida por la teoría. Sostengo que este marco ofrece una reinterpretación de los resultados estándar de distribución de riesgos, prediciendo heterogeneidad entre los hogares en respuesta a los shocks de ingresos. En el segundo artículo, estudio el comportamiento individual y colectivo en juegos de coordinación, donde la información se dispersa a través de una red. Demuestro cómo los cambios en la distribución de las conectividades de la población afectan a los tipos de coordinación en equilibrio, así como la probabilidad de éxito. En el tercer capítulo, analizo un marco de aprendizaje y cambio de personal en el mercado de trabajo. Muestro que emparejamiento selectivo positivo (PAM) se extiende más allá del entorno estable de Eeckhout y Weng (2010) a una situación de incertidumbre residual que exhibe períodos de des-aprendizaje. También extiendo esta configuración para permitir elementos de career concerns y muestro que el equilibrio de PAM sólo puede sostenerse bajo fuertes supuestos.

Preface

This doctoral thesis brings together the results from three independent research projects at the intersection between information economics, development economics and the economics of social interactions. The three essays share a common theme in that the particular structure of interactions between economic agents affects payoffs, and therefore of economic outcomes.

In the first chapter, I investigate mutual insurance arrangements restricted on a social network. My approach solves for Pareto-optimal sharing rules in a situation where exchanges are limited within a given social network. I provide a formal description of the sharing rule between any pair of linked households as a function of their network position. I test the theory on a unique data set of indigenous villages in the Bolivian Amazon, during the years 2004 to 2009. I find that the observed exchanges across families match the network-based sharing rule, and that the theory can account for the deviation from full insurance observed in the data. I argue that this framework provides a reinterpretation of the standard risk sharing results, predicting household heterogeneity in response to income shocks. I show that this network-based variation in consumption behavior is borne out in the data, and that it can be interpreted economically in terms of consumption volatility.

In the second chapter, co-authored with Joan de Martí, I study individual and collective behavior in global games of regime change where information is dispersed through networks. Agents can choose between attacking and not attacking a status quo whose strength is unknown. Communication with neighboring players introduces local correlations in posterior beliefs and also induces more accurate information. We provide general sparseness conditions on networks that allows for asymptotic approximations. We characterize equilibrium behavior in these cases, where the accuracy effect dominates the correlation effect. Following this equilibrium analysis, we show how changes in the distribution of connectivities in the population affect the types of coordination in equilibrium as well as the likelihood of a successful rally. We find that without a public signal strategic incentives align, and the probability of success remains independent of the type of network. With a public signal the network?s degree distribution unambiguously affects the probability of success, although the direction of change is not monotone, and depends crucially on the cost of attack.

Finally, in the third chapter I develop a framework where workers and firms learn continuously about the worker's productivity type, which itself fluctuates randomly and continuously. Workers receive a competitive spot wage and must decide when to switch between firms, given their belief about their own type. Under supermodularity in production, I show that positive assortative matching

(PAM) extends beyond the stable environment of Eeckhout and Weng (2010) to a situation of residual uncertainty that exhibits periods of unlearning, defined as increasing levels of uncertainty along a match. I show that risk-neutrality of workers and skills evolving as a martingale are sufficient to retain PAM. I then extend this setup to allow workers to exert a level of effort subject to classical career concerns (a la Hölmstrom) and find that PAM can only be sustained under strong parameter restrictions.

Contents

List of Figures

List of Tables

Chapter 1

NETWORK-CONSTRAINED RISK SHARING IN VILLAGE ECONOMIES

1.1 Introduction

Vast areas of the developing world rely on informal mechanisms of insurance against random fluctuations in crop-yields and other sources of income. Underdeveloped markets and little financial involvement means that households must often find alternative social arrangements with which to smooth consumption. Typically, these risk sharing arrangements involve the exchanges of goods and services within a village or broader community. A great deal of work has gone into testing the "full risk-sharing hypothesis" under which, if communities are indeed hedging risk efficiently, idiosyncratic and independent movements in income should not correlate with fluctuations in consumption.[1] While this test is widely accepted now as the standard approach to test full insurance, it can only provide evidence for or against Pareto-optimal allocations; it fails, however, to provide an accurate alternative characterization below efficiency. Moreover, most empirical work on the subject has repeatedly rejected full risk sharing in a number of different contexts ranging from India to Tanzania and including Thailand, Peru, and many others.

In this paper I present a complementary interpretation of the risk-sharing test that provides a more detailed account of the type of behavior we might observe when we reject full risk sharing. In particular, I account for local network interactions by constraining Pareto-optimal allocations to a situation where exchanges are limited within a given social arrangement. As a theoretical contribution, I provide a formal description of the sharing rule between any pair of linked households strictly as a function of their network position. I structurally estimate the sharing rule against a unique data set of indigenous communities in the Bolivian Amazon, and I show that this description of sharing behavior does a remarkably good job at describing observed transfers across families. I argue that this framework provides a reinterpretation of the standard risk sharing results, predicting household heterogeneity in response to income shocks. I show that this network-based variation in consumption behavior is borne out in the data, and that it can be interpreted economically in terms of consumption volatitliy. Finally, I show the theory can account for the level of risk sharing observed in the data.

The current framework provides a general approach to modeling mutual insurance organized around local risk sharing groups. It generalizes recent work that has approached within-group insurance largely as an empirical question.[2]

[1]Cochrane (1991) best explained this approach as the cross sectional equivalent to Hall's permanent income hypothesis test, which regressed consumption growth rates over time on ex-ante variables (Hall, 1978). Indeed, under complete borrowing and lending opportunities consumption should not respond, over time, to forecastle shocks, just as it should not respond to idiosyncratic shocks across households under full insurance.

[2]Hayashi et al. (1996) consider whether extended families can be viewed as collective units

Rather than taking groups as separate, perfectly insured communities, I allow for a fully general network with interconnected sharing groups that are specific to each household. I argue that, in this environment, not defining the relevant local sharing group biases the results of classical tests of full insurance. More importantly, I show that controlling for this bias does not eliminate the correlation between consumption and income across households: the network structure generates underlying heterogeneity in sharing behavior, which implies that households' income affects consumption even after appropriately controlling for local aggregates.

I solve a constrained welfare problem in which transfers are limited along a given social structure. The restriction on exchanges means that whatever a household receives from its neighbors cannot be shared further down the network; that is, I assume that income can be split and shared only amongst immediate neighbors. This assumption is meant to capture the relatively low levels of intermediation, relative to direct exchanges, that occur in these types of subsistence economies, where mostly crops and other perishable goods are traded.[3] Alternatively, even when risk-sharing involves the transfer of cash as well, urgent liquidity needs often means households cannot immediately access distant funds that must first be intermediated by the network.[4] While I take the lack of any intermediation as a simplifying assumption, I also show that most results can be sustained when allowing for greater movement of funds.[5]

In this context of no intermediation I solve for the non-contingent (or fixed) sharing rules that maximize welfare. These type of sharing rules specify a fraction of each household's income consumed by each sharing partner, where this fraction is constant across all states of the world. In Section 1.2 I discuss the implications of these type of sharing rules in the context of a very simple example, and I show

sharing risk efficiently. Mazzocco and Saini (2012), for instance, argue that the relevant sharing group in India is the caste, rather than the village. Munshi and Rosenzweig (2009) also find that the caste is the relevant group to explain migration patterns in rural india. More recently Attanasio et al. (2015) test for efficient insurance within extended families in the U.S.

[3]Hooper (2011) for instance mentions it is quite rare to observe the same good exchange hands twice within Tsimane' communities. Similarly, Chiappori et al. (2013), Kinnan and Townsend (2012) and Udry (1994) document that an overwhelming share of the economies they study in Thailand and Nigeria are formed by crops, livestock and other perishable goods.

[4]You can think of this assumption as the complementary version of the assumptions driving the model of Ambrus et al. (2014). In that model, funds can travel indefinitely along the network, but each edge has some exogenous capacity constraint that limits the amount of funds it can intermediate. In this case, intermediation is ruled out, but the amount of funds along any given edge is endogenized. These type of limited interactions are also studied by Bourlés et al. (2015) in the context of altruism in networks, with very different implications.

[5]In Section 1.8.2 I show how to extend results to a general case with network intermediation. Notice that without some limit on how far funds can exchange hands along the network, the welfare problem is unconstrained and full insurance obtains as the unique outcome.

that they allow me to isolate network effects from income distributions in order to obtain simple predictions on sharing behavior across households.

Of course, various other explanations have been provided to account for the failure of full risk sharing in village economies. For instance, a number of papers have argued that incomplete information, limited commitment, heterogeneous preferences, or the presence of outside markets are all capable of generating inefficiencies in mutual insurance mechanisms.[6] In this context, it is worth asking why it makes sense to model social networks as a constraint on the classical welfare problem. Evidence suggests that individuals select into particular social arrangements precisely to mitigate informational frictions and to guarantee compliance, so that within these social arrangements mutual insurance mechanisms function rather well.[7] Moreover, there is strong evidence that informational frictions within these social spaces are relatively unimportant.[8] In this paper I abstain from considering the forces that shape particular social networks. Instead, I take them as given and study the type of efficient outcomes we expect within these restricted environments.

My first main result relates constrained-efficient transfers to a global measure of households' relative importance in the network. This measure reflects a household's direct and indirect interactions along the entire network. In this manner, the proposed measure bears some similarity to previous statistics — for instance Katz-Bonacich or PageRank — that capture higher-order dependencies as they feed back along a given network of connections. The particular flavor of this network measure has to do with the tradeoff faced by the planner between variance and covariance considerations. I find that, for any network, the constrained-efficient exchanges across any pair of households follows a simple relationship between the sender and receiver's network measures.

I show that this framework redefines the relationship between consumption and income in a setting of partial insurance on a network. This has important implications for the standard risk sharing tests. Specifically, the model provides heterogeneous predictions on household's response to income shocks that emerge from households' different positions in the network. This provides insight into the

[6]See for instance Incomplete information (Udry, 1994); Limited Commitment (Ligon et al. 2002); Heterogeneous preferences (Schulhofer-Wohl 2015, Mazzocco & Saini 2015); Outside Markets (Munshi & Rosenzweig 2014, Galeotti et al. 2015, Saidi 2015).

[7]Munshi (2014) describes the general tendency of households to arrange into particular social patterns that avoid certain commitment issues. On the theoretical side, a long list of papers have studied the type of networks that might emerge under limited commitment and similar frictions. Bramoullé & Kranton (2007), Jackson et al. (2012), and Ambrus et al. (2015) are just a few.

[8]In his work on Nigerian communities, for instance, Udry (1994) argues that loan arrangements are very informal, with no collateral, explicit interest rates or repayment dates, and that households know each other well. Hooper (2011) also finds similar evidence of strong informational flows in the Tsimane' networks of Bolivia that I study in this paper.

varied insurance possibilities of households when full risk sharing is rejected. I show that, in certain scenarios, these values can be mapped to important economic features, such as consumption volatility.

Having established the theoretical results, I examine a unique data set of about 250 households in 8 different indigenous villages in the Bolivian Amazon basin, during the years 2004 to 2009. This data is particularly well suited for my analysis as it provides information on the caloric exchanges across pairs of households over time.[9] This allows me to structurally estimate the model by fitting the theoretical relationship between network position and exchanges at the edge level. Moreover, compared to other models of risk sharing networks, I can estimate the model at a much finer level of variation (i.e. using edge-level data), and separately from aggregate considerations on consumption growth. I find that the empirical flows across connected dyads indeed respond to the network structure as the model prescribes. I show that once we account for this restriction on the planner's problem, we can effectively explain all the variation in consumption that correlates with households' income. Finally, I also test the model's implication on network-based heterogeneity of households' response to own income shocks, and I find that the data exhibits the same type of variation that the model prescribes. I do this by constructing a couple of tests that can be applied to many other data sets that include network and income data; it can be tested on a wide range of empirical settings. The results suggest that previous failures of full-risk-sharing tests are best understood by invoking restrictions on bilateral exchanges.

Related Literature

The distribution of uncertainty along social ties has, in the past several years, drawn a lot of interest from economists. Starting with Bramoullé and Kranton (2007a,b) and Bloch et al. (2008), a number of recent contributions — such as Jackson et al. (2012), Billand et al. (2012), Ali and Miller (2013a,b), and Ambrus et al. (2014) — have focused on enforcement concerns and the role of social capital in sustaining cooperative behavior. Most of these studies assume networks serve a dual role as both medium of exchange and social collateral, delivering efficient and stable structures for a set of exogenous, and fixed, bilateral transfers. In other words, most of these papers assume a sharing behavior and find networks that sustain it. Bramoullé and Kranton's (2007a,b) model, for instance, assumes that a connected component equally distributes its surplus independent of the social structure, so that inequality is ruled out. Billand et al. (2012) also assume a sharing behavior whereby high-income households transfer a fixed amount to low-

[9]Exchanges are measured in calories: food is the primary source of income and trade for subsistence economies like the Tsimane' communities studied in this paper.

income neighbors. I take a different view that abstracts from enforceability considerations altogether and instead provides an endogenous prediction of efficient transfers along a network. The focus on the distribution of surplus, and away from enforceability, appears most recently in work by Ambrus et al. (2015) that studies cross-group incentives for social investments. However, their concern has to do with network formation, so they also assume some exogenous split of surplus: bilateral exchanges are assumed to split the total surplus according to the Shapley value, which, in the particular setting they focus on, reduces to equal sharing. Perhaps closest in spirit is the work by Ambrus et al. (2014) that similarly refrains from assuming, a priori, the sharing pattern across connected pairs; they otherwise assume a distribution of "link values" that are perfectly substitutable with consumption, so that coalition-proof transfers are, again, ultimately determined from outside the model. The current paper refrains from engaging with these difficult strategic considerations, and instead solves for a simple constrained-efficient, network-based sharing rule that provides a number of testable implications.

On the empirical side, this paper joins the ranks of a long strand of research devoted to the estimation, and interpretation, of risk-sharing patterns in data. While newer data sets have begun to include social surveys that allow us to test network models directly, the empirical risk-sharing literature has a longer tradition, and one that, with occasional exceptions, has overwhelmingly insisted that communities operate below efficiency. The work of Mace (1991), Cochrane (1991), and Townsend (1994) provided the theoretical foundations for measuring correlations between household income and consumption, which, by now, has become the hallmark of all empirical tests on risk sharing. Since then, a healthy number of studies have sprung up to investigate one or another economic dimension of risk-sharing communities — from the impact of kinship ties on credit constraints in the Philippines (Kinnan and Townsend, 2012) to the decreased social mobility induced by local sharing along caste lines (Munshi and Rosenzweig, 2009). Whatever the particularities, all these studies perform the standard test of full risk-sharing and, together, deliver a cogent narrative that by and large strays away from efficiency. For instance, Ligon (1998) studies a private information alternative to the complete market model and rejects full insurance in rural south India. Fafchamps and Lund (2003) famously reject full insurance for Philippine communities and show that the extent of risk sharing is limited by the extent of interpersonal networks. Mazzocco and Saini (2012) reject full insurance for indian data at the village but not at the caste level, while Munshi and Rosenzweig (2009) reject efficiency at the caste level as well. On their study of investment decisions under exogenous income shocks to networks in rural Mexico, Angelucci et al. (2015) however find that they cannot reject full insurance within extended families. The list is long, though, and more often than not signals of full risk-sharing are absent from a wide range of settings.[10] More importantly, Saidi (2015) studies credit demand in the same Tsimane' indigenous communities that I study and also rejects full insurance.

Finally, the paper also relates to a number of studies that have sought to provide a direct explanation for the repeated failure of efficiency in data. For instance, Ligon et al. (2002) model optimal contracts under limited commitment. They estimate their model on three separate indian villages, and argue that this type of transaction cost accounts for the magnitude of departure from full insurance. More recently, a couple of studies have argued that heterogeneous risk preferences might force an interpretation of full risk-sharing test that is far too pessimistic. (Mazzocco and Saini (2012) or Schulhofer-Wohl (2011)). Schulhofer-Wohl (2012), for instance, argues that if households' variation in risk preferences are cyclical then not accounting for these explicitly introduces an omitted variable bias that pushes the coefficient of own income upwards, leading to false rejections of full insurance. Mazzocco and Saini (2012) have similarly developed empirical tests for heterogeneous preferences and provided a modified empirical procedure to test for efficiency. Most importantly, Fafchamps and Lund (2003) address the failure of efficient insurance in the data by invoking the role of gifts and remittances in risk-sharing and reject mutual insurance at the village level, suggesting instead that households receive transfers from a network of family and friends. Although they don't model network flows explicitly, their findings serve as the principal motivation for this paper.

The remainder of the paper is organized as follows. Section 1.2 goes over the standard test of full insurance and argues how the current setup affects this estimation procedure using a simple example. Section 1.3 introduces the theoretical framework, solves for the efficient sharing behavior, and provides implications for the test of full insurance. Section 1.4 provides background information about the data and summary statistics. In Section 1.5 I structurally estimate the model, and draw a number of testable implications for risk-sharing tests. Section 1.6 concludes.

[10] As Schulhofer-Wohl (2012) reminds us, full insurance has been rejected in data from the United States (Attanasio and Davis 1996; Cochrane 1991; Dynarski and Gruber 1997; Hayashi et al. 1996), Côte d'Ivoire (Deaton 1997), India (Munshi and Rosenzweig 2009; Townsend 1994), Nigeria (Udry 1994), and Thailand (Townsend 1995). Mace (1991) does not reject efficiency in U.S. data, but Nelson (1994) overturns this result.

1.2 Network Constrained Risk Sharing: A Simple Example

In this section I present the canonical model of full risk sharing and I describe the empirical approach that emerges from it to test full insurance from data. I then describe the main assumptions behind this paper and how it refines the concept of sharing groups. I use a very simple example to describe the sharing rules I obtain, and I explain what they claim about the distribution of insurance across the population. Finally, I present the implications of this model on the standard risk-sharing tests and I show that, 1) not defining the appropriate local sharing groups generates biased estimators, and 2) network asymmetries generate varying predictions on the impact of income shocks on consumption, which in turn provides a network story behind the rejection of full insurance. Section 1.3 then generalizes all these arguments to a full fledged model with an arbitrary network and a general income process for households.

1.2.1 Canonical Model

The classical risk sharing models of Cochrane (1991), Mace (1991), and Townsend (1994) solve for the ex-post pareto-optimal allocations by defining a planner problem as follows,

$$\underset{c_i(\omega)}{max} \sum_{\omega} \pi\left(\omega\right) \sum_{i} \eta_i \, u_i\left(c_i\left(\omega\right)\right)$$

where $\pi\left(\omega\right)$ represents the probability of state ω and where η_i represents $i's$ Pareto weight in the welfare function. This problem is subject to the constraint that total consumption not exceed total income in any state of the world, or that $\sum_{i} c_i\left(\omega\right) \leq \sum_{i} y_i\left(\omega\right)$ for all $\omega \in \Omega$, .[11] The first order conditions yield the well known full insurance equations known as Borch's rule. It states that the ratio of marginal utilities across any two agents is constant across states. Formally, we can solve for the problem above and, for any two households i and j, obtain the following expression,

$$\frac{u_i'\left(c_i\left(\omega\right)\right)}{uj'\left(c_j\left(\omega\right)\right)} = \frac{\eta_j}{\eta_i}\ , \quad for\ all\ \omega \in \Omega \tag{1.1}$$

This expression has been used to develop a popular test of full insurance. Indeed, equation (1.1) states that, under full risk sharing, consumption should not respond to idiosyncratic shocks after controlling for aggregate shocks. The following type

[11] Nondecreasing utility functions on consumption means that the constraint will hold with equality.

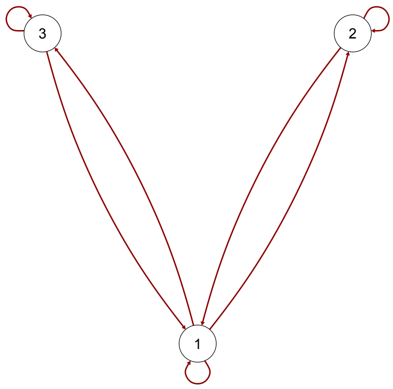

Figure 1.1: A Simple Risk Sharing Economy

of regressions,

$$log(c_{it}) = \alpha_i + \beta_1 log(y_{it}) + \beta_2 log(\bar{y}_t) + \epsilon_{it} \quad (1.2)$$

where $\bar{y}_t$ represents aggregate income, have been used to test for efficient outcomes, in which case $\beta_1 = 0$ and $\beta_2 = 1$. Time and again, β_1 is found to be positive and significant and β_2 below one. Unfortunately, not much can be learned from these results other than the existence, or not, of full insurance. The following approach attempts to give a more nuanced understanding of the type of sharing behavior that might be generating these estimates.

1.2.2 Overlapping Sharing Groups

An important feature of the classical risk sharing model above is that all households form part of the same risk sharing group. In this paper, I relax this assumption by considering the possibility that mutual insurance is local. This allows me to capture a number of relevant intermediation costs that might make it impossible to define a unique sharing group.[12] If these motives are strong, households can only access local risk sharing groups defined by their immediate neighbors (or trading partners).[13]

To fix ideas, consider the economy presented in Figure 1 where households 2 and 3 can only trade with household 1. For simplicity, imagine all households

[12]For instance, the Tsimane' communities that I study in this paper transfer highly perishable goods - mostly prepared food and game. In other contexts where risk sharing involves the transfer of cash as well, urgent liquidity needs means individuals cannot immediately access distant funds that must be intermediated.

[13]In Section 1.8.2 I show how to extend the model to all levels of intermediation. Notice that if intermediation is sufficiently high, all households access the same risk sharing group and efficiency obtains as above.

obtain a random income realization $y_i(\omega)$ that is i.i.d. from some distribution $F(\mu, \sigma^2)$. The lack of intermediation means households must access different, overlapping risk sharing groups — for instance, the risk sharing group of household 2 consists of households 2 and 1 only. Let α_{ij} represent the fraction of $j's$ income consumed by i. The situation of this economy can be written as follows,

$$\begin{array}{lclclcl} c_1(\omega) & = & \alpha_{11}y_1(\omega) & + & \alpha_{12}y_2(\omega) & + & \alpha_{13}y_3(\omega) \\ c_2(\omega) & = & \alpha_{21}y_1(\omega) & + & \alpha_{22}y_2(\omega) & & \\ c_3(\omega) & = & \alpha_{31}y_1(\omega) & + & & + & \alpha_{33}y_3(\omega) \end{array} \tag{1.3}$$

This formulation provides a very tractable way to define the risk sharing rule in this economy by expressing consumption explicitly in terms of bilateral transfers, α_{ij}. Notice that we can describe the canonical model above as the particular case where households 2 and 3 are able to trade with each other because they are directly connected.[14] In this case, all households clearly access the same risk sharing group and efficiency obtains.

In this paper I solve generically for the set of non-contingent sharing rules that maximize welfare in this setup with no intermediation. A non-contingent sharing rule means that the fraction α_{ij} of $j's$ income consumed by i is constant across all states ω.[15] In this paper I show how to solve analytically for this type of sharing rule for any given network. As an example, consider the economy of figure 1, and, in order to make the argument as simple as possible, set all Pareto weights η_i equal and set $\mu^2 = \sigma^2$. Applying the main theoretical result of this paper we obtain the following simple description between a household's consumption and the incomes of its relevant sharing group:

$$\begin{array}{lclclcl} c_1(\omega) & = & \frac{5}{21}y_1(\omega) & + & \frac{9}{21}y_2(\omega) & + & \frac{9}{21}y_3(\omega) \\ c_2(\omega) & = & \frac{8}{21}y_1(\omega) & + & \frac{12}{21}y_2(\omega) & & \\ c_3(\omega) & = & \frac{8}{21}y_1(\omega) & + & & + & \frac{12}{21}y_3(\omega) \end{array} \tag{1.4}$$

This situation is depicted in Figure 2. A great deal can be gleaned already from this very simple example. Notice that households 2 and 3 share a larger fraction of their income with 1 than 1 shares with them ($\frac{9}{21} > \frac{8}{21}$); still household $1's$ relevant sharing group is larger and as a result 1 consumes much less of its own income than 2 or 3 ($\frac{5}{21} < \frac{12}{21}$). Moreover, it is easy to show that consumption volatility associates positively with this value, so that household 1 (with a lower coefficient)

[14] Or, alternatively, by intermediating through household 1.

[15] In 1.8.1 I discuss the alternative assumption that sharing rules are contingent — i.e. $\alpha_{ij}(\omega)$. I show how it restricts the set of states for which the efficient condition (1.1) holds, and I demonstrate the inherent difficulty in isolating general network effects from particular income realizations for these type of contingent sharing rules. I also provide some empirical evidence that suggests informal exchanges in village economies might be closer to a fixed (or non-contingent rule) than to an extremely flexible sharing rule.

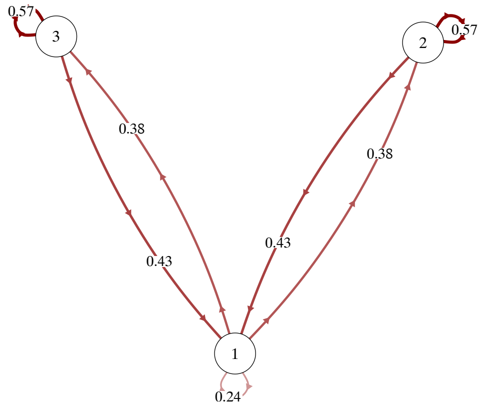

Figure 1.2: The Sharing Rule of a Simple Economy

obtains a less volatile consumption stream than households 2 and 3. This setup therefore provides network-based heterogeneity on households' response to own income shocks and relates it to the distribution of risk sharing opportunities.

1.2.3 Risk Sharing Regressions under Local Insurance

Ultimately, these predictions generate enough information on household consumption to provide reasonable explanations for the rejection of full insurance. I consider how this affects the empirical tests of risk sharing described in the previous section. Let us stick to the simple economy in Figure 1 and consider rewriting equations (1.3) in the form of the classical risk-sharing specification of equation (1.2) with a common aggregate income term,

$$
\begin{aligned}
c_{1t} &= (\alpha_{11} - \alpha_{12})\, y_{1t} + \alpha_{12}\bar{y}_t + \epsilon_{1t} \\
c_{2t} &= (\alpha_{22} - \alpha_{21})\, y_{2t} + \alpha_{21}\bar{y}_t + (\epsilon_{2t} - \alpha_{21} y_{3t}) \\
c_{3t} &= (\alpha_{33} - \alpha_{31})\, y_{3t} + \alpha_{31}\bar{y}_t + (\epsilon_{3t} - \alpha_{31} y_{2t})
\end{aligned}
$$

These equations reflect three important themes of this paper: 1) coefficients on own income are generically different from zero for all households — i.e. $\alpha_{ii} \neq \alpha_{ij}$ 2) these coefficients vary according to households' position in the network, and 3) imposing the common sharing group on all households generates biased estimates: notice the last two equations contain weighted incomes in the error term.[16] The classical risk sharing test in (1.2), pools these equations and obtains

[16] If incomes are positively correlated, then imposing a common aggregate variable biases estimates upwards. Schulhofer-Wolf similarly uncovers a bias in the classical risk-sharing specification that comes from heterogeneity in income preferences. Here, the heterogeneity is induced by positions in social structures. In any case, as I show below, we can adjust for the bias and still expect positive coefficients to income in this setup.

a unique estimate for β_1; given the previous discussion we expect this estimate to be different from zero and positive. In the first column of Table 1 I show the estimates for the simple example of Figure 1 for simulated data.[17] As expected, β_1 is statistically significant and close to 0.2, while the coefficient on the common aggregate income term, β_2 , is statistically lower than 1.

In order to isolate the network effect from the bias in β_1, consider estimating (1.2) with the relevant local sharing group instead. In this case, estimates are no longer biased, but we still obtain heterogeneous estimates, β_i, for the coefficients on own income. As a result, the risk sharing test still delivers positive estimates. To see this rewrite again equations (1.3) in the form of (1.2), but now we allow for household-specific aggregates, $\bar{y}_{it}$, that sum over the incomes of $i's$ local sharing group. In this case we have

$$\begin{aligned} c_{1t} &= (\alpha_{11} - \alpha_{12})\, y_{1t} + \alpha_{12}\bar{y}_{1t} + \epsilon_{1t} \\ c_{2t} &= (\alpha_{22} - \alpha_{21})\, y_{2t} + \alpha_{21}\bar{y}_{2t} + \epsilon_{2t} \\ c_{3t} &= (\alpha_{33} - \alpha_{31})\, y_{3t} + \alpha_{31}\bar{y}_{3t} + \epsilon_{3t} \end{aligned}$$

Because aggregates are now household-specific, the additional terms in the error term disappear and we obtain unbiased estimators. Notice, however, that coefficients to own income are different from zero so long as $\alpha_{ii} - \alpha_{ij} \neq 0$. This implies that the pooled regression will again deliver positive coefficient, β_1, even with the appropriate local aggregates. I present the results to this local sharing group version of equation (1.2) in the right column of Table 1. Again, the coefficient to income is positive, as expected, although estimates decrease by one order of magnitude.

Real world social structures are usually far more complicated than these simple examples. Figure 4 plots two of the networks I build from data in one of eight indigenous communities I study in this paper; the one on the left is built from kinship data and the one on the right on trade data.[18] These networks are orders of magnitude more complicated. Still, I show that the arguments above can be extended generically for any network and general income process across households. Moreover, this unique data set contains information on the transfer of food across households over time, so I am able to structurally estimate the endogenous, network-based sharing rule that I derive in this paper. I find it does a remarkably good job at describing the patterns of exchange across households in these subsistence economies.

[17] I simulate log-normal income data for all three households with $t = 100,000$ and I obtain household consumption as indicated by the sharing rule in (1.4). I then run the standard risk-sharing regression on logged data, controlling for household fixed effects.

[18] Refer to section X for a detailed description of the types of networks constructed and a discussion on the relative merits of each. Refer to Figures 1 and 2 in the appendix for a visual plot of all villages for each of the network types.

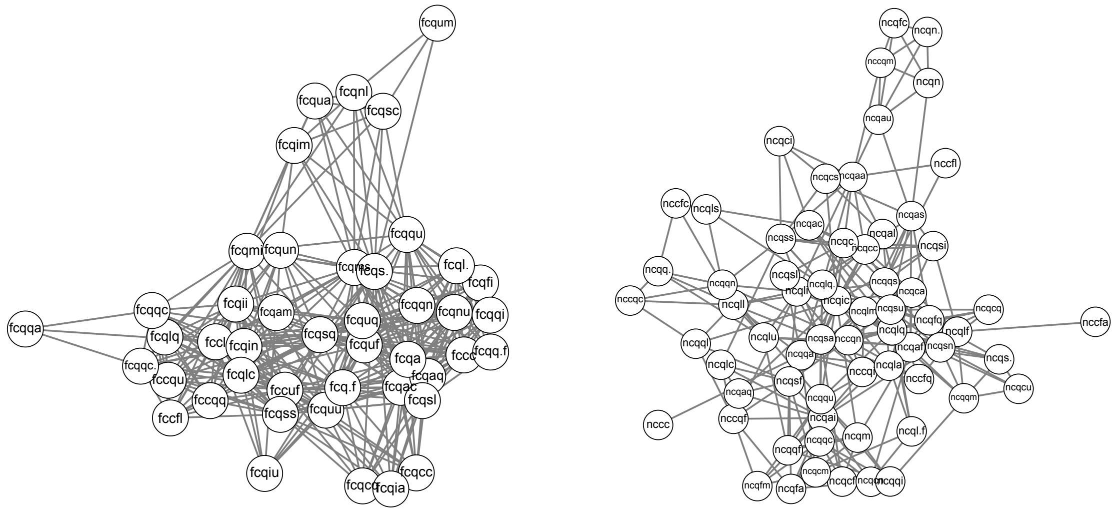

Figure 1.3: Two Tsimane' Villages: (a) Kinship Network (b) Trade Network

1.3 The Model

I study an economy in which households face uncertainty about their income realizations, but may redistribute incomes through a given network of social connections. I characterize efficient transfers as a function of households' position in the network when the movement of funds is restricted. In section 3.1 I describe the theoretical setup. In section 3.2 I solve for the constrained-efficient set of transfers and describe how they relate to the underlying network. In section 3.3 I provide certain properties of the sharing rule and describe its behavior more closely for some simple structures.

1.3.1 Setup

Consider a population of size N arranged in a network $L = (V, E)$, consisting of a set V of households (vertices) and a set E of pairs of elements of V that represent links (edges) across these households. I assume the network is undirected, so that the pair of vertices in E is unordered. It is also useful to define an alternative characterization of this social structure by an adjacency matrix G, where $g_{ij} = 1$ if and only if $\{i, j\}, \{j, i\} \in E$. Each connection can represent a friendship, kinship relation, or other type of social connection between the two parties involved. We will refer to i's neighborhood as the subset of N defined by $N_i = \{j \in N \mid e_{ij} \in E\}$. The degree of a vertex i measures the number of connections of i and is defined as the cardinality of N_i.

All households face risky endowments. Denote the vector of random endowments by $\mathbf{y} = (y_i)_{i \in N}$, drawn from some joint distribution F with mean μ and variance σ^2. I assume a common covariance between the incomes of any two agents and denote it by $\rho \neq 0$. I assume throughout that $|\rho| < \sigma^2$ so that incomes are not perfectly correlated.

Households share incomes along a social network, so that consumption levels will differ, in general, from their income realizations. Incomes can only be exchanged once, so that households consume incomes from immediate neighbors.[19] The shares of neighboring endowments consumed by a given household are defined ex-ante and are non-contingent. Together this implies that a household's consumption in state ω can be defined as a linear combination of neighbors' incomes as,

$$c_i(\omega) = \sum_j g_{ij} \alpha_{ij} y_j(\omega) \tag{1.5}$$

where α_{ij} represents the share of $j's$ endowment that is consumed by i. We will also define $\alpha_i = (\alpha_{ij})_{j \in N_i}$ as the vector of $i's$ incoming shares. By defining the "sharing matrix" $\mathbf{A}$ as $\mathbf{A}_{ij} = g_{ij}\alpha_{ij}$, we can express equation (1.5) in matrix form in the following way, $\mathbf{c} = \mathbf{Ay}$, where I drop the explicit dependency on ω from now on for notational convenience. Of course, the elements of $\mathbf{A}$ represent percentage claims on neighboring incomes and must therefore satisfy a feasibility condition that all claims on a given endowment sum to 1, which can be expressed as $\mathbf{1} = \mathbf{A}'\mathbf{1}$. Finally, I assume all households have quadratic utility functions:

$$u(c_i) = c_i - \frac{1}{2}\gamma c_i^2$$

where γ is the common coefficient of risk aversion.

I now define the planner problem and provide a short discussion on the particular form of the objective function and the constraints.

Definition 1. *The planner problem is defined as,*

$$\max_{\{\alpha_{ij}\}_{ij}} \mathbb{E}\sum_i u(c_i) = \min_{\{\alpha_{ij}\}_{ij}} \sum_i \left(\mu^2 \left(\sum_j g_{ij}\alpha_{ij} \right)^2 + \sigma^2 \sum_j g_{ij}\alpha_{ij}^2 + \rho \sum_{k \neq j} g_{ik} g_{ij} \alpha_{ij} \alpha_{ik} \right) \tag{1.6}$$

subject to $\alpha_{ij} \geq 0$ *for all* $i, j \in N$ *and that* $\sum_i \alpha_{ij} = 1$ *for all* $j \in N$

[19] In Section 1.8.2 I show how this assumption can be relaxed to allow for network intermediation.

The form of equation (1.6) exploits the linear mean-variance tradeoff of expected utility: the first term in brackets corresponds to the squared mean of consumption, while the next two terms correspond to the variance of consumption.[20] The constraints on the planner problem reflect the standard feasibility conditions that shares are positive and sum to one. Finally, since the sum is convex in shares and the constraint set is linear, the maximization is a convex program and the first order conditions completely characterize the optimal solution. In the next section I define these optimality conditions and explore the type of network interactions that are contained in them. I then provide the general solution for any network $\mathbf{G}$ and a general class of distributions F.

1.3.2 Constrained-Efficient Network Flows

Having defined the economy and the welfare problem in the previous section, we are now ready to obtain a description of the sharing rule for any given network. To do this in a way that clarifies the type of network interactions that emerge, I first analyze the planner's optimality condition in some detail. The first order conditions of (1.6) defines the share α_{ij} that i receives from j (for each pair $i, j \in N$) as,

$$\alpha^{\star}_{ij} = g_{ij}(\Lambda_j - \Psi \sum_k g_{ik}\alpha^{\star}_{ik}) \quad for\ all\ i, j \in N \tag{1.7}$$

where $\Lambda_j = \frac{\lambda_j}{2(\sigma^2-\rho)}$, and $\lambda_j > 0$ is the multiplier for $j's$ constraint, and where $\Psi = \frac{\mu^2+\rho}{\sigma^2-\rho} > 0$ captures the shape of the income distribution. It is worthwhile to examine equation (1.7) in some detail. First of all, notice that if i and j are not connected, $g_{ij} = 0$ and i consumes none of $j's$ income. Instead, if $g_{ij} = 1$ then the fraction of $j's$ income consumed by i depends on two terms. The first term, Λ_j, captures the relationship among all of $j's$ shares, as governed by $j's$ constraint, $\sum_i \alpha_{ij} = 1$. It enters positively because a drop in one of $j's$ shares (holding everything else constant) would increase Λ_j, and thus force all of $j's$ shares up to meet the constraint. As such, this term effectively connects all of the first order conditions pertaining to j. For instance, if no other effect existed, Λ_j would set all of $j's$ shares equal to each other. However, in most situations $j's$ shares are not

[20]Notice we can write the planner problem as, $\mathbb{E}u\left(c_i\right) = \mathbb{E}\sum_i \left(c_i\right) - \frac{1}{2}\gamma \sum_i \left(\mathbb{E}\left(c_i\right)^2 + var\left(c_i\right)\right)$ and the first term drops out because aggregate consumption must equal aggregate income by the constraints— i.e. $\sum_i c_i = \sum_i y_i$. Therefore, the planner problem reduces to minimizing $\sum_i \left(\mathbb{E}\left(c_i\right)^2 + var\left(c_i\right)\right)$ which corresponds to the expression in Definition 1. In the appendix I show this problem corresponds to the minimization of expected inequality and I relate it to other similar results for CARA utility in Ambrus et al. (2015).

equal, given the second term in (1.7). This second term determines how all shares received by i affect α_{ij} — the more i receives from some neighbor k the less it receives from j (and vice versa), where the constant Ψ mediates the strength of this response. The value of Ψ captures the relative variance and covariance considerations of the planner: as covariance effects increase (and Ψ increases), $j's$ shares respond more to the value of other shares.[21] To sum up, the share of $j's$ income consumed by i responds, on the one hand, to all shares coming from j (through Λ_j) and, on the other hand, to all shares going to i (through Ψ).

More generally, the second term in (1.7) defines a recursive relationship for α_{ij}. Cutting through the recursivity allows us to reframe the optimality condition (1.7) in terms of the constraints Λ_k as follows[22],

$$\alpha^{\star}_{ij} = g_{ij}\left(\Lambda_j - \frac{\Psi}{1+\Psi d_i}\sum_k g_{ik}\Lambda_k\right) \quad for\ all\ i,j \in N \tag{1.8}$$

Given the arguments above, Λ_k connects all of household $k's$ optimality conditions via $k's$ constraint (if α_{ik} decreases, then α_{jk} increases for all j connected to k, holding everything else constant). Therefore, equation (1.8) expresses α_{ij} not as a function of all shares that i receives (as in (1.7)), but instead as a function of the full set of interactions for each of $i's$ partners. That is, it contains all the indirect interactions that affect α_{ij}. The shape of this expression clarifies the form in which indirect effects — captured by the values of $\mathbf{\Lambda} = (\Lambda_1, \Lambda_2, \ldots, \Lambda_n)$ — feed into the optimality condition of the planner. The challenge consists in determining the exact shape of these indirect effects as a function of the network. It turns out we can obtain a recursive formulation for these constants in the spirit of other well known vertex similarity measures such as Katz-Bonacich or PageRank.[23] This is the content of Proposition 1.3.1.

Proposition 1.3.1. *The constrained-efficient risk sharing agreement for any network defined by G is characterized by a set of transfers given by,*

$$\alpha_{ij} = g_{ij}\left(M_j\left(\Psi,\mathbf{G}\right) - \frac{\Psi}{1+\Psi d_i}\sum_k g_{ik}M_k\left(\Psi,\mathbf{G}\right)\right) \tag{1.9}$$

where $M_i\left(\Psi,\mathbf{G}\right)$ corresponds to i's Weighted Even-Path Centrality (WEPC) de-

[21]Conversely, when controlling the variance becomes more important to the planner than covariance effects, the term Ψ decreases and the planner sets all of $j's$ shares much closer to each other, as demanded by the Λ_j term in (1.7)

[22]This is shown in the proof of Proposition 1.3.1.

[23]See, for instance, Leich, Holme and Newman (2006) for a theoretical account of vertex similarity in networks.

fined recursively as,

$$M_i(\Psi, \mathbf{G}) = \frac{1}{d_i}\left(1 + \sum_{l,k} g_{ik} g_{kl} \frac{\Psi}{1+\Psi d_k} M_l(\Psi, \mathbf{G})\right) \tag{1.10}$$

Proof. See Appendix. □

Proposition 1.3.1 characterizes the full set of shares, $\mathbf{A}(\Psi, \mathbf{G})$, that defines the interior solution to the planner problem for any given network. As discussed above, the solution depends on the parameter Ψ and on the positions of each household in the network. The form in which the network defines the efficient sharing rule has to do with interactions among neighbors of neighbors (in other words, among households located two links apart). To gain some intuition, recall the network interaction terms in equation (1.7): the shares going to household i are substitutes. This implies that households two links apart (with a common neighbor, say, i) interact directly as shown in equation (1.7). But indirect effects play a crucial role here as well. To see this, notice that these two households not only interact through their transfer to i, but also exchange resources with other partners, and these other relations affect what i receives from them, given their constraints that $\sum_i \alpha_{ij} = 1$. This is the main message behind equation (1.8). As a result, each household connected to i not only interacts directly with each other as in (1.7), but they also interact indirectly with others' sharing partners. The recursive definition of network centrality in Proposition 1.3.1 reflects these arguments: it says that $i's$ centrality depends on the centralities of $i's$ neighbors' neighbors (i.e. those households two links apart). Finally, Proposition 1.3.1 says that the sharing rule between any two households depends positively on the sender's measure, and negatively on the sum of measures of the receiver's neighborhood. It is sometimes helpful to think of this tradeoff as capturing the extent to which the sender's indirect interactions in the network cannot be accessed by any other of the receiver's partners.

The previous discussion argues that households at distance two interact directly, but also that households at distance four, six, eight etc.. interact indirectly. With this in mind, I show that the centrality measure captures all these direct and indirect effects across the network. The crucial element in this setting (following the previous arguments) consists of a network statistic that aggregates all *even-length paths* for every household, weighted in some particular way. In other words, Proposition 1.3.2 solves through the recursive definition of Proposition 1.3.1 in order to provide a reinterpretation of network centrality that clarifies the previous discussion.

Proposition 1.3.2. *Household i's WEPC measure corresponds to a weighted sum of all even-length paths starting from i,*

$$M_i(\Psi, \mathbf{G}) = \frac{1}{d_i} + \sum_{q \in \mathbb{N}} \sum_{j \in N} \sum_{\pi_{ij} \in \Pi_{ij}^{2q}} W(\pi_{ij}) \tag{1.11}$$

where the weight of each path $\pi_{ij}^q = (i_0, i_1, i_2, \ldots i_q)$ of length q from i to j (i.e. $i_0 = i$ and $i_q = j$) is given by,

$$W(\pi_{ij}) = \frac{1}{d_{i_0}} \frac{\Psi}{1 + \Psi d_{i_1}} \frac{1}{d_{i_2}} \frac{\Psi}{1 + \Psi d_{i_3}} \cdots \frac{1}{d_{i_q}} \tag{1.12}$$

Proof. See Appendix. □

Recursive measures like the one in equation (1.10) are common in the networks literature. These can be usually expressed similarly as the sum of all weighted paths starting from some household. I refer the reader to Section 1.8.3 for a more detailed and technical discussion of these graph measures and how the WEPC relates to them. In the context of the present discussion, however, it is interesting to note that contrary to other similar measures that weight all paths of a certain length equally, the current measure elicits path-specific weights, as described in equation (1.12). These weights reveal once more how the constraints (which weight all shares evenly as one over the degree, $\frac{1}{d_i}$) are used to connect long chains of interactions, in which two households interact via a common neighbor i through the term $\frac{\Psi}{1+\Psi d_i}$, as shown in equation (1.8).

In the next section I work through some properties of the sharing rule, and I argue the type of effects we expect in the estimation of standard tests for full insurance for some simple networks. In a way, the reader might want to think of this section as a more complete version of section 2, now that the sharing rule has been defined.

1.3.3 Comparative Statics and Implications for the Risk Sharing Test

Section 2 argues, like other recent papers, the importance of specifying the appropriate risk sharing group for each household when running the risk sharing tests. However, it also makes the case that coefficients to own income can generally be different from zero and that these coefficients vary across households, so that full risk sharing is rejected even at the appropriate sharing level; If networks are sufficiently symmetric, though, then households pass the Townsend (1994) test whenever local sharing groups are correctly controlled for. In this section I complement these arguments by working through two simple network structures using the sharing rule developed in the previous section.

First I show how, under symmetric structures, the sharing rule indeed boils down to a simple intuitive sharing behavior that predicts the standard efficiency results, when controlling for aggregate income of the relevant, household-specific sharing group. The most symmetric structure is the regular network — this is a network where every household is connected to k identical households, so all households are in identical positions. Following the arguments of section 2, we write household consumption in the form of risk sharing test

$$c_{it} = (\alpha_{ii} - \alpha_{ij}) y_{it} + \alpha_{ij} \bar{y}_{it} + \epsilon_{it}$$

where $\bar{y}_{it}$ represents aggregate income of $i's$ local sharing group (i.e. $i's$ neighborhood). It is clear that the coefficient to $i's$ own income is zero (as in the Townsend tests) if $\alpha_{ii} = \alpha_{ij}$. This is true of regular networks.

Proposition 1.3.3. *The constrained-efficient sharing arrangement for any regular network with a common degree equal to k corresponds to the equal sharing rule defined as,*

$$\alpha^{\star}_{ij} = g_{ij} \frac{1}{k} \tag{1.13}$$

As a result, the first-best allocation is obtained for complete networks.

Proof. See Appendix. □

Of course the regular network is a very rare and extreme form of symmetry. In reality, social networks are far less structured and will therefore predict widely different transfers for different households. In these other cases, we expect instead that $\alpha_{ii} \neq \alpha_{ij}$ and therefore that the coefficients to own income will be positive and different across households. To take the most extreme example, consider the star network. In this network, one household is connected to all other households, that are otherwise not connected to anyone. In this case, I show the sharing rule simplifies to a simple expression relating to the size of the network and the level of connectivity of each household.

Proposition 1.3.4. *The constrained-efficient sharing arrangement for a star network of size n is given by,*

$$\alpha^{\star}_{ij} = g_{ij} \left(\frac{1}{1 + \left(\frac{n}{2} + 1\right) \Psi} \left(\frac{1}{d_j} + \Psi \right) \right)$$

for all $j \neq i$.

Proof. See Appendix. □

It is easy to see from this expression that flows towards the center consumes a smaller fraction of own income than other households. In fact, the star represents an extreme situation in which the center can very quickly be left to consume none of its own income. [24] Because the center mediates among a great number of other households, it benefits from diversification so long as incomes are sufficiently uncorrelated. Therefore, high centrality translates to lower consumption variance, and therefore the distribution of coefficients to own income across the population obtains a particular interpretation in terms of sharing opportunities — something that is not available in the standard risk sharing tests of Townsend (1994).

Finally, the value of Ψ can also provide drastically different predictions on the type of sharing behavior. Notice that if Ψ tends to zero — for instance for i.i.d. variables ($\rho = 0$) with small ratio $\frac{\mu^2}{\sigma^2}$ — then the optimality conditions imply that a household shares equally with all its neighbors (i.e. that $\alpha_{ij} = g_{ij}\frac{1}{d_j}$).[25] This is not surprising: extremely low values of Ψ represent situations where only minimizing aggregate volatility of uncorrelated earnings is important; this is accomplished by maintaining all shares as equal as possible.[26] This result poses a challenge in identifying the sharing behavior from data. Indeed if the true value of Ψ is low, it might be impossible to distinguish between sharing behavior described generally in Proposition 1.3.1 and a simpler, heuristic behavior such as equal sharing; both prescriptions should perform well as statistical models. In Section 3.4 I explore an alternative theoretical prediction of Proposition 1 in order to strengthen the belief that (1.9), and not (1.13), appropriately describes the sharing patterns of the Tsimane' communities.

On the other hand, as incomes correlate strongly across households the constrained-efficient sharing rule trades off diversification opportunities. In these situations, where Ψ is greater than zero (possibly much greater), the planner's previous tendency to equate shares will lead to a large loss of surplus as strong correlation effects hike up consumption volatility. Now, the incentives move in the opposite direction and reigning in covariances is the primary concern; this is done by keeping all of $i's$ incoming shares as different as possible. In particular, I show below

[24] More precisely, for $n > 4$ the center of the star consumes none of its own income if $\Psi \geq \frac{2}{n(n-4)}$ (for $n < 4$ interior solutions exist for all values of Ψ). This implies that as n increases, the space of parameters that guarantees an interior solution decreases.

[25] To see this notice that equation 1.7 in this case implies $\alpha_{ij} = \Lambda_j$ for all j, meaning that any two households connected to j receive the same fraction of $j's$ endowment. As a result, it must be that $\alpha_{ij} = \frac{1}{d_j}$. This represents equal-sharing.

[26] This is a classical iso-parametric problem of minimizing squares. Notice that for $\Psi \to 0$, the parameter σ^2 dominates over μ^2 and ρ so the planner problem is reduced to minimizing $\sum_{i,j} g_{ij}\alpha_{ij}^2$.

that at this other extreme— that is as Ψ tends to infinity— a household's net shares with each of its partners tend to zero.

Proposition 1.3.5. *As Ψ grows, the net exchange for any two households falls. In the limit, we have that*

$$\lim_{\Psi \to \infty} |\alpha_{ij} - \alpha_{ji}| = 0$$

for all $i \neq j$

Proof. See Appendix. □

Large values of Ψ correspond to situations of negligible net bilateral exchanges. This implies that household's in-shares correspond to its out-shares, that all households consume a convex combination of their partners' incomes.

1.4 Background and Data

In this paper I develop an alternative empirical specification to test risk-haring behavior by fitting (network) constrained-efficient exchanges across pairs of households, and recovering unexplained dependency between income and consumption. To do this I use panel data collected by a team of anthropologists from the years 2004 to 2009 in the small-scale, hunter-gatherer economies of the Tsimane' in the Bolivian Amazon. Before describing the data set in more detail I provide a quick description of the Tsimane' social structure, their economy, and general patterns of exchange (see Hooper (2011) for a much more thorough investigation into the economic life-cycle of the Tsimane').

1.4.1 The Tsimane' Indigenous Communities

The Tsimane' are an indigenous population of about 10 to 20, 000 individuals, residing in the Beni Department in lowland Bolivia. Tsimane' settlements are located primarily along the Maniqui and Quiquibey rivers, their tributaries and nearby forests. The Tsimane' organize primarily around a subsistence economy based on hunting, fishing, and slash-and-burn agricultural production of rice, sweet manioc (or yucca), plantain, and maize. According to Hooper (2011), most families maintain between 1 and 6 fields at one time (an average of 2.9 fields per family) that range in size from 0.1 to 2 hectares (an average of 0.56 hectares per field). While some of this production —- primarily, though not exclusively, rice — is sold to outside nearby markets in San Borja, still, around 95% of Tsimane' subsistence consumption rests on own production and exchanges across families.[27]

[27] See Martin, Melanie et al. (2012)

The Tsimane' social structure is primarily kin-oriented. Closely related nuclear families often reside together in small residential clusters, engaged in high levels of cooperative labor, common and shared meals, and other forms of resource pooling; bilateral exchanges of food across households account for the majority of this form of risk sharing.

The important gains that come from sharing uneven returns to productive effort are not foreign to the Tsimane'. The exchange of food across households forms a significant chunk of economic activity. In previous work on the Tsimane', Hooper (2011) shows clear evidence of reciprocity between families, and across types of goods, suggesting an interest in both attenuating risk and exploiting gains from specialization (see tables 5.1, 5.2 and 5.3 in Hooper (2011)). Around, 99% of the Tsimane' population engage in some form of food sharing at some point in the sample, and only 3.2% form separate trading groups of less than four households. From total production, an average of 5% is sold to outside and the rest is either consumed by the producing family, or exchanged with others. On average, 66% of a household's production is exchanged with other families, while 31% of a family's consumption consists of food received from other households. Genetic relatedness and the age of the household head interact as decisive attributes in determining the patterns of caloric exchanges. Hooper finds that while age alone does not seem to explain transfers, it nonetheless exhibits strong patterns of exchange between closely related families, not between unrelated ones. In terms of relatedness, it alone forms a very good predictor of food sharing. As an example, for two families with 40-year-old parents and zero net meat production, for example, the effect of a 0.1 increase in relatedness on the gross number of meat calories shared from one family to the other is 33.3 calories per day (Hooper, 2011).

1.4.2 The Data

The data comes from field work by a group of anthropologists at the Tsimane Health and Life History Project.[28] A series of field interviews were conducted from the years 2004 to 2009 on 250 families (1245 individuals) residing in 11 different Tsimane' villages. The villages are grouped into four separate regions: "downstream", "forest", "tributary", and "ton'tumsi". Figure 2.1 in Hooper provides a breakdown of the different sample periods and sizes. Each family was interviewed an average of 45.5 times (SD = 20.4), yielding a mean of 92.8 sample days per individual (SD= 40.0). Households were surveyed on average twice per week.

The surveys collected information on how many hours each family member spent laboring in subsistence activities during the preceding two days. These in-

[28]Paul Hooper (Emory), Hillard Kaplan (UNM), and Michael Gurven (UCSB)

clude hunting, fishing, and agricultural work. Quantities of edible products were recorded, and, for each product, interviewees were asked which members of the nuclear family, and which other community members had consumed portions of the product in prepared meals, or had received portions as raw gifts. Families were also asked whether they had received any gifts of food from other households.

For each product, the raw mass in kilograms was calculated from reported quantities based on mean mass measurements derived from field guides and previous research with the Tsimane and other South American foragers. A product's total caloric value was based on estimates of mean dietary calories (assimilated by a human consumer after processing) per kilogram (Hooper, 2011).

In all sample communities, a detailed census was established that provided information on each individual's sex, birth year, and biological parents and grandparents.[29] Consanguineous and affinal relationships between individuals residing in the same community were calculated on the basis of shared genetic ancestry and marriage. Distance between households was constructed, when possible, from GPS data. (See Hooper, 2011) for a detailed account of all data collection procedures).

Together, this information constitutes an unbalanced panel of pairwise calorie exchanges at a frequency that is perhaps too high for this type of analysis. I therefore aggregate the data at a quarterly level. After discarding some pathological cases, I am left with 243 households in 8 different communities, sampled irregularly over 20 quarters, out of which an average household is sampled for 4.5 quarters (not always consecutive).[30]

1.4.3 Constructing Networks

As in most empirical studies of social networks, I confront the usual questions regarding how to define the appropriate underlying (and unobserved) social structures. As explained above, the unit of analysis is considered to be a nuclear family; I also refer to these as households. Although finer, within-household data on exchanges is available, all evidence suggests that these intra-family flows operate efficiently as completely connected (or in any case very dense) networks, so that

[29] Where incomplete, these census data were supplemented with data from demographic interviews described in Gurven et al. (2007). Adult parents and their co-resident dependents (i.e. offspring and adopted dependents) were classified together as nuclear families. Body mass in kilograms, assessed using an electronic standing scale, was available from yearly physical exams conducted by Bolivian physicians and research assistants for 1198 individuals in the sample (96%).

[30] I discard 5 households for which there is no reliable data. They appear to produce nothing and receive nothing throughout the entire sample. I drop another two households for which no reliable data on hours worked exists.

for the purposes of this analysis they are best considered as distinct economic units.[31]

I test the model on three types of networks; each is accompanied with its own set of problems and advantages. The Trade Network establishes a link between two households if ever the two engage in any food sharing. This method of constructing links by "revealed preferences" of course fails to account for additional connections that could exist, but are otherwise not used. One could worry about endogeneity issues coming from this type of network. There are two things to say on this matter: First, the constrained-efficient exchanges I solve for are interior solutions to the planner problem, so the model speaks only to situations where all available links are utilized for some amount of food sharing, no matter how small. In other words, the model makes no predictions about which connections should be used, so taking observed trade as a link is not a huge problem. Secondly, It seems reasonable to assume that if two households share no calories throughout the entire sample then some social cost exists that impedes said relationship.

It is customary in these communities for households to split upon marriage while remaining in the same village.[32] In order to account for this type of network modifications I also construct a dynamic version of the Trade Network. This network constructs links if, at every quarter, households are observed exchanging calories. For this particular network, then, the various centralities computed, and therefore the theoretical predictions on bilateral exchanges, are time dependent. While it might be unreasonable to assume in general that underlying social connections should change often, I show in the next section that in fact these networks show remarkable persistence over time, and in particular that the vast majority of links persist once they appear.

Finally, networks are also built using kinship data. To refrain from the endogeneity issues above I make no judgement on the "appropriate" level of kinship that determines the presence of a link. Instead, I construct links between households that share any level of genetic relatedness. This method poses its own set of concerns, not least of which has to do with missing genetic information for a number of households; this restricts the network artificially. Moreover, while kinship appears repeatedly in sociological work as a crucial determinant of social ties for a wide array of contexts, the Tsimane's population exhibit a disproportionately large share of endogamy[33]. This implies, as I describe in the next section, that while kinship networks might very well capture the appropriate dimension upon which social interaction develops, the level of connectivity might exceed any reasonable, underlying structure that determines insurance.

[31] See for instance, Hooper (2011).

[32] Hooper (2011) documents the creation and destruction of households in the Tsimane' context somewhere between 5 to 10% of households.

[33] See Hooper (2011)

1.4.4 Descriptive and Network Statistics

Table 1.1 provides some descriptive statistics for various demographic and economic household attributes. Some of these represent cross sectional variability in different measures of household wealth, specified in Bolivianos (the national currency) and split into animal (livestock), traditional (non-mechanized productive tools) and modern (technological goods). These are used at times, together with demographic variables such as family size, age and marital status, to control for household-specific economic attributes.[34] More importantly, the average income and consumption variables represent the caloric production and flow data that are used extensively to test the model predictions on bilateral exchanges. These variables are longitudinal, specifying, for every household and every available date, the hours spent in each productive activity, the calories obtained as production, and the ensuing flows of those calories to nearby families. The data set also provides information on hours spent in three general productive activities: Agriculture, Fishing, and Hunting; Leisure is therefore defined as the number of hours in the past two days not spent in any of these activities.

Next I present network statistics for each of the kinship and trade networks that appear visually in figures 1 and 2. I show this information per village since I consider the villages as eight separate network structures. As described above, the kinship network is far denser, exhibiting larger cluster and closeness measures and much lower diameters across most villages. By way of comparing these two networks more rigorously, I also calculate the share of edges of each network that are observed in the other one. I find that about 71% of the connections observed in the data occur between households with some degree of genetic relatedness, while only 35% of households with genetic affiliation actually share food. [35] Finally, I present measures of persistence for the Updated Trade network to show that while it might be worthwhile to allow for certain network changes that come from the movement across villages or the creation of new families, networks are fairly stable over time.

In tables 1.5 to 1.8 I provide some preliminary evidence to support the assumption of no intermediation. Recall that this assumption implies that household consumption is a linear function of neighbors' incomes. As a result, two

[34]I mostly use household fixed effects to control for time-invariant household specific attributes such as these. However, when estimating the model's prediction on bilateral flows, edge-specific intercepts are unfeasible due to limited observations. In these cases I use a battery of controls such as those in Table 1, and others.

[35]I also perform a second measure of network comparison known as the Hamming distance, which measures the number of edges that need to be substituted to turn one network into another. We can see in Table 1.4that across most villages, we need to substitute about a third of all available dyads to move from the trade to the kinship network— in a couple of villages about half of that share is required.

households that share a common neighbor will both consume a fraction of that income, and their consumption will be correlated. Households that are farther from each other, however, will not share any neighbors and their consumption will only co-move as per the underlying covariance across incomes. I therefore estimate consumption of each household against the aggregate consumption of those other households with which it shares a common neighbor (i.e. household in the set N_i^2), and against the aggregate consumption of households with which it does not share a common neighbor (i.e. households not in the set N_i^2). The results indicate that households farther away cannot explain consumption, once we control for aggregate income fluctuations.

1.5 Empirical Analysis

The theory above provides a number of predictions that can be tested directly against data on bilateral exchanges within networks. In this section I first run the standard test of full risk sharing and I find that full insurance is rejected in the Tsimane' data set. I then structurally estimate the sharing rule prescribed by the theory as given in equation (1.9); I find that the constrained-efficient prediction above appropriately describe the type of bilateral exchanges we observe across households. I also show that the model can retrieve the observed deviation from full insurance by estimating the risk sharing test on predicted consumption data. Finally, I test the model's implications on households' heterogeneous response to own income fluctuations. I find that the variation in household's coefficients to income follows the general pattern described by the model.

1.5.1 Test of Full Risk Sharing

In order to assess how well the data conforms with the model, I explore a number of distinct predictions from the theory. Before I do this, however, I first perform the classical risk sharing test of Mace (1991), Cochrane (1991), and Townsend (1994), by running regressions of the form

$$\Delta log\,(c_{it}) = \beta_1 \Delta log\,(y_{it}) + \beta_2 \Delta log(X_{it}) + \tau_{vt} + \epsilon_{it} \tag{1.14}$$

where $\Delta log\,(c_{it})$ and $\Delta log\,(y_{it})$ stand for household consumption and income growth rates respectively, and τ_{vt} represents village-time fixed effects that capture uninsurable aggregate shocks that hit village v at time t. First differencing controls for any idiosyncratic time-invariant characteristic correlated with consumption; I also run some specifications in logs, in which case I add household-level fixed effects instead. Finally, X_{it} captures any other factors that could affect the optimal allocation of consumption and should be controlled for. In particular, for some specifications I control for household leisure over time, which affects consumption if preferences are non-separable and the planner cannot freely transfer leisure across households (Cochrane, 1991). In other specifications, I instead use leisure as an instrument to control for attenuation bias that might come from measurement error in the income variable. Leisure is a suitable instrument as it is undoubtedly correlated with income — households that spend more hours hunting, fishing, or harvesting will collect higher income, all else equal — but, because leisure is a separate survey item, measurement error in leisure arguably does not correlate with error in income (I follow Schulhofer-Wohl (2011) in this approach). All variables are expressed in adult-equivalent terms: I divide by a measure of a household's average adult caloric intake developed for the Tsimane' data set by Hooper (2011); it estimates caloric consumption across gender and age levels, and weights each household's demographic composition accordingly.[36] Standard errors are clustered at the household level.

Recall that we cannot reject the hypothesis of full risk sharing for values of $\beta_1 = 0$ and $\beta_2 = 1$. As shown in Table 1.9, I reject full insurance across all specifications. Coefficients on own income are about $\beta_1 = 0.35$ and statistically significant at the 1% level. Leisure is negative associated with consumption, as expected, but remains non-significant. Controlling for non-separabilities in income and consumption does not change the log estimates and lowers the growth rate estimates only by 0.005. The Instrumental Variables estimator controls for attenuation bias and therefore provides slightly higher estimates both for logs and growth rates; the difference, however, is quite small. All in all, I find a considerable correlation between consumption and own-income, consistent with previous studies in similar settings. Although the magnitude of this association varies across studies, a value of 0.35 falls well within the expected range. For instance, Munshi and Rosenzweig (2009) estimate values between 0.17 and 0.26 for Indian data, while Cochrane (1991) finds values between 0.1 and 0.2 in the PSID; Kinnan (2014) finds values ranging from 0.07 to 0.3 for Thai data, depending on the type of estimation.[37] Overall, the results square fairly well with the literature and unequivocally reject full insurance. The theory provides new ways to think of partial insurance within a network context and help us understand the type of behavior that exists when we reject full insurance. To bring the main theoretical predictions to data, I first estimate the income process; I then fit the sharing rule to data.

[36]See table 2.2 in Hooper (2011) for a more detailed explanation of this adult consumption measure

[37]Saidi (2015) finds that the magnitude of departure from efficiency is smaller than mine in the Tsimane' communities. His data set comes from an entirely different survey corresponding to non-overlapping sets of Tsimane' villages. Moreover, he defines income from the sale of goods and labor as a separate survey item, whereas consumption here is income plus transfers, and therefore more tightly correlated.

1.5.2 Estimating the Income Process

Before taking my model to data, it is necessary to obtain an estimate for the income process. In this section I develop estimates of $\Psi = \frac{\mu^2+\rho}{\rho-\sigma^2}$ from panel data on the income processes of households. I carry out two different estimation procedures that deliver similar results. I first perform a very simple non parametric approach that uses basic moment estimators. I assume that income is described by two transitory shocks, an aggregate and an idiosyncratic one,

$$y_{it} = \kappa_t + \epsilon_{it}$$

with $\kappa_t \sim iid\,(\mu, \sigma_\kappa^2)$ and $\epsilon_{it} \sim iid\,(0, \sigma_\epsilon^2)$. I take the cross sectional average of income as an estimator of the aggregate shock, $\hat{\kappa_t} = \frac{1}{n}\sum_i y_{it}$, and I calculate the variability of this estimator to obtain an estimate of its variance $\hat{\sigma}_\kappa^2 = var(\hat{\kappa}_t)$. I obtain the variability in the residuals, $y_{it} - \hat{\kappa_t} = \hat{\epsilon}_{it}$, as an estimator of $\hat{\sigma}_\epsilon^2$. Finally, I compute the mean μ, the variance σ^2, and the common covariance term ρ, as follows: $\mu = \frac{1}{T}\sum_t \hat{k}_t$, $\sigma^2 = \hat{\sigma}_\kappa^2 + \hat{\sigma}_\epsilon^2$, and $\rho = \hat{\sigma}_\kappa^2$. These values deliver an estimate of Ψ equal to 0.97 ± 0.2.

I also perform a more sophisticated estimation procedure, availing myself of a vast and well established literature on estimating earnings processes from data.[38] The estimation assumes a state space model for the income process. Income is assumed to follow an aggregate shock, a temporary shock, and a persistent shock.

$$\begin{aligned} y_{it} &= \kappa_t + \eta_{it} + \nu_{it} \\ \eta_{i,t} &= c + \gamma\eta_{i,t-1} + \epsilon_{it} \end{aligned}$$

This model is estimated by GMM. More information is given in Section 1.8.6, where I go over the estimation details and I show that I obtain values of Ψ close to those obtained with the more naive approach of the previous paragraph.

1.5.3 Structural Estimation of Network Flows

Having constructed networks and estimated the underlying income process, we are now ready to fit the model's sharing rules against the Tsimane's data set. To bring the model to data, recall the closed form expression for the constrained-efficient sharing rule as a linear function of global network measures shown in equation (1.9),

$$\alpha_{ij}^\star = g_{ij}\left(M_j\left(\Psi, \mathbf{G}\right) - \frac{\Psi}{1+\Psi d_i}\sum_k g_{ik} M_k\left(\Psi, \mathbf{G}\right)\right) \tag{1.15}$$

[38] See, for instance, Lillard and Weiss (1979), MaCurdy (1982), Nakata & Tonetti (2015)

and assume that shares are measured every period with additive error: observed bilateral shares are $\alpha_{ijt}^{obs} = \alpha_{ijt} + \epsilon_{ijt}$. Then the constrained-efficient sharing rule proposed in this paper implies the following relationship,

$$\alpha_{ijt}^{obs} = \beta_1 M_j + \beta_2 M_{N_i} + \epsilon_{ijt} \tag{1.16}$$

where $M_{N_i} = \frac{\Psi}{1-\Psi d_i} \sum_k g_{ik} M_k (\Psi, \mathbf{G})$ aggregates the WEPC centrality measure across of all of $i's$ neighbors. Under such a specification the theory requires that $\beta_1 = 1$, and $\beta_2 = -1$. I control for village-level shocks by allowing for village-time specific intercepts and control for a number of household-specific attributes, such as household size, total wealth, marital status, and average age of the household heads.[39] I show in Table 1.10 that we obtain estimates statistically indistinguishable at the 5% level from $\beta_1 = 1$ and $\beta_2 = -1$ for the Kinship, Trade, and Updated Trade networks, using a value of $\Psi = 0.9$ as estimated in the previous section. Values for the Trade and Update Trade networks are statistically indistinguishable at the 1% level. Moreover, in Figures 1.6 to 1.8 I plot the two coefficients for a wide range of Ψ values to clarify that these results are fairly robust to possible estimation errors. Both coefficients are statistically different from zeros. We see that β_2 clearly takes on values close to -1 for values of Ψ away from zero and that for most values it is statistically indistinguishable from -1 at the 95% confidence level. Already in this first direct approach the model performs remarkably well at describing the observed relationship between network structure and household exchanges.

I also perform a more demanding test of the model. Ideally, I would like to estimate the sharing rule separately for each pair of households over time, and obtain distinct coefficients to each of the centrality measures in equation (1.15). The lack of sufficient longitudinal data, however, precludes this type of analysis. As an alternative, notice that the second term in equation (1.15) varies according to the degree of the receiving household. I exploit this variability by splitting the population by their degree, d_i, and redefining M_{N_i} in (1.16) as, $M_{N_i} = \sum_k g_{ik} M_k (\Psi, \mathbf{G})$. In this case, a successful model would obtain a negative coefficient of $\beta_2 (d_i) = -\frac{\Psi}{1+\Psi d_i}$ that increases with the degree. Figures 1.10 and 1.11 plot $\beta_2 (d_i)$ against the degree of each group. I also plot the curve representing the theoretical prediction of $\frac{\Psi}{1+\Psi d_i}$. The positive relationship between this coefficient and the degree of the receiving household is clear. Moreover, the model and theoretical predictions move in a similar fashion and reasonably close to each other.

[39] Because I assume networks to be undirected, any edge that only sustains unilateral exchanges is complemented by adding a flow equal to zero in the opposite direction for any period in which the household in question obtains positive production.

Although the model seems to fit the sharing data quite well, we might still worry that other network centrality measures could also predict similar results, undermining the model's predictive power. After all, it is well known that network centrality measures often correlate strongly. In 1.8.5 I discuss the statistical relationship between the centrality measure proposed here and several other familiar candidates in the network literature. I show that these other measures are not strongly correlated with the WEPC centrality that I propose, and more importantly that they fail to explain the patterns of exchange observed in the Tsimane' data.

1.5.4 Revisiting the Risk-Sharing Test

In this section I show that the positive coefficient on own income obtained in the test of full insurance of Section 1.5.1 can be interpreted as capturing the bilateral sharing arrangement proposed in this paper. To do this I run the income data through my model to obtain household consumption under the constrained-efficient arrangement that I propose. I then estimate the risk sharing test of equation (1.14) with predicted, rather than observed, consumption data. A successful theory of partial insurance would retrieve the same coefficients to own income as those observed in the data in section 1.5.1.

More concretely, I use the sharing rule of Proposition 1.3.1 to calculate the consumption level of each household in every period, given income, as,

$$\hat{c}_{it} = \sum_j g_{ij} \left(M_j - \frac{\Psi}{1 + \Psi d_i} \sum_k g_{ik} M_k \right) y_{jt}$$

where I have dropped the explicit dependency on $\mathbf{G}$ and Ψ for convenience. This equation defines household consumption using the proposed sharing rule of the model. Expected consumption data is used, in lieu of the actual consumption, to test whether the type of variability in household consumption behavior proposed by this model can replicate the departure from full efficiency observed in the data.

The results are presented in Tables 1.12 to 1.14 for the different networks being analyzed, and for the value of Ψ estimated in section 4.3. Under my model's predictions, and for the output data available for the Tsimane', the coefficient on own income corresponding to the Trade Network shown in Table 1.13 oscillates between 0.15 ± 0.023 for OLS to about 0.25 ± 0.06 for IV estimates. Compared to the value of about 0.35 ± 0.07 that we obtain in Table 1.9, the model seems to slightly underestimate the empirical loading on own income for this network structure. However, it is worth noting that these differences are not large, and that for the IV estimates the difference is statistically not significant. The results for the kinship network presented in Table 1.12 show estimates far too low to resemble the magnitude of departure from efficiency in Table 1.9; the specification in growth rates provides OLS estimates that are non significant, suggesting full insurance under network constraints. This is not particularly surprising given that, as mentioned above, kinship networks are excessively dense, so that the planner problem is far less constrained than in the other network structures.[40] On the other hand, the results for the Updated Trade network in Table 1.14 are statistically indistinguishable from the estimates in Table 1.9. Although these coefficients, again, lie below the values provided by data, the differences now are negligible—sometimes as small as 0.01 — and therefore are all statistically insignificant.

1.5.5 Underlying Heterogeneity in Consumption

The environment I describe not only accounts for the type of coefficients we obtain when we reject full insurance, but, more importantly, defines a complete distribution of these coefficients based on network measures. Moreover, as argued in Section 2, the size of these coefficients can provide information about the relative sharing opportunities of each household in certain environments. Indeed, households with lower coefficients to their own income process are far more central than others, and as a result obtain in general smoother consumption paths. In other words, while the previous section showed that the model can generate a common coefficient that reflects the observed departure from efficiency, the following estimation procedure implies that the theory also provides insight into the type of asymmetric insurance possibilities affecting households as a result of their social situation.

Recall that a household's share of its own income left for consumption can be described, for each i, as,

$$\alpha_{ii} = M_i\left(\Psi, \mathbf{G}\right) - \frac{\Psi}{1+\Psi d_i}\sum_k g_{ik} M_k\left(\Psi, \mathbf{G}\right) \tag{1.17}$$

Retrieving precisely these values from data would require estimating each household's theoretical consumption, as described in (1.17), independently. The short time dimension of the panel unfortunately prohibits this type of analysis. Moreover, the variation across these values is often small for "similar" nodes and would be difficult to extract from the inherent noise in data. Instead, I decide to rank the population according to (1.17) and then split the population into equally sized groups. I then estimate equation (1.14) separately for each group; This gives me enough variability both within and across groups to effectively measure the expected positive difference across successive groups. [41]

[40]For this same reason this network performs worst in the structural estimation of bilateral exchanges of section 4.2 for the value of Ψ estimated from data.

[41]In order to allow for as much intergroup variability as possible, group size was kept as small as

The results are shown in Figures 1.12 and 1.13. The positive trend across groups is evident for all networks, and is especially pronounced for the trade and updated trade networks. In all these cases, while any two consecutive groups might show little variation, the overall increase from the first to the last group is generally about 0.5, and in all but the kinship network the difference is statistically significant. In other words, the positive association between income and consumption found for the Tsimane' data set can be further decomposed into those households that, by the overall social arrangement, consume more or less of their own income ex-post.

1.6 Conclusion

Time and again, evidence collected from risk-sharing communities in the developing world has concluded that households in these type of arrangements are only partially insured against random fluctuations in income. In this paper I argue that these insurance mechanisms overwhelmingly perform below full efficiency precisely because networks of interactions are not completely connected. I show that if the underlying social structures are accounted for when deriving constrained-efficient exchanges, then observed trades across pairs of households is well described by the theory, and the distance from the Pareto frontier can be obtained.

I propose a constrained-efficient framework that relaxes a crucial assumption in the classical risk-sharing literature, which allows all households to trade with each other. Instead, I restrict the movement of goods along a given set of social relations and I derive a full analytical description of the exchanges between any two households as a function of their network position. I show that exchanges are determined by a global network measure that accumulates all direct and indirect interactions along the entire network. In other words, this theory endogenizes pairwise sharing behavior along any given network. The theory is useful in providing a rich description of the type of partial insurance we might expect if we believe network constraints are a relevant friction keeping communities below full efficiency. More importantly, it can be easily tested in a number of different settings, as long as income and network data is available. In this sense, it is capable of providing testable predictions at the pairwise level, generating much more detailed variation on the exchanges generating consumption streams.

I test the theory with data from Tsimane' indigenous communities in the Bolivian Amazon. I structurally estimate the constrained-efficient sharing rule against bilateral exchanges observed across Tsimane' households and find that the theory does a good job of fitting empirical sharing behavior. Moreover, predicted

possible, while retaining enough observations to provide efficient standard errors. Average group size was 35 households per group, leading to a total of 7 groups.

consumption profiles generate the type of inefficiencies observed in these communities, and other important implications on the distribution of insurance levels across different households are also observed in data. Overall, evidence from Tsimane' communities suggest that accounting for incomplete social structures goes a long way to explain the type of partial insurance mechanisms operating more broadly in village economies.

Of course a number of other elements have been proposed that surely form part of a full description of these complicated social arrangements. For instance, Ligon et al. (2000) have studied the presence of limited commitment in these informal exchanges and argue that incentive constraints under dynamic contracts indeed lead to partial insurance similar to that first observed by Townsend (1994) for Indian villages. More recently, Schulhofer-Wohl (2015) and Mazzocco and Saini (2012) have stressed that heterogeneous preferences might lead one to overestimate the failure of full insurance. I believe these views and the one I propose here are complementary and together build a richer story of informal insurance. Indeed my model refrains from considering these and many other interesting dimensions, and I try and stay as close as possible to the classical setup of risk-sharing proposed by Mace (1991), Cochrane (1991), and Townsend (1994), while at the same time allowing me to engage directly with a general network structure.

A network description of exchanges like the one proposed here holds great promise for identifying vulnerable households, or determining superior social arrangements. After all, one of the advantages of modeling social interactions explicitly in this context is that it provides a great deal of heterogeneity on consumption volatility and inequality both across households and networks. It would be interesting to know, for instance, which arrangements perform better than others, and whether we can generally identify households that, if removed, would most affect the sharing opportunities of the entire community. Indeed network models like this one have already answered these type of questions in other settings, such as Ballester et al. (2007) which do a similar exercise for criminal networks. Although I provide some tentative results on the ranking of households by consumption volatility in section 1.8.4, the ordering is only partial and I am currently working on new results. The challenge here, with respect to Ballester et al. (2007), has to do with the complicated weighting scheme for paths that emerges in this setting, and which is absent in the Bonacich centrality or other similar recursive network measures. Indeed a lot of the existing tools to make progress on this front utilize the convenient geometric weighting of Bonacich, but I am working on a recursive formulation of WEPC that would allow me to make progress nonetheless.

Another ambitious proposal that emerges from this analysis seeks to structurally estimate the underlying social structure. In other words, if we agree the model performs well in this context, and might therefore be a good proxy for the type of sharing behavior of the Tsimane', then an exciting step forward would utilize the theory's predictions in order to structurally back out an estimate of the true underlying structure. This approach is not without its own set of challenges, not least of which is that the model requires inputing an entire network described by an n^2 dimensional object. Extracting this from data is not simple. However, there have been some recent developments by Manresa (2015) on estimating the structure of interactions from panel data using a pooled lasso estimator that might be very useful. If we can frame the spillover effects in an amenable way, it might be possible to identify the most likely structure from within the class of sparse networks.

Perhaps the most promising step forward involves more general results on the welfare implications from my theory. Indeed, a theoretical result that relates network-based heterogeneity in consumption behavior to more general welfare implications would provide a clear economic interpretation to the coefficients of empirical risk sharing tests. In other words, beyond rejecting or not full insurance, the theory could allow data to speak more clearly on the distribution of welfare across the population when full insurance is rejected. The distribution of households' response to income shocks that this paper predicts could then be mapped directly to a normative implication on welfare. It would form an important contribution, and would come full circle towards a new interpretation of empirical risk sharing test.

1.7 Tables and Figures

This section presents all tables and figures in the order in which they appear in the text. Some additional tables and figures can be found in the Online Appendix.

	n	mean	sd	median	min	max
HH Size	245	5.22	3.04	5.00	1.00	16.00
Mean Age Head of HH	245	36.96	16.33	34.00	14.50	86.00
No. of Dependents	245	3.12	2.74	2.00	0.00	13.00
% Rice Sold on Market	184	32.03	25.54	29.23	0.00	95.00
Animal Wealth (Bolivianos)	199	1478.28	3500.81	455.00	0.00	28250.00
Traditional Wealth (Bolivianos)	199	1179.94	1078.11	743.00	0.00	5917.50
Modern Wealth (Bolivianos)	199	3582.21	2356.61	3348.64	184.68	10726.22
Total Wealth (Bolivianos)	199	6259.17	5089.62	5403.40	363.20	35582.94
Avg. Income (Calories)	244	925.63	694.70	750.30	0.00	3896.01
Avg. Consumption (Calories)	246	888.02	694.77	707.66	46.67	4527.67
Avg. Out Flow (Calories)	242	353.53	536.24	174.18	0.00	4365.31
Avg. In Flow (Calories)	246	286.84	415.54	167.88	0.00	3441.35
Avg. Leisure (Hours)	244	46.22	1.48	46.61	35.97	48.00

Table 1.1: Household Summary Statistics: *Variables expressed in adult-equivalent terms. Averages taken over periods where data is available*

	n	Edges	Avg.Degree	Diameter	Density	Cluster	Avg.Between	Avg.Closeness
1	27	73	5.407	6	0.193	0.596	0.026	0.066
2	38	121	6.368	5	0.163	0.441	0.055	0.344
3	11	45	8.182	2	0.682	0.676	0.042	0.753
4	20	46	4.600	6	0.219	0.484	0.058	0.115
5	13	51	7.846	3	0.560	0.624	0.054	0.642
6	27	189	14.000	3	0.500	0.627	0.024	0.645
7	46	122	5.304	10	0.113	0.357	0.064	0.205
8	65	320	9.846	5	0.149	0.315	0.022	0.422

Table 1.2: Network Statistics Per Village: Trade Network

	n	Edges	Avg.Degree	Diameter	Density	Cluster	Avg.Between	Avg.Closeness
1	26	194	14.923	5	0.287	0.687	0.043	0.229
2	38	292	15.368	4	0.202	0.806	0.017	0.077
3	11	111	20.182	2	0.917	0.940	0.010	0.928
4	20	110	11.000	5	0.275	0.810	0.029	0.092
5	13	121	18.615	3	0.716	0.871	0.030	0.779
6	26	210	16.154	4	0.311	0.781	0.025	0.155
7	44	768	34.909	4	0.397	0.691	0.016	0.609
8	64	1594	49.812	5	0.389	0.810	0.013	0.581

Table 1.3: Network Statistics Per Village: Kinship Network

	Hamming Distance	Normalized Hamming Distance
1	100	0.285
2	116	0.165
3	18	0.327
4	29	0.153
5	30	0.385
6	128	0.365
7	320	0.309
8	804	0.387

Table 1.4: Hamming Distance per Village between Trade and Kinship Networks

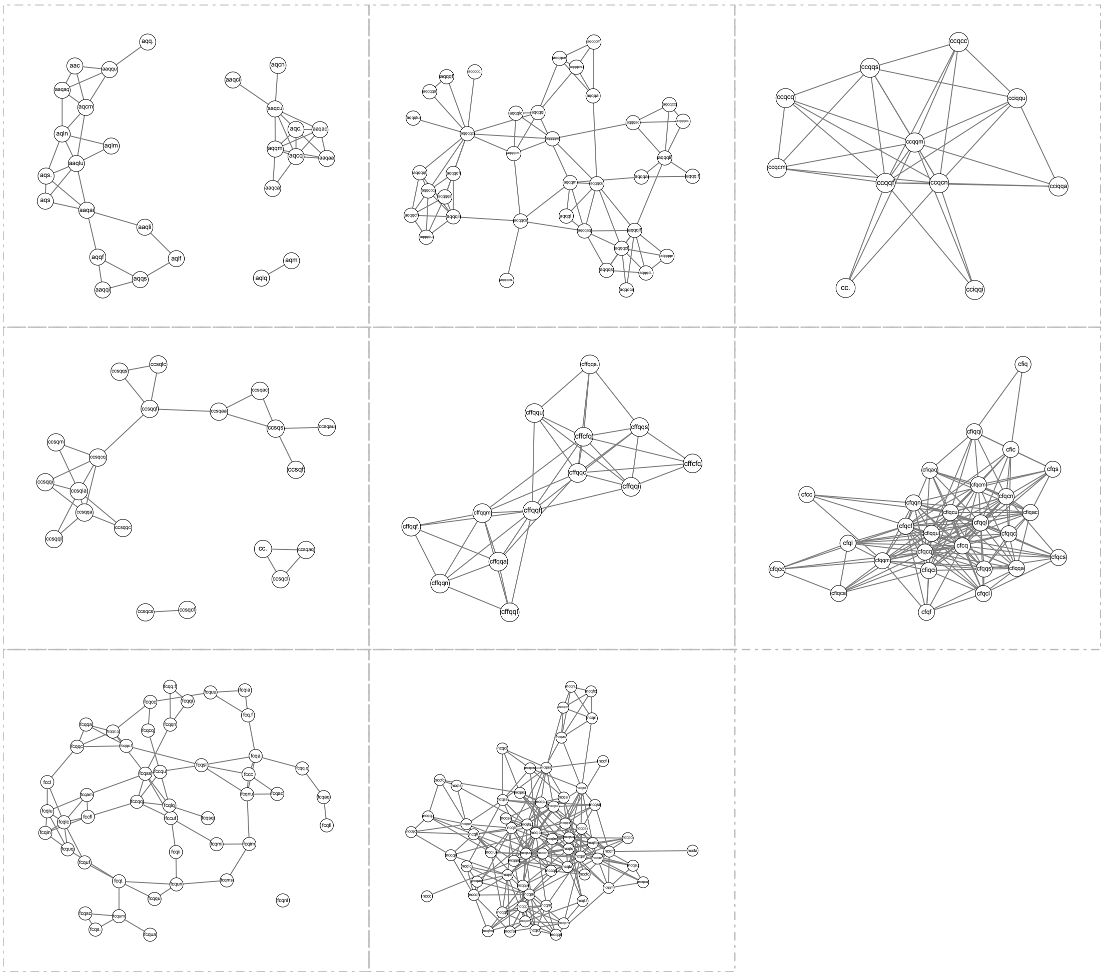

Figure 1.4: Trade Network: Link exists if households exchange food at any point in the sample.

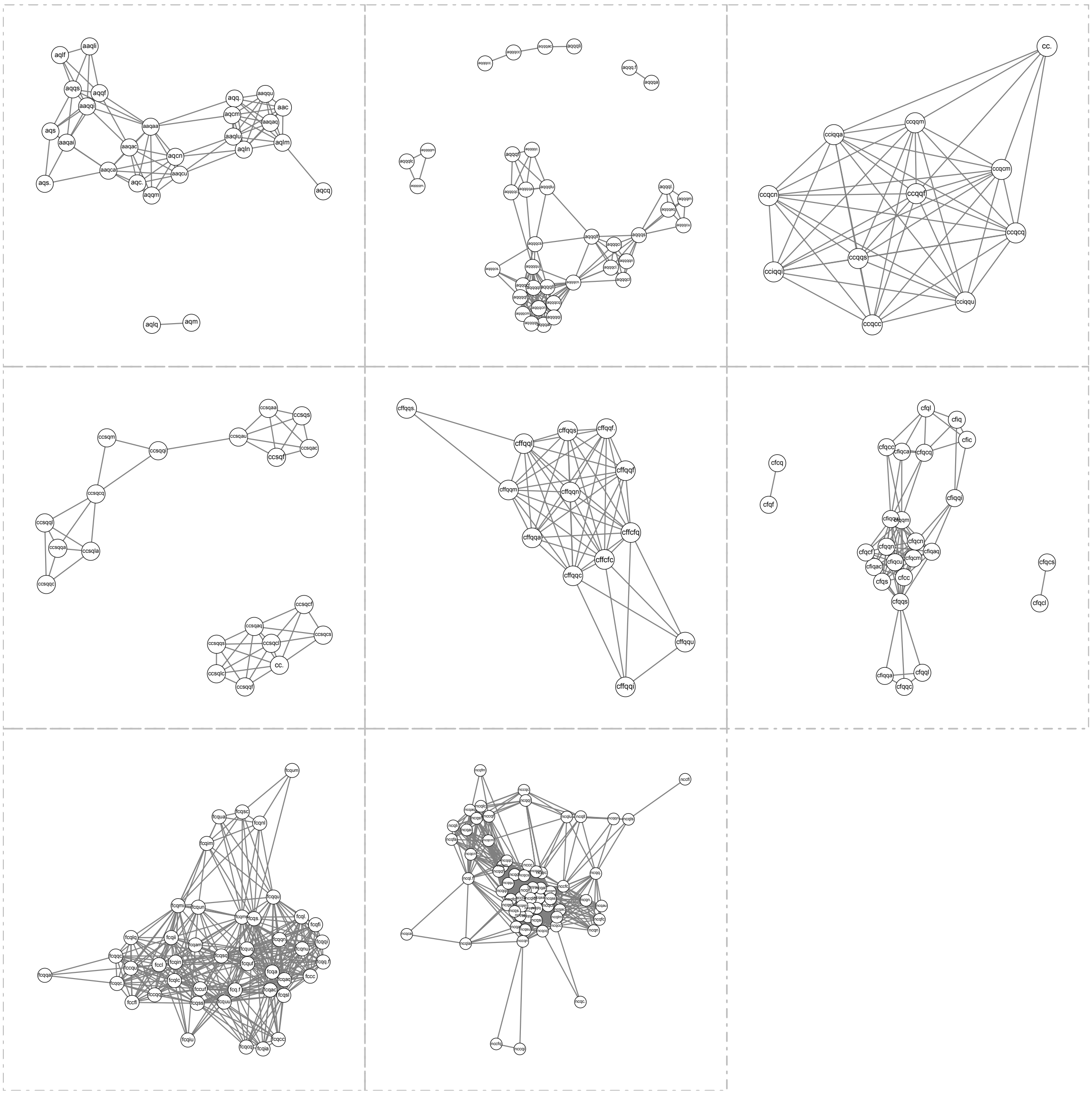

Figure 1.5: Kinship Network: Link exists if Mean Genetic Relation is above 0

Table 1.5: Local Correlations: Trade Network

	Dependent variable:					
	log(consumption)			Δ log(consumption)		
	No IV		Instrument: Leisure	No IV		Instrument: Leisure
	(1)	(2)	(3)	(4)	(5)	(6)
log(income)	0.362***	0.360***	0.375***			
	(0.039)	(0.040)	(0.039)			
log(leisure)		−0.344				
		(2.452)				
log(consumption $\in N_i^2$)	0.298***	0.299***	0.295***			
	(0.107)	(0.108)	(0.106)			
log(consumption $\notin N_i^2$)	0.214	0.215	0.215			
	(0.145)	(0.145)	(0.143)			
Δ log(income)				0.378***	0.376***	0.396***
				(0.036)	(0.037)	(0.037)
Δ log(leisure)					−0.463	
					(1.933)	
Δ log(consumption $\in N_i^2$)				0.117	0.117	0.117
				(0.103)	(0.102)	(0.102)
Δ log(consumption $\notin N_i^2$)				0.096	0.095	0.099
				(0.114)	(0.115)	(0.111)
Household Fixed Effects	Y	Y	Y	N	N	N
Village-Time Fixed Effects	Y	Y	Y	Y	Y	Y
Observations	922	922	922	692	692	692
R^2	0.470	0.470	0.470	0.485	0.485	0.485
Adjusted R^2	0.350	0.349	0.349	0.463	0.462	0.463

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Variables constructed as log(1+x) to admit zeros in cons. and inc. data
Values in parentheses are standard errors clustered at the household level.
Instrumental Variables estimation uses Balestra-Varadharajan-Krishnakumar's transformation.

Table 1.6: Local Correlations: Kinship Network

	Dependent variable:					
	log(consumption)			Δ log(consumption)		
	No IV		Instrument: Leisure	No IV		Instrument: Leisure
	(1)	(2)	(3)	(4)	(5)	(6)
log(1 + inc)	0.361***	0.365***	0.326***			
	(0.040)	(0.042)	(0.040)			
log(leisure)		0.873				
		(2.041)				
log(consumption $\in N_i^2$)	−0.439***	−0.447***	−0.444***			
	(0.165)	(0.168)	(0.167)			
log(consumption $\notin N_i^2$)	−0.114	−0.112	−0.120			
	(0.108)	(0.106)	(0.108)			
Δ log(income)				0.363***	0.363***	0.367***
				(0.038)	(0.039)	(0.038)
Δ log(leisure)					−0.077	
					(1.632)	
Δ log(consumption $\in N_i^2$)				−0.437***	−0.437***	−0.436***
				(0.155)	(0.156)	(0.155)
Δ log(consumption $\notin N_i^2$)				−0.160	−0.160	−0.159
				(0.108)	(0.107)	(0.108)
Household Fixed Effects	Y	Y	Y	N	N	N
Village-Time Fixed Effects	Y	Y	Y	Y	Y	Y
Observations	878	878	878	669	669	669
R^2	0.477	0.478	0.476	0.506	0.506	0.506
Adjusted R^2	0.358	0.357	0.357	0.481	0.480	0.481

Note: *p<0.1; **p<0.05; ***p<0.01

Variables constructed as log(1+x) to admit zeros in cons. and inc. data

Values in parentheses are standard errors clustered at the household level.

Instrumental Variables estimation uses Balestra-Varadharajan-Krishnakumar's transformation.

Table 1.7: Local Correlations: Trade Network

	Dependent variable:					
	log(consumption)			Δ log(comsumption)		
	No IV		Instrument: Leisure	No IV		Instrument: Leisure
	(1)	(2)	(3)	(4)	(5)	(6)
log(income)	0.353***	0.350***	0.378***			
	(0.039)	(0.040)	(0.039)			
log(leisure)		−0.659				
		(2.450)				
log(consumption $\in N_i$)	0.235***	0.236***	0.232***			
	(0.077)	(0.077)	(0.076)			
log(consumption $\in N_i^2 \cap \notin N_i$)	−0.077	−0.077	−0.081			
	(0.064)	(0.065)	(0.064)			
log(consumption $\notin N_i^2$)	0.155	0.155	0.157			
	(0.144)	(0.144)	(0.139)			
Δ log(income)				0.369***	0.369***	0.368***
				(0.036)	(0.038)	(0.036)
Δ log(leisure)					0.013	
					(1.972)	
Δ log(consumption $\in N_i$)				0.190**	0.190**	0.190**
				(0.080)	(0.081)	(0.080)
Δ log(consumption $\in N_i^2 \cap \notin N_i$)				−0.128*	−0.128*	−0.128*
				(0.069)	(0.069)	(0.069)
Δ log(consumption $\notin N_i^2$)				0.071	0.071	0.071
				(0.108)	(0.108)	(0.108)
Household Fixed Effects	Y	Y	Y	N	N	N
Village-Time Fixed Effects	Y	Y	Y	Y	Y	Y
Observations	914	914	914	682	682	682
R^2	0.471	0.471	0.470	0.501	0.501	0.501
Adjusted R^2	0.349	0.349	0.349	0.476	0.476	0.476

Note: *p<0.1; **p<0.05; ***p<0.01

Variables constructed as log(1+x) to admit zeros in cons. and inc. data

Values in parentheses are standard errors clustered at the household level.

Instrumental Variables estimation uses Balestra-Varadharajan-Krishnakumar's transformation.

Table 1.8: Local Correlations: Kinship Network (Kinship)

	Dependent variable:					
	log(consumption)			Δ log(consumption)		
	No IV		Instrument: Leisure	No IV		Instrument: Leisure
	(1)	(2)	(3)	(4)	(5)	(6)
log(income)	0.357^{***}	0.360^{***}	0.328^{***}			
	(0.040)	(0.042)	(0.040)			
log(leisure)		0.733				
		(1.938)				
log(consumption $\in N_i$)	-0.047	-0.053	-0.049			
	(0.107)	(0.102)	(0.106)			
log(consumption $\in N_i^2 \cap \notin N_i$)	-0.375^{***}	-0.378^{***}	-0.388^{***}			
	(0.120)	(0.120)	(0.118)			
log(consumption $\notin N_i^2$)	-0.098	-0.096	-0.103			
	(0.114)	(0.112)	(0.113)			
Δ log(income)				0.363^{***}	0.363^{***}	0.362^{***}
				(0.038)	(0.039)	(0.038)
Δ log(leisure)					0.012	
					(1.688)	
Δ log(consumption $\in N_i$)				-0.099	-0.100	-0.100
				(0.105)	(0.104)	(0.105)
Δ log(consumption $\in N_i^2 \cap \notin N_i$)				-0.268^{**}	-0.268^{**}	-0.268^{**}
				(0.105)	(0.105)	(0.105)
Δ log(consumption $\notin N_i^2$)				-0.151	-0.151	-0.151
				(0.106)	(0.105)	(0.106)
Household Fixed Effects	Y	Y	Y	N	N	N
Village-Time Fixed Effects	Y	Y	Y	Y	Y	Y
Observations	874	874	874	665	665	665
R^2	0.476	0.477	0.476	0.503	0.503	0.503
Adjusted R^2	0.356	0.356	0.355	0.477	0.476	0.477

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Variables constructed as log(1+x) to admit zeros in cons. and inc. data
Values in parentheses are standard errors clustered at the household level.
Instrumental Variables estimation uses Balestra-Varadharajan-Krishnakumar's transformation.

Table 1.9: Full Risk Sharing Test

	Dependent variable:					
	log(consumption)			Δ log(consumption)		
	OLS		IV: Leisure	OLS		IV: Leisure
	(1)	(2)	(3)	(4)	(5)	(6)
log(1 + inc)	0.355***	0.355***	0.357***			
	(0.036)	(0.020)	(0.036)			
log(leisure)		−0.043				
		(1.260)				
Δ log(income)				0.360***	0.355***	0.402***
				(0.034)	(0.035)	(0.036)
Δ log(leisure)					−1.070	
					(1.635)	
Household Fixed Effects	Y	Y	Y	N	N	N
Village-Time Fixed Effects	Y	Y	Y	Y	Y	Y
Observations	1,129	1,129	1,129	854	854	854
R^2	0.452	0.452	0.452	0.469	0.469	0.467
Adjusted R^2	0.339	0.339	0.339	0.449	0.449	0.448

Note: *p<0.1; **p<0.05; ***p<0.01

Variables constructed as log(1+x) to admit zeros in cons. and inc. data

Values in parentheses are standard errors clustered at the household level.

Instrumental Variables estimation uses Balestra-Varadharajan-Krishnakumar's transformation.

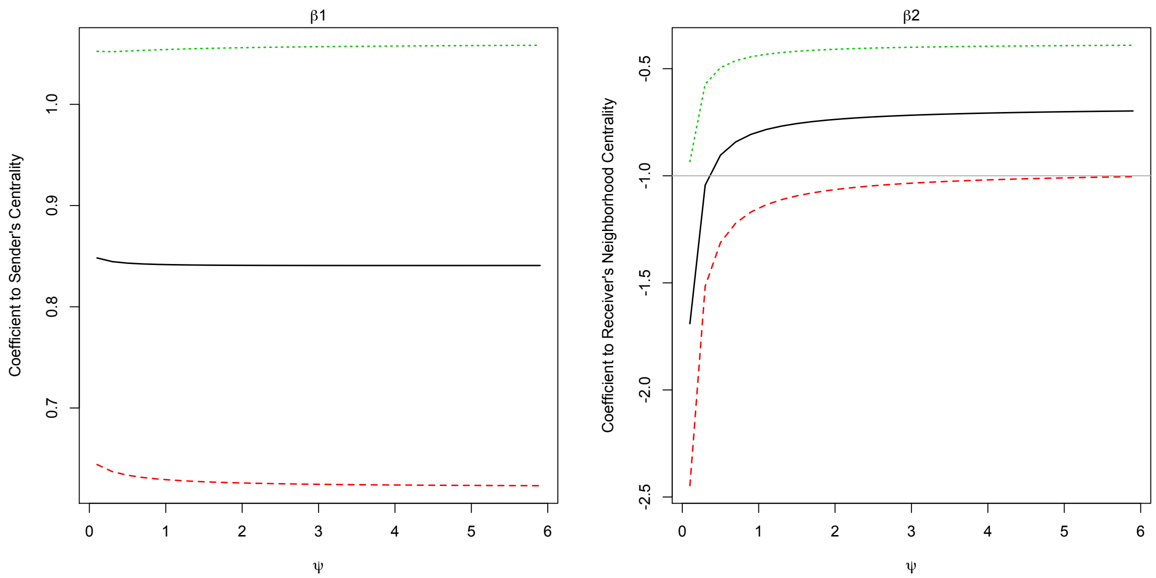

Figure 1.6: Coefficients to Network Centralities in Regression of Edge-Level Exchanges: Trade Network (Households younger than 4 0)

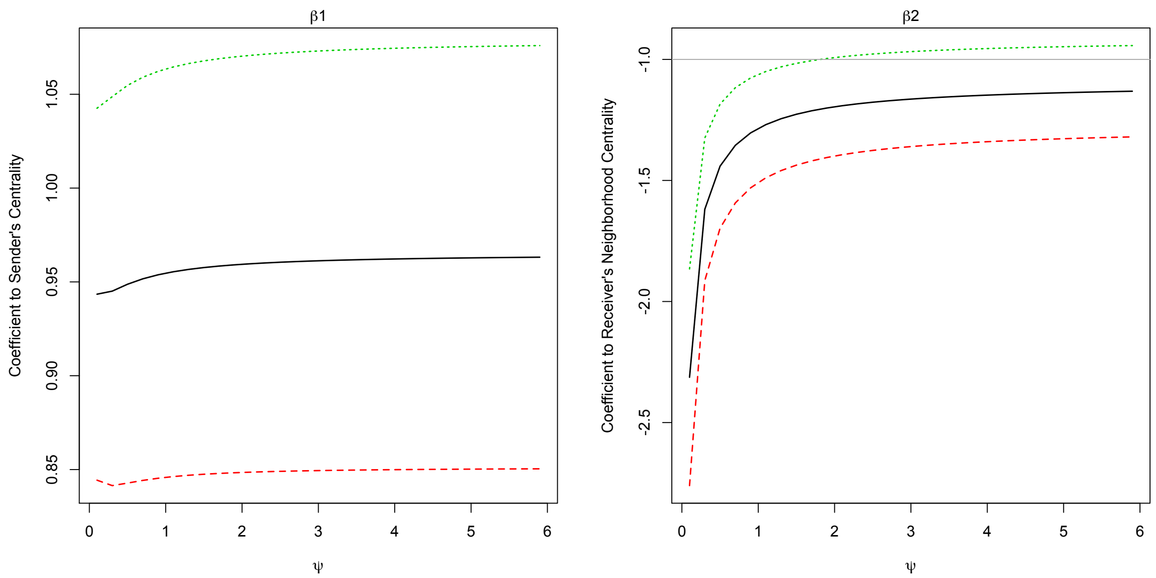

Figure 1.7: Coefficients to Network Centralities in Regression of Edge-Level Exchanges: Kinship Network

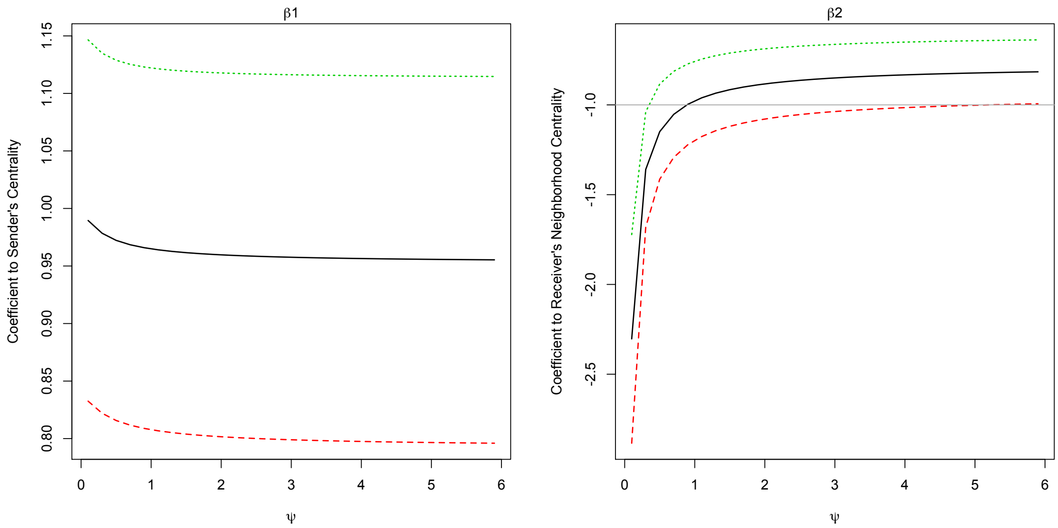

Figure 1.8: Coefficients to Network Centralities in Regression of Edge-Level Exchanges: Updated Trade Network

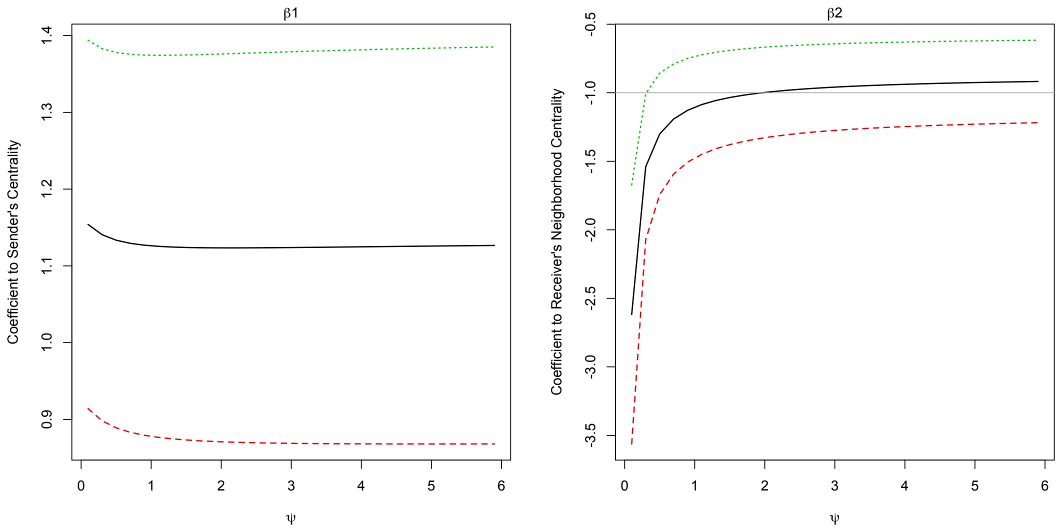

Figure 1.9: Coefficients to Network Centralities in Regression of Edge-Level Exchanges: Updated Trade Network (Households younger than 4 0)

Table 1.10: Regression of Edge-Level Exchanges on Predicted Sharing Rule ($\Psi = 0.9$)

	Dependent variable:		
	α_{ij} (i.e. share from i to j)		
	Kinship Network	Trade Network	Updated Trade Network
	(1)	(2)	(3)
$M_j(\mathbf{G}, \mathbf{\Psi})$	0.878***	0.841***	1.109***
	(0.044)	(0.111)	(0.093)
$M_{N_i}(\mathbf{G}, \mathbf{\Psi})$	$-$1.324***	$-$0.697***	$-$0.946***
	(0.113)	(0.156)	(0.133)
Village-Time Fixed Effects	Y	Y	Y
Observations	11,943	2,059	1,730
Adjusted R^2	0.092	0.180	0.202

Note: *p<0.1; **p<0.05; ***p<0.01
Values in parentheses are standard errors clustered at the household level.

Table 1.11: Regression of Edge-Level Exchanges on Alternative Local Measure

	Dependent variable:		
	α_{ij} (i.e. share from i to j)		
	Trade Network	Kinship Network	Updated Trade Network
	(1)	(2)	(3)
$\frac{\lvert N_i^2(\mathbf{G})\rvert}{\sum_i g_{ij}\lvert N_i^2(\mathbf{G})\rvert}$	0.096**	0.167***	$-$0.007
	(0.049)	(0.037)	(0.054)
Village-Time Fixed Effects	Y	Y	Y
Observations	5,586	11,931	2,742
Adjusted R^2	0.057	0.043	0.080

Note: *p<0.1; **p<0.05; ***p<0.01
Values in parentheses are standard errors clustered at the household level.

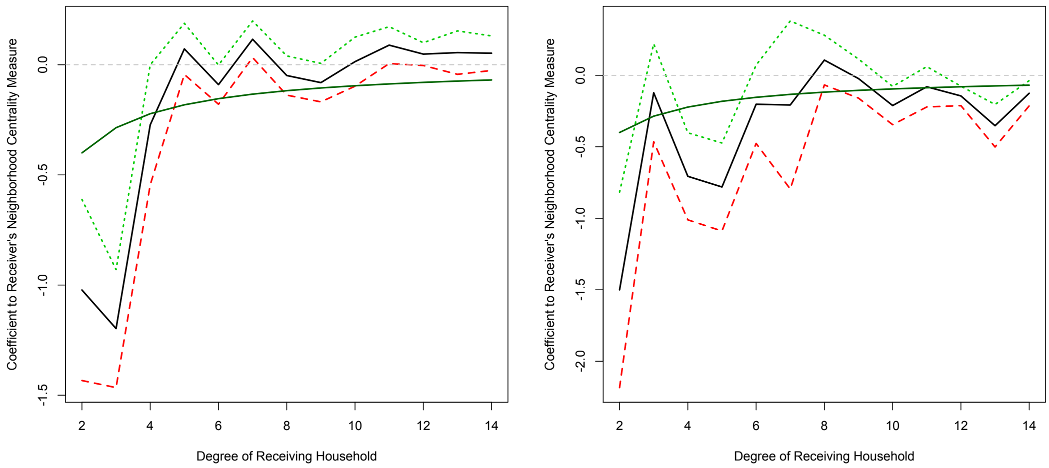

Figure 1.10: Coefficient β_2 as a function of Receiver's degree. Panel A: Trade Network. Panel B: Kinship Network

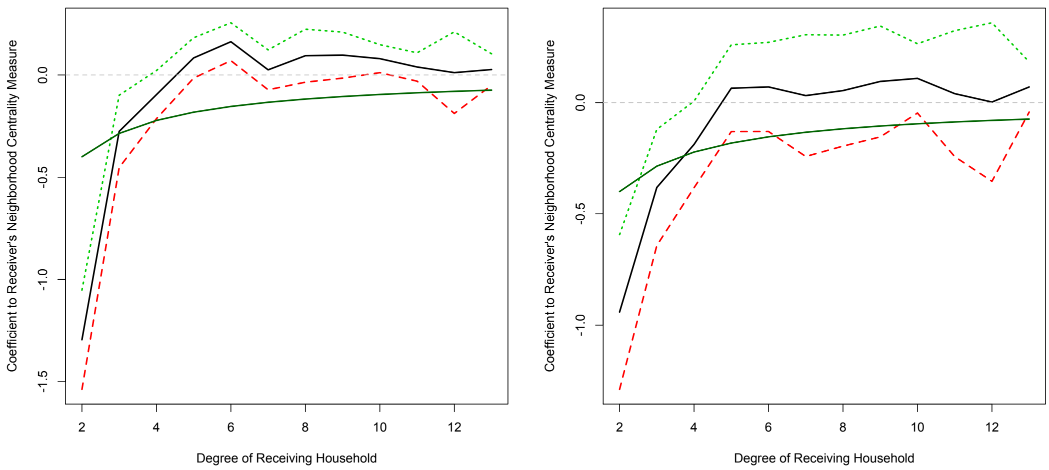

Figure 1.11: Coefficient β_2 as a function of Receiver's degree. Panel A: Updated Trade Network. Panel B: Updated Trade Network (Households younger than 40)

Table 1.12: Full Risk Sharing Test for Model Consumption Data: Kinship Network

	Dependent variable:					
	log(Predicted Consumption)			Δ log(Predicted Consumption)		
	OLS		IV: Leisure	OLS		IV: Leisure
	(1)	(2)	(3)	(4)	(5)	(6)
log(income)	0.062***	0.056***	0.158***			
	(0.016)	(0.016)	(0.021)			
log(leisure)		−1.282*				
		(0.711)				
Δ log(income)				0.026	0.017	0.141***
				(0.018)	(0.018)	(0.026)
Δ log(leisure)					−1.928**	
					(0.773)	
Household Fixed Effects	Y	Y	Y	N	N	N
Village-Time Fixed Effects	Y	Y	Y	Y	Y	Y
Observations	1,100	1,100	1,100	817	817	817
R^2	0.406	0.408	0.372	0.334	0.340	0.279
Adjusted R^2	0.302	0.303	0.277	0.319	0.324	0.266

Note: *p<0.1; **p<0.05; ***p<0.01
Variables constructed as log(1+x) to admit zeros in cons. and inc. data
Values in parentheses are standard errors clustered at the household level.
Instrumental Variables estimation uses Balestra-Varadharajan-Krishnakumar's transformation.

Table 1.13: Full Risk Sharing Test for Model Consumption Data: Trade Network

	Dependent variable:					
	log(Predicted Consumption)			Δ log(Predicted Consumption)		
	OLS		IV: Leisure	OLS		IV: Leisure
	(1)	(2)	(3)	(4)	(5)	(6)
log(1 + inc)	0.154***	0.148***	0.248***			
	(0.022)	(0.023)	(0.026)			
log(leisure)		−1.260*				
		(0.718)				
Δ log(income)				0.133***	0.124***	0.243***
				(0.025)	(0.025)	(0.031)
Δ log(leisure)					−1.851**	
					(0.883)	
Household Fixed Effects	Y	Y	Y	N	N	N
Village-Time Fixed Effects	Y	Y	Y	Y	Y	Y
Observations	1,096	1,096	1,096	812	812	812
R^2	0.398	0.399	0.380	0.343	0.347	0.317
Adjusted R^2	0.295	0.296	0.282	0.327	0.330	0.303

Note: *p<0.1; **p<0.05; ***p<0.01

Variables constructed as log(1+x) to admit zeros in cons. and inc. data

Values in parentheses are standard errors clustered at the household level.

Instrumental Variables estimation uses Balestra-Varadharajan-Krishnakumar's transformation.

Table 1.14: Full Risk Sharing Test for Model Consumption Data: Updated Trade Network

	Dependent variable:					
	log(Predicted Consumption)			Δ log(Predicted Consumption)		
	OLS		IV: Leisure	OLS		IV: Leisure
	(1)	(2)	(3)	(4)	(5)	(6)
log(1 + inc)	0.316***	0.312***	0.397***			
	(0.042)	(0.043)	(0.046)			
log(leisure)		−0.948				
		(1.152)				
Δ log(income)				0.286***	0.278***	0.392***
				(0.043)	(0.044)	(0.049)
Δ log(leisure)					−1.607	
					(1.485)	
Household Fixed Effects	Y	Y	Y	N	N	N
Village-Time Fixed Effects	Y	Y	Y	Y	Y	Y
Observations	1,086	1,086	1,086	800	800	800
R^2	0.401	0.402	0.396	0.360	0.362	0.351
Adjusted R^2	0.297	0.298	0.294	0.344	0.345	0.336

Note: *p<0.1; **p<0.05; ***p<0.01
Variables constructed as log(1+x) to admit zeros in cons. and inc. data
Values in parentheses are standard errors clustered at the household level.
Instrumental Variables estimation uses Balestra-Varadharajan-Krishnakumar's transformation.

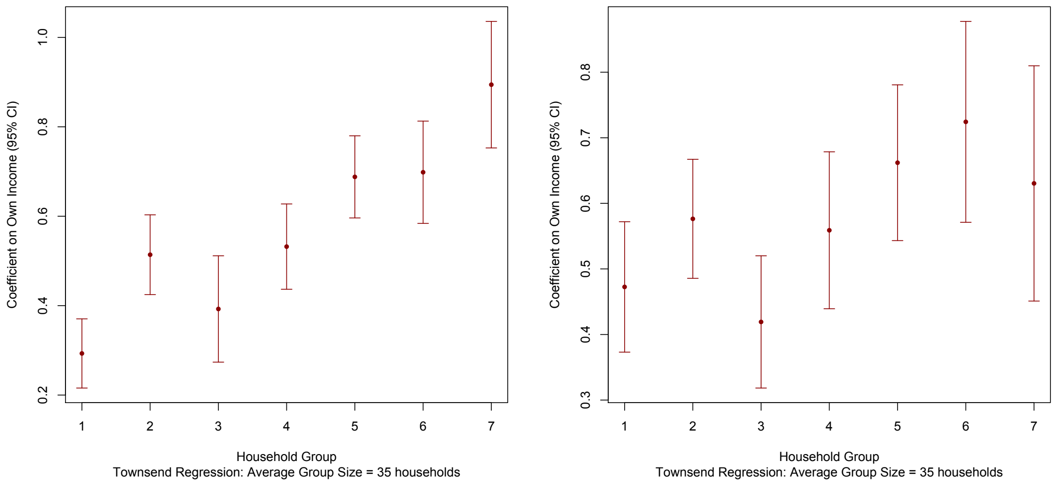

Figure 1.12: Coefficients and Confidence Intervals for Equation 1.14 Partitioning Population according to Centrality Measure. Panel A: Trade Network. Panel B: Kinship Network

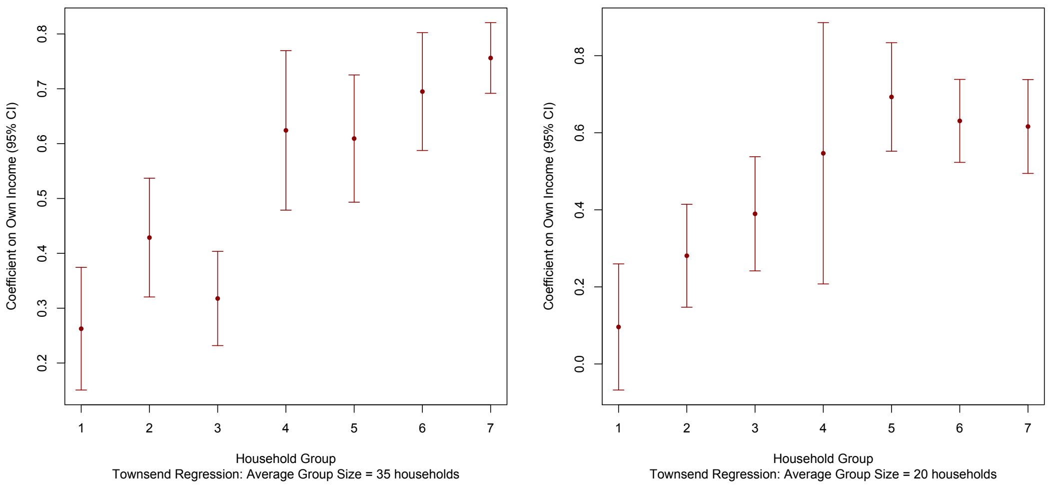

Figure 1.13: Coefficients and Confidence Intervals for Equation 1.14 Partitioning Population according to Centrality Measure. Panel A: Updated Trade Network. Panel B: Updated Trade Network (Age ¡ 40)

1.8 Additional Results

1.8.1 Contingent Sharing Rules

Consider the set of contingent sharing rules that maximize welfare for the simple economy in Figure 1. Shares from i to j now depend on the state of the world ω; intuitively, the distribution of income in each state will determine the sharing opportunities. For instance, if household 1 obtains an income $y_1(\omega)$ larger than $y_2(\omega)$ and $y_3(\omega)$ then funds can readily be redistributed so that the efficient condition of equation (1.1) holds for that particular state ω. However, it is easy to see that this will not be possible for all states: for instance if $y_1(\omega) < \bar{y}(\omega) - y_2(\omega)$ then the income of household 1 is not large enough to transfer the required resources to household 2.

Intuitively, because sharing groups are local, household consumption is bounded above by equation (1.3). A good way to think of this setting more generally is by imagining that the lack of intermediation essentially sets capacity constraints on what each node can transfer to its neighbors. This type of environment is explored in Ambrus et al. (2014) in the context of credit constraints on a network with exogenous link values. Here, rather than limiting the transfer across nodes by

As a result, only partial insurance is possible and the ratio of marginal utilities is not constant across all states. As we have just seen, the ratio is constant only for a subset of states where the income of the intermediating household is sufficiently large. More generally we can define a set $\bar{\Omega}(\mathbf{G}) \subseteq \Omega$, such that for a given network $\mathbf{G}$ it provides the subset of Ω such that full insurance is possible. The previous discussion signals the inherent difficulty in isolating general network effects from particular income realizations for these type of contingent sharing rules. Not surprisingly, a large part of the literature on risk sharing networks have dealt with fixed (or non-contingent) sharing rules.

1.8.2 A Model with Network Intermediation

In this section I show how to extend the current setup to a allow for network intermediation. In particular, I relax the assumption of no-intermediation to a general case where households can access income from households at some distance k (the setup analyzed in the main text corresponds to the situation where $k = 1$). To simplify the arguments, let $k = 2$ in what follows. However, all the arguments below apply for all values of k . In this scenario, consumption by household i is a linear function, not only of incomes of neighbors (as before), but also of the income of neighbors' neighbors (i.e. those households two links away from i) as follows,

$$c_i(\omega) = \sum_{jk} g_{ik} g_{kj} \alpha_{ik} \alpha_{kj} y_j(\omega)$$

It is easy to see that the relationship between consumption and income is still defined by my model's predictions as given in Proposition 1.3.1, but where the primitive of the model now is not the original network $\mathbf{G}$, but rather a new network $\tilde{\mathbf{G}}$ that is built from $\mathbf{G}$ by connecting all households that are two links apart. Indeed, while the exact trade is hard to describe analytically as a function of the true network $\mathbf{G}$, it is nonetheless very easy to describe as a function of the network $\tilde{\mathbf{G}}$: it depends on the sharing rule of Proposition1.3.1. In other words, the theoretical predictions of my model allow for general descriptions of intermediation, and the general arguments on the implications for consumption behavior follow through. The only caveat is that it is now difficult to characterize the type of exchanges (in this case exchanges of exchanges) that will lead to an efficient solution, but the efficient solution, in terms of consumption responses to income shocks, can be described precisely by my model. Finally, notice that there is an upper limit on the amount of intermediation that generates a situation of partial insurance. Indeed, if k is larger than the minimum distance separating any two households, then intermediation is sufficiently large that full insurance is retrieved.

1.8.3 Discussion of Weighted Even Path Centrality

Proposition 1.3.1 is powerful because it provides a full description of efficient sharing-behavior under restricted bilateral exchanges for all possible social networks and distributional parameters. As a concrete prescription of network flows to be tested against data, it suffices, and, as we will see in section 3, performs reasonably well. In this section I describe the expression of the WEPC in equation (1.10) in more detail and discuss what it can tell us about the optimal network shares.

It turns out that recursive expressions like the one in (1.10) are found often in network analysis. These measures attempt to quantify associations between vertices based solely on the structure of connections. For instance, in their well-known work on strategic complementarities in networks, Ballester et al. (2007) show that equilibrium actions depend on a similar recursive measure known as Bonacich Centrality. More recently, Banerjee et al. (2012) have sought to identify individuals in the network that are best placed to diffuse information on micro-credit opportunities in India. They find that participation is higher if those first informed have higher eigenvector centrality. [42] It is a matter of fact that global

[42] Already at the beginning of the internet boom, a number of algorithms surfaced that allowed users to rank websites by their significance in the broader world wide web network. Procedures

network measures such as these always appear in situations with entangled interactions along a set of connections. All of these measures can be expressed generically as

$$B_i = c + \gamma \sum_k g_{ik} B_k \tag{1.18}$$

for some constant c and with $|\gamma| < 1$. This expression essentially says that $i's$ measure depends linearly on the sum of measures that are connected to i. Let us distinguish two crucial differences with respect to the WEPC measure defined in (1.10). First of all, notice that equation (1.10) does not sum over all measures that i is connected to, but instead sums over all measures that $i's$ partners are connected to. In other words, the WEPC is defined recursively at distance two, not one. This is not entirely rare in network analysis and in fact appears in some work on vertex similarity by Jeh and Widom (2002). It has also appeared in newer page-ranking algorithms, such as the HITS algorithm.[43] Secondly, notice that, in contrast to equation (1.18), the WEPC does not weight all neighboring measures with a common parameter γ. Instead, the measures at distance two are weighted by the degree of the household that serves as a bridge between them. So for example, imagine two households k and i are both linked to a third household l, but are not linked to each other. Then, $k's$ measure will enter the definition of $i's$ measure, weighed by the degree of l. Moreover, notice that the particular weight has the familiar form of equation (1.8) that captures the full extent of $l's$ local interactions with its partners. Remember that $M_i(\mathbf{G}, \Psi)$ captures all of $i's$ indirect interactions along the network. Following earlier discussions, these indirect interactions affect i only in so far as they alter the sharing behavior of those at distance two from i (with whom i actually interacts). It only seems natural, then, that $i's$ measure, $M_i(\mathbf{G}, \Psi)$, is defined recursively from the measures, $M_l(\mathbf{G}, \Psi)$, for all households l that that i directly interacts with (i.e. those that are at distance two from i). In other words, the exact shape in the recursive definition of $M_i(\mathbf{G}, \Psi)$ spells out quite clearly the preceding discussion on how indirect interactions appear in the local tradeoffs of each household. I show next that the recursiveness in (1.10) can be undone into an expression that holds a lot more meaning in terms of a household's network position.

An important property of expressions like (1.18) is that, for appropriate values of ϕ — particularly for $\gamma < \frac{1}{\nu_1}$ for ν_1 the largest eigenvalue of $\mathbf{G}$ — we can write them as,

$$\mathbf{S} = \mathbf{I} + \gamma \mathbf{G} + \gamma^2 \mathbf{G^2} + \dots$$

---

such as PageRank and HITS algorithm also refined measures recursively throughout the network.

[43] In the HTIS algorithm, a webpage is given both an authority and a hubness score, with the property that a website's authority is determined by the sum of the hubness scores of other websites it links to, while a website's hubness is determined by the sum of the authorities of websites it is linked by. This implies that each one of this measures is defined recursively at distance two.

In other words, all these recursive measures are often expressed as the sum of all paths starting from i, which can be written as $B_i = \sum\limits_{j \in N} \sum\limits_{q=1}^{\infty} \gamma^q g_{ij}^q$. This framework provides a much more natural way to think of the network statistic containing information on the importance of each household in the general social structure. After all, it measures a household's accessibility by aggregating all locations that can be reached by it at any length. Since we have just seen that the WEPC is a particular version of these general recursive measures, it should not be surprising that it too can be written as the accumulation of paths. Proposition 1.3.2 indeed shows that we can similarly think of the WEPC measure accumulating such paths, subject to the caveats discussed above: that only even paths are accumulated, and that the weighting scheme for each path is particular to that path. Technically, the current setting asks us to solve a modified version of these fixed points on a graph that looks like, $(\mathbf{D} - \mathbf{G}\Psi\mathbf{G})^{-1}$. I show in Proposition 1.3.2 that we can write this as,

$$\mathbf{D}^{-1}\left(\mathbf{I} + \left(\mathbf{A} + \mathbf{A}\mathbf{D}^{-1}\mathbf{A} + \mathbf{A}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-\mathbf{1}}\mathbf{A} + \mathbf{A}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-\mathbf{1}}\mathbf{A}\mathbf{D}^{-1}\mathbf{A} + \ldots\right)\mathbf{D}^{-1}\right)$$

Unlike Bonacich and other types of centralities that weight all paths of a certain length equally, in this scenario each path elicits a specific set of weights, determined by the connectivity of each individual involved in that particular path. The type of weighting scheme in equation 1.3.2 can be thought of in terms of the accumulated local interactions mentioned above. Recall that indirect interactions only represent the concatenation of various direct interactions linked together by the network constraints. This can be gleaned from equation 1.3.2 where the weights $\frac{\Psi}{1+\Psi d_k}$ capture all the households in a given path engaged in direct interactions and the weighs $\frac{1}{d_i}$ capture the connecting household's constraint. This weighting scheme marks a crucial distinction vis-à-vis other measures, in that additional paths does not guarantee an increase in households' network measure.

1.8.4 Individual and Aggregate Volatility

In the context of bilateral risk-sharing in networks a natural concern seeks to distinguish amongst those households that, from their structural position in a broader social arrangement, obtain smoother consumption streams than others. Said differently, we can ask how a household's consumption variability relates to its location in the network. Proposition 1.3.1 defines transfers and therefore establishes, for every household, a particular linear combination of neighboring incomes that enter its consumption. The particularly complicated form of these transfers makes it difficult to obtain an intuitive translation from network position to consumption variance. In the following result I provide a first, partial attempt at ranking household's consumption variance from network characteristics; I am currently working on expanding this into a complete, intuitive ordering of variances on networks. Notice we can write the variance of consumption of household i as ,

$$var\left(c_i\right) = \left(\sigma^2 - \rho\right)\alpha_i'\alpha_i + \rho\mathbf{1}'\alpha_i\alpha_i'\mathbf{1}$$

where $\alpha_i = \left(\alpha_{i1}, \alpha_{i2}, \ldots \alpha_{in}\right)'$ is defined as in (1.9). We want to find the household $i^\star$ such that $var\left(c_{i^\star}\right) \geq var\left(c_i\right)$ for all $i \neq i^\star$. Using the expression for exchanges in (1.9) This next result allows us to rank variances when endowments are independent across households.

Proposition 1.8.1. *Let* $\mathbf{H}_i = \left(diag\left(\mathbf{G}_i\right) - \frac{\Psi}{1+\Psi d_i}\mathbf{G}_i'\mathbf{G}_i\right)$. *If* $\rho = 0$ *and* $\mathbf{H}_i^2 - \mathbf{H}_j^2$ *is positive semi-definite, then* $var\left(c_i\right) > var\left(c_j\right)$

Proof. See Appendix □

The rather technical form of this result precludes a straightforward interpretation on the distribution of consumption volatility. In any case, it provides a testable prediction on individual consumption volatility that is fairly quickly checked in data. I am in the process of extending this result and testing it on the Tsimane' data set.

Aggregate volatility of an entire village is perhaps even more important than distinguishing amongst individual variances. After all, policy considerations can emerge from a deeper understanding of what social arrangements are more conducive to better insurance opportunities. In this respect, we might want to know 1) what type of network is persistently less volatile than another, or 2) what individual, when removed, reduces volatility the most. After some manipulations, I obtain a "useful" form that should allow me to conclude something about which networks are prone to higher aggregate volatility. Specifically, I have that,

$$\sum_i var\left(c_i\right) \propto \mathbf{1}'\left(\mathbf{D} - \mathbf{G}\mathbf{\Psi}\mathbf{G}\right)^{-1}\left(\mathbf{D} - \mathbf{G}\mathbf{\Psi}\mathbf{\Gamma}\mathbf{G}\right)\left(\mathbf{D} - \mathbf{G}\mathbf{\Psi}\mathbf{G}\right)^{-1}\mathbf{1}$$

where $\mathbf{\Gamma}$ is a diagonal matrix similar to $\mathbf{\Psi}$, such that $\mathbf{\Gamma}_{ii} = \frac{2+\Psi d_i}{1+\Psi d_i}$. The familiar quadratic form, although a complicated function of the adjacency matrix, might conceal some useful properties that might allow me to answer these two questions.

1.8.5 Alternative Centrality Measures

As in most network papers that prescribe a centrality-based prediction on behavior, a natural concern is that in fact other similar measures might be as successful in explaining data, so that the predictions of the model are rendered mute. Most times this is dealt with by running a horse race against other measures and showing that the theory's predictions indeed outperform other measures. To do this, I first show that the WEPC centrality presented in this paper is only weakly correlated with other well known global network measures, such as Bonacich centrality or eigenvalue centrality. Correlation with Bonacich is about 0.32 for the Kinship Network and about 0.28 for the Trade Network. Correlation with eigenvalue centrality is about 0.46 for the Kinship Network and about 0.38 for the Trade Network. These values are not large, moreover if we substitute these measures for the WEPC in the expression for the sharing rule in equation (1.9), we obtain non-significant, and even negative, result. Of course, these values are not normalized, so they predict shares that fail to satisfy the constraint (i.e. outside the interval $[0, 1]$ and/or don't sum to one).

A possible objection, therefore, may be that any other linear function of arbitrary network measures that both defines values in $[0, 1]$ and satisfies the feasibility constraint, $\sum\limits_{j}\alpha_{ij} = 1$, would deliver similar results; said differently, one could ask if (1.16) estimates nothing but a simple accounting identity of cross-claims on a network. Indeed, while there exist many such matrices $\mathbf{A}(\mathbf{G})$ that satisfy the feasibility constraints, the estimation procedure could fail to distinguish amongst them, delivering "appropriate" fits to vastly different predictions. To test this I estimate a simple, intuitive alternative to equation (1.9) that only captures local node characteristics. Specifically I consider the possibility that the share of $j's$ endowment consumed by i is determined entirely by the size of $i's$ $k - neighborhood$ relative the total of all $k - neighborhoods$ of all of $j's$ neighbors. This measure captures $i's$ relative importance within $j's$ sphere of influence similar to (1.9), but under a reduced, local notion of importance. We can express this sharing behavior as,

$$\alpha_{ij} = \frac{\left|N_i^k(\mathbf{G})\right|}{\sum\limits_{i} g_{ij}\left|N_i^k(\mathbf{G})\right|}$$

where $|A|$ denotes the cardinality of set A. Indeed it is not difficult to see that this first-order measure provides predictions within the unit interval and satisfies the budget constraint. If the structural estimation of the sharing rule only captures an accounting identity, the estimation of

$$\alpha_{ijt}^{obs} = \beta_1 L_{ij} + \epsilon_{ijt} \tag{1.19}$$

where $L_{ij} = \frac{|N_i(\mathbf{G})|}{\sum\limits_{i} g_{ij}|N_i(\mathbf{G})|}$, should undoubtedly produce $\beta_1 = 1$. I show results for the $2 - neighborhood$ in table 1.11, although similar results hold for all k values tested (all below 10). The results show that indeed β_1 is either not significant, or far from one. Instead, relative neighborhood size fails to correlate with the sharing behavior of the Tsimane' in any reasonable manner that would indicate that other local measures can do as good a job at describing pairwise exchanges.

1.8.6 Estimating the Income Process

Before estimating the income process I control for predictable components. Although the data set contains a number of time-invariant demographic statistics for each household, the only time-varying, household-specific attribute that predicts the level of income is hours worked. Therefore, I run the following first-stage regression of log household income on hours invested in productive activities, together with household and village-time fixed effects,

$$log\left(y_{i,t}\right) = h_{i,t} + \tau_{vt} + \delta_i + \epsilon_{it} \tag{1.20}$$

I choose to allow for household-specific intercepts rather than introducing a long, but still incomplete, list of household demographic traits. I obtain a residual income process for household i from (1.20) that I use as my unpredictable component of income in order to estimate the parameter Ψ.

The next step requires that we define a process for residual income,

$$\begin{aligned} \tilde{y}_{it} &= \eta_{it} + \nu_{it} \\ \eta_{i,t} &= c + \gamma\eta_{i,t-1} + \epsilon_{it} \end{aligned}$$

where $\tilde{y}_{i,t}$ is the residual from a log income regression for an individual i at time t, η_{it} is the persistent component of income and is assumed to follow an AR(1) process, ν_{it} is the transitory component of income, ϵ_{it} is the shock to the persistent component of household income. Finally, $\upsilon_{it} \sim (0, \sigma_\upsilon^2)$, $\epsilon_{it} \sim (0, \sigma_\epsilon^2)$, $\eta_{i,0} \sim (0, \sigma_0^2)$ and are independent of each other for all i and t. The parameter vector to estimate is $\theta = (\gamma, c, \sigma_\epsilon^2, \sigma_\upsilon^2)$. Notice that we don't make any distributional assumptions on the error terms besides defining first and second moments.

Before estimating the vector θ I relate its elements to the parameter of interest, Ψ. Recall that $\Psi = \frac{\mu^2+\rho}{\sigma^2-\rho}$ where μ, σ^2, and ρ represented the mean, variance and common covariance term of the joint distribution of income across households. Given the description on residual income above we can conclude the following relationship between the parameters of θ and the parameters that form the value of Ψ:

$$\begin{aligned} \mu &= \tfrac{c}{1-\gamma} \\ \sigma^2 &= \sigma_\epsilon^2 \tfrac{1}{1-\gamma^2} + \sigma_\nu^2 \\ \rho &= \gamma \tfrac{c}{1-\gamma} \end{aligned}$$

Computing the cross-sectional covariances between period t and period $t+k$ (for all t and k) produces a total of $\frac{T(T+1)}{2}$ distinct moment conditions that relate residual income and distributional parameters.[44] In particular, if we write down the moment $m_{tk}(\theta)$ between agents at time t and $t+k$ we have,

$$\begin{aligned} m_{tk}(\theta) &= \mathbb{E}\left[y_{i,t} \cdot y_{i,t+k}\right] \\ &= \mathbb{E}\left[(\eta_{it} + \nu_{it})(\eta_{i,t+k} + \nu_{i.t+k})\right] \\ &= \begin{cases} \mathbb{E}\left[\eta_{it}^2\right] + \sigma_\nu^2 & if\ k = 0 \\ \gamma^k \mathbb{E}\left[\eta_{it}^2\right] + \left(1-\gamma^k\right)\mathbb{E}\left[\eta_{it}\right]^2 & if\ k > 0 \end{cases} \end{aligned}$$

where

$$\mathbb{E}\left[\eta_{it}^2\right] = \sigma_\epsilon^2 \frac{1}{1-\gamma^2} + \mathbb{E}\left[\eta_{it}\right]^2$$

and $\mathbb{E}[\eta_{it}] = \frac{c}{1-\gamma}$ is the typical expression for the mean of an AR(1) process. The above expressions represent an over-identified system for θ, so moment conditions cannot be solved explicitly.[45] As usual in these cases, we look for the vector θ that minimizes the distance between theoretical moments and their empirical counterparts,

$$\hat{\theta} = \min_{\theta} \left(M(\theta) - \hat{M}\right)' \mathbf{W} \left(M(\theta) - \hat{M}\right)$$

where $M(\theta)$ and $\hat{M}$ stack the moment conditions and the sample analogs respectively, and where $\mathbf{W}$ is a weighting matrix. Following the general trend in the literature I take $\mathbf{W}$ as the identity matrix.[46] The non-linear GMM estimation delivers estimates of $\hat{\theta} = \left(\hat{\gamma}, \hat{c}, \hat{\sigma_\epsilon^2}, \hat{\sigma_\upsilon^2}\right) = (0.981, 0.00932, 0.521 * 10^{-5}, 2.018)$. This represents a negligible shock to the persistent component of income, and a strong persistence parameter. the variance to transitory shocks is about 2. This implies that household income can best be thought mostly of transitory shock with a small common intercept. Using the expressions above we find that $\mu = 0.93$, $\sigma^2 = 2.019$, and $\rho = 0.86$. Together this implies an estimated value of $\Psi = 1.474$. This value is close to other values found using more rudimentary estimates of simpler models in the main text.

[44] Notice the matrix of moment conditions is symmetric (i.e. $m_{tk}(\theta) = m_{kt}(\theta)$) so we only calculate the lower triangular part of the matrix, consistent of $\frac{T(T+1)}{2}$ distinct terms.

[45] In unbalanced panels like this one, moreover, we might estimate less conditions since it might very well happen that no household is present both in period t and $t+k$. Formally, we estimate the available moment conditions defined as, $\mathbb{E}\left[\lambda_{i,t,k}(\hat{m}_{tk} - m_{tk}(\theta))\right]$ where $\lambda_{i,t,k}$ equals 1 only if i is present at t and $t+k$, and is 0 otherwise, and where $\hat{m}_{t,k} = \frac{1}{I_{t,k}} \sum_{i=1}^{I_{t,k}} y_{i,t} y_{i,t+k}$, with $I_{t,k} = \sum_i \lambda_{i,t,k}$.

[46] Altonji and Segal (1996) show that the optimal weighting matrix introduces significant small sample bias. They study the small sample properties of the GMM estimator with several alternative weighting matrices and recommend using the identity matrix.

1.9 Proofs

Lemma 1.9.1. *Under quadratic utility, there exists no ex-ante conflict between efficiency and equity. If* $L(\mathbf{C})$ *is a network component, the ex ante Pareto-efficient risk-sharing arrangement among agents in* $\mathbf{C}$ *minimizes expected cross-sectional variability in consumption. Formally,*

$$max \sum_{i \in \mathbf{C}} \mathbb{E}u(c_i) = min \mathbb{E} \sum_{i \in \mathbf{C}} (c_i - \bar{c})^2$$

and corresponds to solving the following mean and variance relation,

$$min \sum_{i \in \mathbf{C}} var(c_i) + \mathbb{E}(c_i)^2 \quad \textit{subject to} \quad \sum_{i \in \mathbf{C}} c_i(\omega) = \sum_{i \in \mathbf{C}} y_i(\omega) \quad \textit{for every state } \omega \tag{1.21}$$

Proof of Lemma Lemma 1.9.1

Consider the minimization of expected cross-sectional variability in consumption defined as $\mathbb{E}\left[\sum_i (c_i - \bar{c})^2\right]$, where $\bar{c} = \frac{1}{N}\sum_i c_i$ represents the average consumption. This is equivalent to minimizing $\sum_i c_i^2 - \frac{1}{n}\left(\sum_i c_i\right)^2$. Since we have that $\sum c_i = \sum y_i$ the second term drops out of the optimization problem. As a result, the problem reduces to minimizing $\mathbb{E}\sum c_i^2$. Notice that the welfare problem under quadratic utility corresponds to minimizing $\mathbb{E}\sum_i c_i - \frac{1}{2}\gamma c_i^2$. Distributing the sum and imposing the feasibility condition that $\sum c_i = \sum y_i$, implies that the planner essentially minimizes $\mathbb{E}\sum c_i^2$. For the second statement, notice that $\sum_i \mathbb{E}c_i^2 = \sum_i var(c_i) + \mathbb{E}(c_i)^2$ follows from the definition of variance and the linearity of the expectations operator.

Proof of Proposition 1.3.1

Proof. Recall the optimality conditions given in equation (1.7), that characterize the dependency of shares across the network,

$$\alpha_{ij} = g_{ij}(\Lambda_j - \Psi \sum_k g_{ik}\alpha_{ik}) \tag{1.22}$$

where $\Psi = \frac{\mu^2+\rho}{\sigma^2-\rho}$, $\Lambda_j = \frac{\lambda_j}{2(\sigma^2-\rho)}$ and λ_j is the constant to the constraint on $j's$ outgoing shares. We will rewrite these in matrix form by defining first the vector of i's incoming shares $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{in})'$ and the vector of constraint multipliers $\mathbf{\Lambda} = (\Lambda_1, \Lambda_2, \ldots, \Lambda_n)'$. Let $\mathbf{G}_i$ represent the i^{th} row of $\mathbf{G}$ and let $\hat{\Psi} = -\Psi$. Then, equation (1.22) can be written as,

$$\left(\mathbf{I} - \hat{\Psi}\mathbf{G}_i'\mathbf{G}_i\right)\alpha_i = diag\left(\mathbf{G}_i'\mathbf{G}_i\right)\mathbf{\Lambda}$$

and, because the matrix on the left-hand-side is full-rank, we can offer the following formulation,

$$\alpha_i = \left(\mathbf{I} - \hat{\Psi}\mathbf{P}_i\right)^{-1} diag\left(\mathbf{P}_i\right)\mathbf{\Lambda} \tag{1.23}$$

where I set $\mathbf{P}_i = \mathbf{G}_i'\mathbf{G_i}$ for ease of notation. Now, if the value of $\hat{\Psi}$ is such that $\hat{\Psi} < \frac{1}{\nu_{max}}$ where ν_{max} is the leading eigenvalue of $\mathbf{P}_i$ then we can write the following relation,

$$\left(\mathbf{I} - \hat{\Psi}\mathbf{P}_i\right)^{-1} = \sum_{k=0}^{\infty}\hat{\Psi}^k\mathbf{P}_i^k$$

The condition on $\hat{\Psi}$ holds for all matrices $\mathbf{P}_i$ since, by the Perron-Frobenius theorem, $0 < \underset{j}{min}\sum_k\mathbf{P}_{i;kj} \leq \nu_{max} \leq \underset{j}{max}\sum_k\mathbf{P}_{i;kj}$ and $\hat{\Psi} < 0$. Now, because the matrix $\mathbf{P}_i$ is idempotent up to a scalar corresponding to the degree of i — i.e. $\mathbf{P}_i^k = d_i^{k-1}\mathbf{P}_i$ — then we can simplify the above series in the following way,

$$\sum_{k=0}^{\infty}\hat{\Psi}^k\mathbf{P}_i^k = \mathbf{I} - \frac{\Psi}{1+\Psi d_i}\mathbf{P}_i$$

Finally, it can be easily checked that $\mathbf{P}_i \cdot diag\left(\mathbf{P_i}\right) = \mathbf{P}_i$, which means we can rewrite equation (1.23) as,

$$\alpha_i = \left(diag\left(\mathbf{P}_i\right) - \frac{\Psi}{1+\Psi d_i}\mathbf{P}_i\right)\mathbf{\Lambda} \tag{1.24}$$

where we still have to solve for $\mathbf{\Lambda}$ to obtain a closed form solution of α_i. To do this notice that 1.24 allows us to rewrite $j's$ constraint as,

$$1 = \sum_i \alpha_{ij} = d_j\Lambda_j - \sum_i g_{ij}\left(\frac{\Psi}{1+\Psi d_i}\sum_k g_{ik}\Lambda_k\right)$$

which implies that

$$\Lambda_j = \frac{1}{d_j} + \frac{1}{d_j}\sum_{i,k} g_{ji}g_{ik}\frac{\Psi}{1+\Psi d_i}\Lambda_k$$

□

Proof of Proposition 1.3.2

Proof. If we impose the feasibility constraints on the vector equation (1.24) we obtain that,

$$\mathbf{1} = \sum_i \left(diag\left(\mathbf{P}_i\right) - \frac{\Psi}{1+\Psi d_i}\mathbf{P}_i \right) \mathbf{\Lambda}$$

where $\mathbf{1}$ is an $n - vector$ of ones. The properties of $\mathbf{P}_i$ means we can rewrite equation (1.24) as a function of the original matrix $\mathbf{G}$ in the following way,

$$\alpha_i = \left(diag\left(\mathbf{P}_i\right) - \frac{\Psi}{1+\Psi d_i}\mathbf{P}_i \right) \left(\mathbf{D} - \mathbf{G}\mathbf{\Psi}\mathbf{G}\right)^{-1} \mathbf{1}$$

where $\mathbf{\Psi}$ is a diagonal matrix with $\mathbf{\Psi}_{ii} = \frac{\Psi}{1+\Psi d_i}$ and $\Psi_{ij} = 0$ for all $i \neq j$. This provides a closed form solution of the constrained-efficient flows on any given network.

Finally, to arrive at the result we solve for the inverse matrix above as a series of powers of $\mathbf{G}$. The following formulation allows us to do so

$$\left(\mathbf{D} - \mathbf{G}\mathbf{\Psi}\mathbf{G}\right)^{-1} = \left(\mathbf{D}^{\frac{1}{2}}\left(\mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{G}\mathbf{\Psi}\mathbf{G}\mathbf{D}^{-\frac{1}{2}}\right)\mathbf{D}^{\frac{1}{2}}\right)^{-1} = \mathbf{D}^{-\frac{1}{2}}\left(\mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{G}\mathbf{\Psi}\mathbf{G}\mathbf{D}^{-\frac{1}{2}}\right)^{-1}\mathbf{D}^{-\frac{1}{2}}$$

the middle term being inverted can be expressed as a geometric series as long as ?? . Provided this is so, the following relation holds,

$$\left(\mathbf{D} - \mathbf{G}\mathbf{\Psi}\mathbf{G}\right)^{-1} = \mathbf{D}^{-1} + \mathbf{D}^{-\frac{1}{2}}\sum_{k=1}^{\infty}\left(\mathbf{D}^{-\frac{1}{2}}\mathbf{G}\mathbf{\Psi}\mathbf{G}\mathbf{D}^{-\frac{1}{2}}\right)^k\mathbf{D}^{-\frac{1}{2}}$$

which can be understood as accumulating weighted even powers of the adjacency matrix as follows,

$$\mathbf{D}^{-1}\left(\mathbf{I} + \left(\mathbf{A} + \mathbf{A}\mathbf{D}^{-1}\mathbf{A} + \mathbf{A}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-\mathbf{1}}\mathbf{A} + \mathbf{A}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-\mathbf{1}}\mathbf{A}\mathbf{D}^{-1}\mathbf{A} + \ldots\right)\mathbf{D}^{-1}\right)$$

where we set $\mathbf{A} = \mathbf{G}\mathbf{\Psi}\mathbf{G}$ to ease notation. So the sharing rule weights all even powers between i and j through the matrix $\hat{\mathbf{\Psi}}$ that appears between the product of $\mathbf{G}$. This can be surmised in the above expression and can be written as follows. Consider the set of all paths of length q between i and j under $\mathbf{G}$ as

$$\Pi_{ij}^q\left(\mathbf{G}\right) = \left\{\left\{i_0, i_1, i_2, \ldots i_q\right\} \mid i_0 = i,\ i_q = j\ and\ g_{n,n+1} = 1\ for\ n = 0, 1, \ldots q-1\right\}$$

for every $\pi_{ij} \in \Pi_{ij}^q\left(\mathbf{G}\right)$ let $W\left(\pi\right)$ define the weights associated to this path. It is not difficult to see that,

$$W\left(\pi_{ij}\right) = \frac{1}{d_i}\frac{\mu^2+\rho}{\sigma^2-\rho+\left(\mu^2+\rho\right)d_{i_1}}\frac{1}{d_{i_2}}\frac{\mu^2+\rho}{\sigma^2-\rho+\left(\mu^2+\rho\right)d_{i_3}}\cdots\frac{1}{d_j}$$

Finally, let M_i represent the i^{th} element of the vector $(\mathbf{D} - \mathbf{G\Psi G})^{-1}\mathbf{1}$. Then, $M_i = \frac{1}{d_1} + \sum_j M_{ij}$, where

$$M_{ij} = \sum_{q=1}^{\infty} \sum_{\pi \in \Pi_{ij}^{2q}} W(\pi_{ij})$$

$\square$

Proof of Proposition 1.3.3

Proof. Assume on the contrary that $\alpha_{ij} \neq \frac{1}{d}$. A regular network has the property that $M_i(\Psi, \mathbf{G}_{reg}) = M_j(\Psi, \mathbf{G}_{reg}) = M(\Psi, \mathbf{G}_{reg})$ for all i and j. In that case, we can write the assumption that equal sharing is not the solutions as,

$$\alpha_{ij}^{\star} = M(\Psi, \mathbf{G}_{reg}) \left(1 - \frac{\Psi d}{1 + \Psi d}\right) \neq \frac{1}{d}$$

which implies that

$$M(\Psi, \mathbf{G}_{reg}) \neq \frac{1}{d} + \Psi$$

Using the result from Proposition X, we can also write $M(\Psi, \mathbf{G}_{reg})$ as,

$$M(\Psi, \mathbf{G}_{reg}) = \frac{1}{d} + \sum_{q \in \mathbb{N}} \sum_{j \in N} \sum_{\pi_{ij} \in \Pi_{ij}^{2q}} W(\pi_{ij})$$

where $W(\pi)$ corresponds to a particular weighting scheme for paths between i and j. So if equal sharing is not the solution for a regular network, then $\Psi \neq \sum_{q \in \mathbb{N}} \sum_{j \in N} \sum_{\pi_{ij} \in \Pi_{ij}^{2q}} W(\pi_{ij})$. I show next that in fact they are equal.

By the symmetry of the complete network, we know that, $W\left(\pi_{ij}^{2q}\right) = \left(\frac{1}{d}\right)^{q+1} \left(\frac{\Psi}{1+\Psi d}\right)^{q}$ for any path of length $2q$ between i and j. Finally it is just a matter of finding how many such paths there are. Let $\Pi_j^q = \bigcup_i \Pi_{ij}^q$ and Π_{ij}^q is the set of all paths of length q between i and j. Define $|A|$ as the cardinality of set A. Then we can write that,

$$\sum_{q \in \mathbb{N}} \sum_{j \in N} \sum_{\pi_{ij} \in \Pi_{ij}^{2q}} W(\pi_{ij}) = \frac{1}{d} \sum_{q \in \mathbb{N}} \left|\Pi_j^{2q}\right| \left(\frac{1}{d}\right)^q \left(\frac{\Psi}{1 + \Psi d}\right)^q$$

The value of $\left|\Pi_j^{2q}\right|$ corresponds to the number of paths of length $2q$ starting from j. All paths of length $2q$ contain $2q+1$ nodes, so this is equivalent to the number of ways to assign d values to each of the $2q$ remaining values (once we fix j). This is a standard assignment problem in combinatorics and the solution is well known and equal to d^{2q}. This means that we can write the following,

$$\frac{1}{d}\sum_{q=1}^{\infty}d^{2q}\left(\frac{1}{d}\right)^{q}\left(\frac{\Psi}{1+\Psi d}\right)^{q}=\frac{1}{d}\sum_{q=1}^{\infty}\left(\frac{d\Psi}{1+\Psi d}\right)^{q}=\Psi$$

where the second equality comes from the convergence of the geometric series. This contradicts the original assumption that $\alpha_{ij}\neq\frac{1}{d}$ and proves the result. □

Let us define a household's neighborhood centrality as a weighted average of all of its neighboring centralities as follows,

$$M_{N_i}\left(\mathbf{G},\Psi\right)=\frac{\Psi}{1+\Psi d_i}\sum_{k}g_{ik}M_k\left(\mathbf{G},\Psi\right) \tag{1.25}$$

This term appears in the constrained-efficient solution to all of $i's$ incoming shares and weights the total position of all of $i's$ neighbors by the connectivity of i. The following two lemmas derive properties of this neighborhood centrality and are used in a couple of proofs in the paper.

Lemma 1.9.2. *The WEPC of agent i can be expressed as a function of the neighborhood centralities of all its neighbors. In other words,*

$$M_i\left(\mathbf{G},\Psi\right)=\frac{1}{d_i}\left(1+\sum_{k}g_{ik}M_{N_k}\left(\mathbf{G},\Psi\right)\right)$$

where $M_{N_k}=\frac{\Psi}{1+\Psi d_i}\sum_{k}g_{ik}M_k$

Proof. Recall the two-step recursive expression of WEPC from equation 1.10,

$$M_i\left(\Psi,\mathbf{G}\right)=\frac{1}{d_i}\left(1+\sum_{l,k}g_{ik}g_{kl}\frac{\Psi}{1+\Psi d_k}M_l\left(\Psi,\mathbf{G}\right)\right) \tag{1.26}$$

the second term in brackets above can be rewritten as

$$\sum_{j}g_{ij}\frac{\Psi}{1+\Psi d_j}\sum_{k}g_{jk}M_k\left(\mathbf{G},\Psi\right)$$

and using the definition of M_{N_j} in equation 1.25 we obtain the expression. □

Lemma 1.9.3. *let $n(C)$ equal the total number of households in any connected component C, then the average neighborhood centrality over that component is always equal to Ψ. Formally,*

$$\sum_{i \in C} M_{N_i} = \Psi n(C)$$

for all C.

Proof. Using the budget constraint and our constrained-efficient solution, $\alpha^{\star}_{ij}$ in equation (1.9), we have that

$$n(C) = \sum_{i \in C} \sum_{j} g_{ij} \alpha_{ij} = \sum_{i \in C} \sum_{j} g_{ij} \left(M_j - M_{N_i}\right) = \sum_{i \in C} \sum_{j} g_{ij} M_j - \sum_{i \in C} M_{N_i} \sum_{j} g_{ij}$$

Now using the definition of M_{N_i} in equation 1.25, we have the following relationship

$$n(C) = \sum_{i \in C} M_{N_i} \left(\frac{1 + \Psi d_i}{\Psi} - d_i \right)$$

rearranging we get the result. □

Lemmas 2 and 3 together imply the following useful result,

Lemma 1.9.4. *The WEPC of a household i that is connected to all other households in a component $C_i(\mathbf{G})$ is constant across all networks and equal to*

$$M_i(\mathbf{G}, \Psi) = \frac{1}{d_i} + \Psi$$

for all $\mathbf{G}$ whenever $d_i = |C_i(\mathbf{G})|$.

Proof. Straightforward. □

Proof of Proposition 1.3.4

Proof. let h represent the center (or hub) of the star and s represent the peripheral households (or spokes). Then, define the transfer from h to s as $\alpha_{sh} = M_h(\mathbf{G}, \Psi) - M_{N_s}(\mathbf{G}, \Psi)$ using the constrained-efficient solution of equation 1.9 and the definition of neighborhood centrality in equation (1.25). By Lemma 2 we can rewrite this as

$$\begin{aligned} \alpha_{sh} &= \tfrac{1}{d_h}\left(1 + (n-1) M_{N_s} + M_{N_h}\right) - M_{N_s} \\ &= \tfrac{1}{d_h}\left(1 + M_{N_h}\right) + M_{N_s}\left(\tfrac{n-1}{d_h} - 1\right) \end{aligned}$$

and by lemma 3 we have that

$$M_{N_s} = \frac{n\Psi - M_{N_h}}{n-1}$$

which allows us to express α_{sh} only as a function of M_{N_h}.

Finally, since h by definition is connected to all other players in the network, we can use corollary 1 together with the constraint on the shares sent by h to obtain the following useful relationship between M_{N_h} and α_{sh}

$$1 - (n-1)\,\alpha_{sh} = \frac{1}{d_h} + \Psi - M_{N_h}$$

this implies that we can express α_{sh} uniquely as a function of parameters, Ψ, n and d_h, as follows,

$$\alpha_{sh} = \frac{1}{d_h}\left(\frac{1}{d_h} + \Psi + (n-1)\,\alpha_{sh}\right) + \frac{n\Psi - \frac{1}{d_h} + \Psi + (n-1)\,\alpha_{sh} - 1}{n-1}\left(\frac{n-1}{d_h} - 1\right)$$

rearranging we get that

$$\alpha^{\star}_{sh} = \frac{1}{1 + \left(\frac{n}{2} + 1\right)\Psi}\left(\frac{1}{d_h} + \Psi\right)$$

this proves the result for transfers from h to s. Similar steps show that Proposition 1.3.4 also holds for transfers from s to h □

Proof of Proposition 1.8.1

Proof. We write down the variability in consumption as $var\,(c_i) \propto \alpha_i'\alpha_i$ where

$$\alpha_i = \left(diag\,(\mathbf{G}_i) - \frac{\Psi}{1+\Psi d_i}\mathbf{G}_i'\mathbf{G}_i\right)(\mathbf{D} - \mathbf{G\Psi G})^{-1}\,\mathbf{1}$$

let $\mathbf{H}_i = \left(diag\,(\mathbf{G}_i) - \frac{\Psi}{1+\Psi d_i}\mathbf{G}_i'\mathbf{G}_i\right)$. Then we have that,

$$var\,(c_i) = \left((\mathbf{D} - \mathbf{G\Psi G})^{-1}\,\mathbf{1}\right)'\mathbf{H}_i'\mathbf{H}_i\,(\mathbf{D} - \mathbf{G\Psi G})^{-1}\,\mathbf{1} \qquad (1.27)$$

$\mathbf{H}_i$ is symmetric so $\mathbf{H}_i'\mathbf{H}_i = \mathbf{H}_i^2$. Equation (1.27) is a quadratic form on $\mathbf{H}_i^2$. The result follows from standard properties of quadratic forms. □

Chapter 2

REGIME CHANGE IN LARGE INFORMATION NETWORKS

Joint with Joan de Martí, Universitat Pompeu Fabra

2.1 Introduction

Global games of regime change describe coordination games of incomplete information in which the status quo- i.e. a currency peg, a bank's balance sheet, or a political regime, - is abandoned when a sufficient fraction of the population attacks it. So far all previous work has treated the population as an infinite, homogeneous, mass of individuals, each with a private noisy signal of the fundamentals, without any regard for the potential patterns of communication that admittedly exist amongst individuals.[1] In this paper we propose an approach to introduce networks of communication within this class of games so as to draw some implications the network architecture might impose on the outcomes of the game.

Our model assumes a population of individuals connected according to some social network, each with an independent noisy signal on the underlying strength of the status quo. This fundamental value essentially describes the minimal fraction of individuals that are necessary for regime change. Individuals engage in one round of communication in which each agent truthfully shares her signal to her immediate neighbors. More connected agents, then, will receive a larger set of signals and attain a more precise observation of the fundamental. Each agent can then either choose to attack or not attack the status quo. Attacking can yield a positive payoff- if regime change is successful- or a negative payoff- if it is not. Not attacking always yields a 0 payoff. Payoffs are discontinuous both in the state and other players' actions so that the payoff structure does not exhibit strategic complementarities per se. The equilibrium is defined in threshold strategies, whereby a player chooses to attack if she observes a signal realization below some cutoff value.

To avoid dealing with intractable correlation effects that plague the system for a generic class of networks, we focus instead on networks where each agent's neighborhood is sufficiently small relative to the entire population, such that these correlation effects are negligible. In the end, we are left with an infinite population split into various partitions of varying connectivity (or degree) and therefore of varying precision in their private information. We identify the degree distribution of the network as a crucial determinant of individual strategies and aggregate behavior, and compare the equilibrium outcomes for various degree distributions ranked according to different measures of Stochastic Dominance. Moreover, we present some results on uniqueness of equilibrium that relate meaningfully to previous models with a homogeneous population.

In the case of a diffuse common prior (i.e. no public information) we prove that the probability of successful regime change does not vary with the degree dis-

[1] See for instance: Angeletos, Hellwig and Pavan (2006, 2007), Angeletos and Werning (2006), and Edmond(2011)

tribution: As remarked by Vives (2005) and others, maximal strategic uncertainty with respect to others' behavior induces "flatter" best response so that individuals are less concerned with the aggregate composition of the population. It turns out that each individual responds to common beliefs about aggregate behavior by selecting a threshold strategy commensurate with the probability that their private signal is extreme. This implies that players with smaller tails in their private signals compensate by selecting larger thresholds at exactly the proportions that offset any differences across individuals. As a result, everyone in the population (regardless of their degree) is equally propense to attack the regime, so that altering the proportion of highly connected individuals does not affect the aggregate share of belligerents. This does not mean, however, that all equilibria are identical. While the success probability remains unchanged, the size and composition of attacks conditional on success/failure indeed respond to the varying connectivity of the population. Populations with a larger (smaller) share of highly connected individuals will exhibit larger (smaller) successful attacks and smaller (larger) failed attacks.

Introducing public information has important implications for equilibrium outcomes: the degree distribution becomes an instrumental determinant of the success probability, albeit with surprising considerations. It turns out that the cost endured by attackers in the event of failure defines the direction of the comparative statics and sheds some light into the strategic interactions of a population with varying informativeness of the relevant state of the world. When costs are sufficiently low, increasing the connectivity of the population (by way of First Order Stochastic Dominance) actually lowers the success probability. The opposite is true for sufficiently high costs. This surprising result reflects the fact that less informed individuals place less weight on their private signal and respond more to aggregate behavior. As costs fall less connected individuals increase their propensity to attack far more than more connected agents. Moreover this difference increases and becomes arbitrarily large as the costs of revolt approach zero.

The model attains uniqueness for a larger set of parameter values than under a homogeneous population (i.e. without incorporating a communication network). This is explained by noticing that the presence of varying degrees in this model essentially translates into a convex combination of weights placed on the public signal. As more weight is placed on high degree individuals (who in turn pay less attention to the public signal) then we retain uniqueness for smaller public-signal variances than the previous models allowed.

Finally the paper provides a methodological contribution to the literature by identifying sufficient conditions on network sparseness that allow for an approximation of large networks by an infinite population partitioned by their connectivity. Not only do we gain tractability and insight but we are able to isolate the effect of connectivity on informativeness by disregarding the local correlations induced by the network. We show these conditions are quite general and applicable to a wide range of network global games and we hope these can be useful for future research in the area.

2.2 Literature

Our paper mainly contributes to two general strains of literature: that on global games pioneered by Carlsson & Van Damme (1993) and Morris & Shin (2002), and on network theory. The paper essentially extends the static version of the global game of regime change put forth by Angeletos et al. (2007) in considering the role played by the exchange of private information within a network. Indeed the presence of local communication lays the ground for a number of additional questions on the role of connectivity in coordination not adressed in the basic model of Angeletos et al. Others, such as Edmond (2011) and Bueno de Mesquita (2011), have similarly dealt with discrete action global games in large populations, tackling diverse aspects such as the possibility of strategic action by the status quo, but neglecting the impact of connectivity in shaping posterior beliefs and equilibrium actions when coordinating aggregate behavior. Our paper shares with Hellwig (2002) and Angeletos & Werning (2006) the study of the interaction of private and public information in determining uniqueness of coordinating equilibria.

More precise efforts to model regime change with heterogeneous agents, such as Chwe(2000) and Guimaraes and Morris (2005), distinguish agents' possibly different action spaces contingent on types or their network position in sequential action games, but similarly make no effort to model varying connectivity and its role in the sharing of private information. Moreover, the latter focus on continuous action spaces, which disregards some of the inherent complications of correlated signals in threshold equilibria with finite players. We show that these considerations are not innocent, and that the strategic impact of connectivity on equilibrium outcomes is far from obvious. Most recently, an attempt to adress the role of networks by Dahleh et al. (2012) has provided a partial characterization for finite populations. Their results are silent to non-regular network structures and their focus on multiplicity is strangely at odds with the solution concept employed. Finally, Hassanpour (2010) provides an applied study that underscores the empirical importance of these types of models in recent experiences with large scale coordinated attacks on regimes. His theoretical model, however, allows for continuous belief updating a la deGroot, which fundamentally undermines the impact of limited local communication in coordinated attacks.

The network literature includes a rich tradition of modelling communication. Bloch & Dutta (2009) propose a model of network formation where agents can choose to invest in links of communication with varying degrees. Their work establishes stable and efficient architectures, rather than exploring the impact of exchange on games of coordination. Galeotti et al. (2010) consider a model where individuals are partially informed about the structure of the social network and provide results characterizing how the network structure shape individual behaviour and payoffs. Finally, Hagenback & Koessler (2009) consider strategic communication in networks by modelling a cheap talk communication stage within networks .

We assume in this paper that communication is truthful and limited to direct network neighbors. This is a modelling assumption shared with Calvó-Armengol & de Martí (2007, 2009) that deals directly with the role of communication networks in a class of global games with continuous quadratic payoffs. They provide a knowledge index that essentially compounds higher-order expectations in order to map beliefs into actions. Regime change models, however, are discrete action games that require a consideration of the entire posterior distributions. As such, a new approach that resolves the underlying correlations is warranted.

The recent paper by Barberà and Jackson (2016) characterizes the set of monotone threshold equilibria for a discrete version of a similar collective action game, with an infinite number of players but with a finite number of possible signals. The analysis reveals non-linearities in the participation decisions and non-monotonicities in the participation rate if players can receive more signals. Their model assumes that all individuals receive the same number of signals and, in network terms, this could represent the case of a regular network where all players have exactly the same number of connections.

2.3 Model

This section develops a model that builds on the original model of regime change by Angeletos et al. (2007) by introducing a stage of communication in networks loosely inspired by the work of Calvó-Armengol & de-Martí (2007). By approximating large networks with an infinite population the model essentially introduces heterogeneous variances to the original analysis by Angeletos et al. which allows for a different set of comparative statics exercises.

2.3.1 Actions, Payoffs and Network

There is a population N of individuals connected according to some network G to be specified below. Each agent takes an action $a_i \in \{0, 1\}$ where $a_i = 1$ will represent an attack on the status quo. The payoffs are as follows:

	Regime Change $(A \geq \theta)$	Failure $(A < \theta)$
$a_i = 1$	$1 - c$	$-c$
$a_i = 0$	0	0

where θ is some exogenous parameter, $A = \frac{1}{N}\sum_{n} a_i$ is the proportion of the population that chooses to attack the status quo and $c \in (0,1)$ represents the cost of attack.

There is a network G that captures the communication process. We assume $g_{ij} = g_{ji}$ (undirected channels) and $g_{ij} \in (0,1)$, with $g_{ij} = 1$ meaning that agents i and j communicate with each other. For computational simplicity me let $g_{ii} = 1$. We define the neighborhood of i as $N_i = \{j \in N | g_{ij} = 1\}$ and we denote its cardinality as the degree (that is, $d_i = |N_i|$) . Let N_d represent the number of individuals in the network with degree d. Finally, let $D = \underset{i=1,\ldots N}{max} \{d_i\}$

2.3.2 Information, Communication, and Belief Formation

Agents have a common prior belief θ_0 with a corresponding variance σ_0^2 (can be diffuse or not). Each agent then receives an i.i.d. signal $x_i = \theta + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma^2)$. There is one round of truthful communication in which each agent transmits their signal to his/her neighbors. Alternatively you could think of the network as describing a technology whereby being connected to someone implies that their private information is readily available. In any case, after the dispersal of information, each agent i contains a vector composed of d_i independent signals with which to update beliefs about the strength of the status quo, θ, using Bayes' rule. Clearly, in the measure that agents' signal vectors overlap their posterior beliefs about fundamentals are correlated, meaning that agents that share most of their neighbors will also have very similar information; in the limit, a complete network corresponds to a situation of common knowledge.

Formally, agent i forms the following posterior distribution of θ conditioned on the entire vector of private signals,

$$\theta \mid \mathbf{x}_i \sim N\left(\frac{\sigma^2}{\sigma^2 + d_i\sigma_0^2}\theta_0 + \frac{\sigma_0^2}{\sigma^2 + d_i\sigma_0^2}\langle \mathbf{x_i}, \mathbf{1_{d_i}} \rangle, \frac{\sigma^2\sigma_0^2}{\sigma^2 + d_i\sigma_0^2}\right) \tag{2.1}$$

where $\langle \cdot \rangle$ represents the dot product of two vectors and $\mathbf{1_{d_i}}$ is a vector of ones of dimension d_i. This updating process is instrumental to equilibrium since it refines agents' beliefs about the probability of success (holding everyone's equilibrium behavior fixed) and therefore allows players to obtain their optimal best response.

Now, since we are dealing with payoffs that are discontinuous in actions, our equilibrium must be defined in monotone cutoff strategies. In other words, players will choose to attack if their vector of signals is "below" some optimal frontier, to be specified below. As will become clear later, the shape of this frontier responds directly to the strength of correlation effects: when correlations are strong (i.e. in the case of finite networks) the frontier will be highly non-linear, while negligible correlation effects (in the case of infinite population) will lead to a linear frontier and allow for a tractable solution to the model.

2.3.3 Strategies

Because this is a game of incomplete information and finite actions, the equilibrium is defined in terms of threshold strategies on private signals such that if player i receives a vector of signal realization below some frontier she then takes action 1, and conversely takes action 0 if she observes a private signal above said bounds. Of course the value and shape of the threshold will make player i indifferent between the two available actions, which given our payoff structure implies that $P\left(A \geq \theta \mid \mathbf{x_i}\right) = c$ where the posterior probability is calculated by Bayes' rule as indicated in equation (2.1). As such, player i will choose to attack the status quo for all signal vectors in the set,

$$B_i(A) = \{\mathbf{x_i} \mid P\left(A \geq \theta \mid \mathbf{x_i}\right) \leq c\} \tag{2.2}$$

where we can now express the equilibrium size of attack, A, in terms of equilibrium threshold strategies as,

$$A = \frac{1}{n}\sum_{n} a_i = \frac{1}{n}\sum_{i}^{n} \mathbf{1}_{\{\mathbf{x_i} \in B_i\}} \tag{2.3}$$

where $\mathbf{1}$ here represents the indicator function. The set $B_i(A)$ contains all vectors of signal realizations that induces agent i to attack the status quo. Recall, however, that agent i knows some of these realizations are observed, as well, by other neighboring players. This means that if she observes, say, a very high realization for the signal of an influential neighbor, she can infer many others have observed it as well. Then she might expect that the share of attackers, A, will be low, and will therefore need to observe much much lower realizations in her remaining signals such that she believes a low A is sufficient for success. In other words, the upper boundary of the set B_i can take on strange, non-linearities corresponding to the network topology and ensuing correlations. The following section works through an example of this sort, but in general these considerations complicate the analysis considerably and it is precisely what we try to avoid with an infinite population approach.

The size of attack, A, is essentially a binomial random variable (or a sum of bernoulli random variables). But the correlations implicit in the network structure guarantees these are not independent bernoulli draws. This means that a player's position in the network admittedly affect her expectation about A. To see this notice that players are correlated amongst individuals up to two links apart (I am correlated with my friend's friend since both of us received my friend's signal). This guarantees that my belief about the possible states in which, say, $A = \frac{1}{3}$ is not the same as someone else with a different set of neighbors (and thus a different set of correlations). This seems to imply that any two individuals (even of the same degree) can arrive at radically different beliefs about the possible value of A. Of course, common knowledge of the network structure would guarantee that every player knows everyone's correlation structure when calculating their thresholds, so that in equilibrium every player would know each other's threshold strategies perfectly and in fact would end up calculating the exact same distribution for A.[2]

With all this we can then formally define a Bayesian Nash Equilibrium of this model as a situation where given everyone else's strategies, B_{-i}, player i forms beliefs about the size of attack A, and given A and the vector of incoming signals, $\mathbf{x_i}$, player i chooses optimal strategy B_i.

Definition. An equilibrium corresponds to a complete strategy profile $B = (B_1 \ldots, B_n)$ such that equations (2.2) and (2.3) hold simultaneously.

Calculating this equilibrium, however, implies going through every possible network structure, calculating each player's correlation to obtain a distribution of A, and then for each possible realization of A compute the equilibrium threshold for each player. The computational difficulty explodes as N increases.[3] In order to circumvent these issues and arrive at a tractable model, this paper formally approximates a large network with an infinite population by showing that these correlation effects become negligible in the limit. Before we proceed, however, it is worth exploring the types of equilibria we might expect in finite networks.

2.4 A Finite Network Example

In this section we seek to underline the correlation effects that guide equilibrium behavior in finite networks by working through an amenable example with three

[2]This model deals with very large networks and the idea that the entire geometry is somehow known by everyone is untenable. Fortunately it is not necessary. It suffices that players all know the degree distribution of the network and that they have a common prior belief about the likelihood of each particular architecture that is possible given this degree distribution.

[3]It is important to note that in this scenario we would be left with a distribution of finite support, not a fixed share, which would preclude any clear cut prediction about which values of θ guarantee success and which guarantee failure in equilibrium. Equilibrium in that case would assign to each value of θ a probability of success.

players. Given the difficulties in solving the model for a general finite network of size n, the next section will show that these correlation effects disappear for large (and sufficiently sparse) networks; this will allow us to solve the model asymptotically. In any case, we will stress that the following intuition is analogous to the behavior we might expect for any large, but finite, network.

In order to solve the model with finite agents we must consider the possibility that players a priori will not weight all signals equally when defining equilibrium strategies- i.e. the upper boundary of set B_i is not linear. In the infinite population scenario agents simply take an average of all signals in anticipation of the negligible impact of correlations. This in turn means that the position in the network turns out to be irrelevant (all that matters is the degree of each player) and all incoming signals are equally useful when calculating posteriors. But with finite agents the position in the network is crucial in determining equilibrium strategies. As an example, a player might choose to weight one of his neighbor's signals more if this neighbor happens to be in a privileged position- i.e. a "hub"'s signal gets read by a large share of the total population. This implies that depending on others' equilibrium strategies, best response functions may take on different shapes corresponding to different weights placed on each signal. Formally, best responses here are not a linear mapping of all incoming signals (as is the case for infinite players) but instead will be shaped by the relative position of each neighbors who transmitted each signal.

To begin fixing ideas, consider the game described above played by three agents (call them a, b, and c) connected in a star-like network as shown in figure 1. It should be clear from the communication process that after the signals have dispersed through the network all agents are correlated to each other, and in particular agents b and c are correlated vis-a-vis a's signal.[4] An equilibrium here corresponds to a vector of equilibrium strategies $B = (B_a, B_b, B_c)$ defined as in equation (2.2) and where the share of agents that attack, A, is given by the relation $A(\theta) = \frac{1}{3}\sum_i \mathbf{1}_{\{\mathbf{x_i} \in B_i^*\}}$.

As the hub, a's private information will take on a heftier share of others' best response correspondences. To develop some intuition consider how player b finds her optimal strategy. Given the strategies of other players fixed, imagine b observes signals (x_a, x_b) such that x_a is very large. Although player b has no direct contact with c, she knows c has also observed this signal and can reason that it will drive c's posterior beliefs about θ upward. Then, given c's beliefs about A fixed, player b can argue that c will need to observe a very low realization of x_c (the only remaining signal observed by c) in order for $P(A \geq \theta \mid \mathbf{x_c}) \leq c$, which

[4]Because communication occurs for one round only, correlations emerge across players with at most 2 degrees of separation. If we added a fourth player d connected to c, then b and c would lie 3 links apart and would not be correlated in their posterior estimates of θ

we know induces c to attack. In sum, a high realization of x_a gives player b some confidence that c is very unlikely to attack, even as b and c are not directly connected. It should be clear that b can perform similar reasoning with respect to a's equilibrium behavior ex-post. Now consider what happens when b observes a very low realization of x_a instead. In this scenario b will argue that c will attack more often than before because, keeping others' strategies fixed, b can reason that this low realization of x_a will induce c's beliefs about θ downward and thus $P(A \geq \theta \mid \mathbf{x_c}) \leq c$ holds for larger realizations of x_c than before. Together, these arguments suggest that low values of x_a raise b's belief about c's (and a's) propensity to attack, while high values of x_a lower these beliefs.

This reasoning will affect b's equilibrium strategy because it affects her belief about the aggregate size of attack A. Since networks here are small, forming expectations about other's equilibrium action will in fact affect the belief about A and therefore will impact best responses. To see this notice that out indifference condition $Pr(A > \theta \mid \mathbf{x_i}) = c$, which defines the boundary of our set B_i, can be reformulated as

$$\int_{-\infty}^{\infty} Pr(A(\theta; B) > \theta) \, Pr(\theta \mid \mathbf{x}_i^\star) \, d\theta = c$$

where the dependence of A on players' equilibrium strategies $B = (B_1 \ldots, B_n)$ is made explicit. After rearranging and noticing that necessarily $A \in [0, 1]$ gives the following meaningful expression

$$\int_{0}^{1} Pr(A(\theta; B) > \theta) \, Pr(\theta \mid \mathbf{x}_\mathbf{i}^\star) \, d\theta = c - Pr(\theta < 0 \mid \mathbf{x}_i^\star) \tag{2.4}$$

Notice this equation essentially defines a best response correspondence $B_i(B_{-i})$ for each i that is non-linear following the intuition above: when determining the set of vectors $\mathbf{x}_\mathbf{i}^\star$ that satisfy the indifference condition, players must take into account that signal realizations will not only affect their inference of θ as determined by $Pr(\theta \mid \mathbf{x}_i^\star)$ but also how they will affect their beliefs about aggregate equilibrium behavior as defined by $Pr(A(\theta; B) > \theta)$. While all signals will contribute equally to statistical inference on the state of the world (the first effect), the previous discussion makes clear that signals will nonetheless have a varying impact on the belief about aggregate behavior in equilibrium (the second effect).

In later sections we show that under large networks this entire reasoning may be disregarded: the fact that c is more or less likely to attack (given an observation of a common signal x_a) will not affect b's belief about A when the population tends to infinity.[5] As a result, there is no strategic reason for players to take into account the correlation structures generated by the network; Indeed, a's signal is no more valuable than b's signal if the information it provides about c's equilibrium behavior does not move b's beliefs about A. More formally, since beliefs about A remain constant ex-post we will see that the above equation can be simplified by identifying regions $\left(-\infty, \hat{\theta}\right)$ where success occurs with probability 1 and regions $\left(\hat{\theta}, \infty\right)$ where success never occurs, so that the above equation can be rewritten as

$$\int_{-\infty}^{\hat{\theta}} Pr\left(\theta \mid \mathbf{x}_i^{\star}\right) = Pr\left(\theta < \hat{\theta} \mid \mathbf{x}_i^{\star}\right) = c \tag{2.5}$$

which allows us to solve the model analytically- more on that later. In short, when signals cannot inform players on aggregate behavior they merely retain their role in statistical inference (updating beliefs about θ), and they are all equally valuable in this sense.

Returning to b's equilibrium strategy, consider how her reasoning above affects her best response correspondence. We have seen that a high realization of x_a will not only force b's posterior beliefs about θ upward, but it will also allow b to conclude that both a and c will now attack less often in equilibrium, so that aggregate behavior A goes down. This will make b more reluctant to engage in attacking the status quo, unless his remaining signal x_b is extremely favorable for success. What is important here is that although we could make the same argument for a high realization of x_b, this signal will only allow b to form beliefs about a's equilibrium behavior, while x_a allows b to form beliefs about the equilibrium behavior of both a and c. In other words, a's signal moves b's beliefs about aggregate behavior much more than her own signal and so b will react more strongly to it. This unequal weighting that results from correlations in the network is what generates nonlinear strategies in finite networks as shown in figure 2.

We may now proceed with some general results for finite networks that, although not precise and analytical, nonetheless capture the basic strategic insight of the model when correlations are significant.

2.5 A Network Approximation

The reader can think that a network in this model generates two main effects: imbuing the system with correlation and allowing for the pooling of information.

[5] In truth, it is not sufficient that the population tend to infinity. We must also ensure that no player remains too central as population grows or else her signal would indeed move beliefs on aggregate behavior. This sparseness condition will be specified in the next section.

This paper focuses on the latter by assuming that we are in sufficiently large networks with only local correlation effects, such that they become strategically irrelevant. Then, the only effect of the network is that agents with larger degree have more precise signals. Of course not all network architectures exhibit a sufficiently local correlation- consider the complete network where correlation is maximal across all players. Clearly we are after a condition on sparseness that guarantees local correlation effects. So before we proceed let us formalize the admissible structures.

Formally, let N_d be the subset of players with degree d. Then we can extend the previous definition of A by partitioning the total population n into degrees in the following way,

$$A = \frac{1}{n}\sum_{j}^{N} 1_{\left\{\mathbf{x_j} \in B_j\right\}} = \frac{1}{n}\sum_{d=1}^{D}\sum_{j \in N_d} 1_{\left\{\mathbf{x_j} \in B_j\right\}}$$

Notice that each for each $j \in N_d$ the random variable $1_{\left\{\mathbf{x_j} \in B_j\right\}}$ is a weakly dependent Bernoulli with probability parameter $P(\mathbf{x_j} \in B_d)$. This symmetry across players of the same degree arises from the fact we deal with asymptotically large populations, where local correlations become strategically negligible. This "price-taking" effect implies that the identity of each incoming signal does not condition the behavior of A, so that in equilibrium players' position in the network becomes irrelevant and all that matters is the impact of degree on the precision of the posterior distribution. Moreover, since each incoming signal now has the same value, players will weight them all the same in a simple average, and the upper boundary of B_i is linear with a slope of -1. That is, to maintain indifference between attacking or not, a low realization of one signal can be offset by a higher realization of another by exactly the same amount, something that did not occur in the finite case above.(see figure) Formally what this means is that strategies can now be formulated in terms of the average of all incoming signals as such,

$$B_i = \left\{\mathbf{x}_i \mid \frac{1}{d_i}\left\langle \mathbf{x}_i, \mathbf{1_{d_i}} \right\rangle \leq x^\star \right\}$$

where we have that $x^\star$ satisfies our indifference condition $P\left(A \geq \theta \mid x^\star\right) = c$. [6] All this allows us to rewrite the above equation as,

$$\frac{1}{n}\sum_{d=1}^{D}\sum_{j \in N_d} 1_{\left\{\mathbf{x_j} \in B_d^\star\right\}} = \sum_{d=1}^{D}\frac{n_d}{n}\left(\frac{1}{n_d}\sum_{j \in N_d} 1_{\left\{\bar{x}_j \leq x_d^\star\right\}}\right)$$

[6] Notice the probability now is conditioned on the average realization, rather than on the entire vector. A cursory look at equation (2.1) should reveal that these are equivalent formulations

where $\bar{x}_j = \frac{1}{d_j}\left\langle \mathbf{x}_j, \mathbf{1_{d_j}} \right\rangle$. Now, if we can apply the Law of Large Numbers to the terms in brackets the share of players with degree d to attack the status quo converges from a binomial variable to a fixed share $P(\bar{x}_d \leq x_d^*)$. In that case we obtain that

$$A \underset{n\to\infty}{\longrightarrow} \sum_{d=1}^{D} P_d P(\bar{x}_d \leq x_d^*) \tag{2.6}$$

where P_d is the share of the population with degree d. Of course the elements summed in the bracketed terms are not independent random variables so the classical LLN does not apply. Instead we rely on LLN for weakly dependent random variables. The weak correlation structures that allow for LLN also condition the admissible network architecture by requiring a minimum level of sparseness. This is formalized in the following proposition that places conditions on the growth rate of the maximal degrees in order to assure that these limit properties hold in equilibrium.

Proposition 2.5.1. *Let* $d_1 = max\,(D)$ *and* $d_2 = max\,(D \setminus \{d_1\})$*. If* $d_1 \cdot d_2 \in o\,(n)$*, then equation* (2.6) *holds.*

Proof. See the Supplementary Appendix for a proof and a more general methodological result on convergence for a wide class of correlation structures. $\square$

By placing a bound on the growth rate of maximal degrees, proposition 2.5.1 essentially forces us to concentrate on networks that are sufficiently sparse in that no one individual contains too many links relative the size of the population. Intuitively, if we are trying to coordinate with a large set of individuals then our local correlations with a small subset of the population becomes strategically insignificant and we can disregard them when developing our optimal strategies. As such, all incoming signals are equally valuable and the upper frontier of the set B_i is linear. This is the key step in the model, and what allows us to proceed with an analytical solution that focuses on the network effect on precision and ignores correlations. As a counter example consider a star network. Clearly the maximal degree in this network grows linearly with n so that the proposition fails. In this case it is fairly straightforward that correlations remain a crucial determinant of equilibrium for all population sizes: all spokes know that everyone else has observed the hub's signal, so its' realization will move everyone's beliefs about A much more than their own realization. In other words, the frontier of B_i is not linear because correlation effects imply that not all signals are equally valuable.

If we are willing to focus on the networks prescribed by the proposition above, we can proceed and think of the network as summarized by the degree distribution. We are left with an infinite population that is split into D partitions with proportions P_d for $d = 1, \ldots, D$. Each partition has measure 1 and each agent of partition $d \in \{1, 2, \ldots, D\}$ will have degree d. In this sense, the following model essentially extends the standard global game of regime change to a population with heterogeneous variances.

2.6 Equilibrium With Diffuse Prior

Agents have a diffuse prior over the state of the world (i.e. θ is distributed uniformly over the real line). Next, suppose there is a degree-specific threshold strategy $x_d^\star \in \Re$ such that each agent with degree d attacks if and only if $x \leq x_d^\star$. The measure of agents who attack is given by

$$A(\theta) = \sum_{d=1}^{D} P_d \cdot \Phi\left(\frac{\sqrt{d}}{\sigma}\left(x_d^\star - \theta\right)\right)$$

where Φ is the CDF of the standard normal. Because $A(\theta)$ is decreasing in θ we can be sure there exists a value $\hat{\theta}$ that is a fixed point of $A(\theta)$. Formally, $\hat{\theta}$ solves,

$$A(\hat{\theta}) = \sum_{d=1}^{D} P_d \cdot \Phi\left(\frac{\sqrt{d}}{\sigma}\left(x_d^\star - \hat{\theta}\right)\right) = \hat{\theta} \tag{2.7}$$

Finally, notice that there is regime change whenever $\theta \leq \hat{\theta}$. Standard Bayesian updating implies that the posterior expectation for an agent i of degree d that receives a signal realization x_i is $\theta \mid x_i \sim N(x_i, \frac{\sigma^2}{d})$. Therefore, to this particular agent, the probability of regime change is given by, $pr(\theta \leq \hat{\theta} \mid x_i) = \Phi\left(\frac{\sqrt{d}}{\sigma}\left(\hat{\theta} - x_i\right)\right)$. The agent will find it optimal to attach wheneverthe posterior probability of regime change is greater than the marginal cost from attacking c, or whenever, $x_i \leq x_d^\star$ where $x_d^\star$ solves,

$$\Phi\left(\frac{\sqrt{d}}{\sigma}\left(\hat{\theta} - x_d^\star\right)\right) = c \qquad for \;\; d = 1, \ldots, D \tag{2.8}$$

A Bayesian Equilibrium is a $D + 1 - tuple$, $\left(\hat{\theta}; x_1^\star, x_2^\star, \ldots, x_D^\star\right)$ that solves equations (2.7) and (2.8). Notice players are strategic, and respond to everyone else's strategy through the parameter $\hat{\theta}$. We are essentially describing a coordination game of incomplete information with heterogeneous precisions.

Strategic response to others' strategies does not mean, however, that we face pathological, corner equilibria where everyone chooses never (always) to attack. Consider the response of individual i to a population where every other player chooses never to attack (i.e. everyone chooses $x_j^\star = -\infty$). In that case $A(\theta) = 0$

for all values of θ so that $\hat{\theta} = 0$. But consider introducing this to equation (2.8); Because $c > 0$ player i will choose an equilibrium threshold $x_i^\star$ bounded away from $-\infty$, so that never attacking is not an equilibrium strategy. In fact this will constitute an equilibrium if and only if $c = 1$. Intuitively, even though everyone else is choosing not to attack, there is still a positive probability that $\theta \leq 0$ (recall that player i's posterior belief about theta is $\theta \mid x_i \sim N(x_i, \sigma^2)$) so player i will choose to attack for some positive probability as a result.[7]. This is true as well when considering the equilibrium where everyone always attacks (i.e. this will only be an equilibrium for $c = 0$). Together this implies we have a unique equilibrium. Our first result establishes the existence of this unique equilibrium by explicitly solving for $\hat{\theta}(c, \mathbf{p})$ in equation (2.7), and then establishes its lack of response to the degree distribution.

Proposition 2.6.1. *There exists a unique equilibrium to this static one-shot game in monotone strategies. The probability of regime change, $Pr(\theta < \hat{\theta})$, does not change with degree distribution* **p** *and decreases with c.*

Proof. First rewrite equation (2.8) to get the best reply function for a player i of degree$d$,

$$x_d^\star = \hat{\theta} - \frac{\sigma}{\sqrt{d}}\Phi^{-1}(c)$$

Then substituting this into equation (2.7) gives

$$A(\hat{\theta}) = \sum_{d=1}^{D} P_d \cdot \Phi\left(\frac{\sqrt{d}}{\sigma}\left(\hat{\theta} - \frac{\sigma}{\sqrt{d}}\Phi^{-1}(c) - \hat{\theta}\right)\right) = \sum_{d=1}^{D} P_d \cdot \Phi\left(\frac{\sqrt{d}}{\sigma}\left(-\frac{\sigma}{\sqrt{d}}\Phi^{-1}(c)\right)\right)$$

$$= \sum_{d=1}^{D} P_d \cdot \Phi\left(-\Phi^{-1}(c)\right) = \sum_{d=1}^{D} P_d\left(1 - \Phi\left(\Phi^{-1}(c)\right)\right) = \sum_{d=1}^{D} P_d(1-c) = 1 - c = \hat{\theta}$$

Then $\hat{\theta} = 1 - c$ is independent of the degree distribution $(p_1, p_2, \ldots, p_D)$ and clearly decreases with c. This implies that the probability of regime change $Pr\left(\theta \leq \hat{\theta}\right)$ is also invariant to the degree distribution. Finally, we can calculate threshold strategies $(x_1^\star, x_2^\star, \ldots, x_D^\star)$ that define the equilibrium by substituting in for $\hat{\theta}$ to obtain $x_d^\star = 1 - c - \frac{\sigma}{\sqrt{d}}\Phi^{-1}(c)$. □

[7] With a bounded distribution of private noise this corner equilibria could be retrieved. To see this imagine instead that $\epsilon_i \sim U(\underline{\epsilon}, .\overline{\epsilon})$. Players' posterior beliefs about θ would be $\theta \mid x_i \sim U(x_i - \underline{\epsilon}, x_i + \overline{\epsilon})$. If $\underline{\epsilon}$ and $\overline{\epsilon}$ are sufficiently close, we cannot rule out a situation where all players believe the probability that $\theta = 0$ is 0 (alternatively 1), so that each player reacts to $\hat{\theta} = 0$ ($\hat{\theta} = 1$) by choosing an equilibrium strategy consistent with never (always) attacking the regime, which means choosing $x_i^\star = -\infty$ ($x_i^\star = +\infty$).

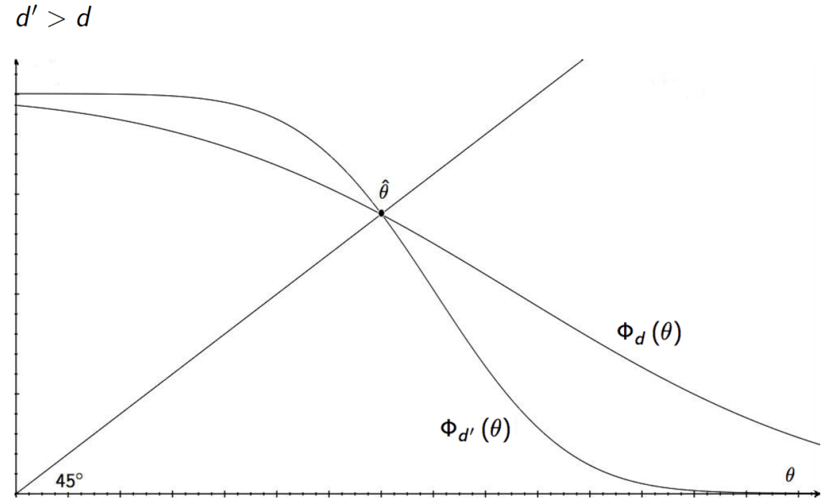

Figure 2.1: Size of Attack as a function of θ

This result at first glance is not at all intuitive. In fact, we would expect the degree distribution to have an impact on the probability of regime change. It turns out that the range of θ where attacks are successful does not change as we alter the average connectivity of the society. Why? With a diffuse prior, players pay no attention to prior information; only private signals feed the inference of posterior beliefs. As a result, the strategic uncertainty of each agent is maximal with respect to the behavior of others. Formally, each player's higher-order beliefs on θ characterized by equation (2.8) have the same shape as the commonly observed distributions of private signals that sum in equation (2.7). Players have no way of improving on these beliefs. So even though different degrees select different threshold strategies, it turns out that precisely at the value of θ where the size of attack is the smallest successful attack, the propensity to take action is identical across the entire population. This is important because it implies that shuffling the distribution of degrees will not modify where the smallest successful attack is defined, so it will not modify the range of θ where successful attacks begin. As a result, the probability of observing a success will also remain fixed.

It is important to stress that threshold strategies are not identical. In fact threshold strategies depend crucially on d. However, the behavior of the population in equilibrium cannot be glimpsed directly from these values. Instead, the share of individuals of degree d that decide to attack the status quo in equilibrium is defined by $Pr\left(x_d \leq x_d^\star \mid \theta\right)$, and this need not obey the ordering of threshold strategies $x_d^\star$. To see this consider the case of $c > \frac{1}{2}$. Players with a larger degree have a tighter posterior of θ and will choose a strictly higher threshold defined by $x_d^\star = \hat{\theta} - \frac{\sigma}{\sqrt{d}}\Phi^{-1}\left(c\right)$. However, notice that for the same threshold $x^\star$, these high degree individuals will observe $x \leq x^{\star}$ less often than low degree individuals. It turns out that at $\hat{\theta}$ high degree players will choose a threshold that is larger by the amount that exactly compensates the lower probability of observing a signal to the left of said threshold. As a result, even though it is true that $x_1^{\star} < x_2^{\star} < \cdots < x_D^{\star}$, in equilibrium we have that $Pr(x_1 \leq x_1^{\star} \mid \hat{\theta}) = Pr(x_2 \leq x_2^{\star} \mid \hat{\theta}) = \cdots = Pr(x_D \leq x_D^{\star} \mid \hat{\theta})$. This striking result comes from the fact that the shape of the posterior belief about θ (which chooses the threshold $x_d^{\star}$) and the shape of the distribution of signals (which determines $Pr\left(x_d \leq x_d^{\star}\right)$) are exactly the same. Once we introduce a prior with finite variance players make us of it as a public coordinating device and higher order beliefs about θ will depart from higher order beliefs about $x_d \sim N\left(\theta, \frac{\sigma^2}{d}\right)$. As we will see, however, too little strategic uncertainty in the form of too low a prior variance, leads to multiplicity by increasing strategic complementarities of the model.

Proposition 2.6.1 should not convey the idea that the propensity to attack is independent of the degree distribution. This is only true at the point $\hat{\theta} = A\left(\hat{\theta}\right)$, which as mentioned above implies that the likelihood of observing a successful attack remains fixed for all $\mathbf{p}$. However, the equilibrium is defined for all values of $\theta \neq \hat{\theta}$. In these cases, the size of the attack responds directly to the weights chosen for each degree. In other words, conditional on the attack being too small to be successful, its size will vary with the degree distribution. The same applies for attacks large enough to be successful. In order to gain some intuition, Figure 2.1 plots the size of attack $A\left(\theta\right)$ for two different populations: one with $p_d = 1$ and the other with $p_{d'} = 1$ (where $d' > d$). You can think that a population with positive weight on both these degrees will have an $A\left(\theta\right)$ line somewhere in between. What is important to note first is that proposition 2.6.1 can be thought of as stating that these two lines (and in fact all other lines for all degrees $d = 1, 2, \ldots, D$) intersect at the $45°$ line. It should be clear that any convex combination of these two functions (i.e. for any degree distribution) will always cross the $45°$ at the same point of intersection $\hat{\theta} = 1 - c$. For all values different form $\hat{\theta}$, however, the two curves take on quite different values and it is here where the degree distribution will determine the size of attacks. This is the content of our next result.

Proposition 2.6.2. *Define $A_{\mathbf{p}}\left(\theta\right)$ as the equilibrium size of attack under degree distribution $\mathbf{p}$ and state of the world θ. Let $\mathbf{p}'$ FOSD $\mathbf{p}$ then:*

- *for $\theta > \hat{\theta}$ (failure) $A_{\mathbf{p}}\left(\theta\right) > A_{\mathbf{p}'}\left(\theta\right)$ and $A_{\mathbf{p}}\left(\theta\right) - A_{\mathbf{p}'}\left(\theta\right)$ increases with θ*
- *for θ¡$\hat{\theta}$ (success) $A_{\mathbf{p}}\left(\theta\right) < A_{\mathbf{p}'}\left(\theta\right)$ and $A_{\mathbf{p}'}\left(\theta\right) - A_{\mathbf{p}}\left(\theta\right)$ increases with θ*

Moreover, this effect is largest whenever $c = \frac{1}{2}$ and decreases monotonically as $c \rightarrow \{0, 1\}$

Proof. Consider the equilibrium size of attack by plugging in equilibrium threshold strategies found in proposition 1 into the definition of $A(\theta)$

$$A(\theta) = \sum_{d=1}^{D} P_d \cdot \Phi\left(\frac{\sqrt{d}}{\sigma}\left(1 - c - \frac{\sigma}{\sqrt{d}}\Phi^{-1}(c) - \theta\right)\right) =$$

$$\sum_{d=1}^{D} P_d \cdot \Phi\left(\frac{\sqrt{d}}{\sigma}(1 - c - \theta) - \Phi^{-1}(c)\right)$$

Notice that whenever $\theta > \hat{\theta}$ ($\theta < \hat{\theta}$) then the factor multiplying $\frac{\sqrt{d}}{\sigma}$ above is positive (negative) so that the argument of Φ is larger (smaller) for a larger degree. Then, by shifting weight to larger degrees (FOSD) we increase the weights on those terms in the summation that are larger (smaller) leading to a total value of $A(\theta)$ that is larger (smaller). For the second part of the proof notice that for $c = \frac{1}{2}$ we have $\Phi^{-1}(c) = 0$ and that as c tends to the extremes $\Phi^{-1}(c)$ tends to $\pm\infty$. □

Proposition 2.6.2 essentially states that in a population with equal weights, unsuccessful attacks are composed by a majority of less connected individuals, while successful attacks are composed by a majority of more connected individuals. This confirms our intuition that more connected individuals, because they are more informed, miscoordinate less often. After all, they obtain a more precise estimate of the true parameter, so it only makes sense that once θ is too large to guarantee success they retreat from attacking in larger shares. Graphically, you can see that for $\theta > \hat{\theta}$ the slope of $A(\theta)$ is steeper for the more connected individual.

What seems harder to reconcile, however, is that the difference in performance across degrees diminishes monotonically as the costs of revolt become more extreme. The intuition here is that the informational advantage of more informed individuals is greatest when the costs of attack are the least extreme. In other words, for costs neither too high nor too low more informed individuals shirk from attacking much more quickly as θ rises, creating a large advantage vis-a-vis the less informed. But for extreme costs these individuals respond less to the value of θ and choose to attack more or less the same for all state of the world (after all, either the costs are so low that attacking is almost always a better option, or too high to attack regardless of the state of the world).

2.7 Equilibrium With Non-Diffuse Prior

Next we turn to the case where player's hold a finite-variance prior about the state of the world, and must therefore incorporate it into their posterior beliefs.

It turns out that in this scenario the previous results change dramatically. In particular, with the presence of a public signal the degree distribution completely determines the probability of success. Most interestingly, this comparative static is not monotone- so that more average connectivity means a greater probability of success- and instead depends on the cost of attack, c. When the costs are high more connectivity translates to a larger share of success, but the opposite is true for sufficiently low costs.

To begin the analysis, notice that if the prior is not diffuse, and instead follows a normal distribution,

$$\theta \sim \mathcal{N}(\theta_0, \sigma_0^2)$$

then the posterior distribution of the state of the world, θ, looks like,

$$\theta | x_i \sim \mathcal{N}(\frac{d_i \sigma_0^2 x_i + \sigma^2 \theta_0}{d_i \sigma_0^2 + \sigma^2}, \frac{\sigma_0^2 \sigma^2}{d_i \sigma_0^2 + \sigma^2})$$

The computation of $A(\theta)$ remains unaffected to the . However, the last part of the analysis changes and leads to

$$x_d^* = (1 + R_d)\hat{\theta} - \frac{\sigma}{\sqrt{d}} \sqrt{1 + R_d} \Phi^{-1}(c) - R_d \theta_0. \tag{2.9}$$

where

$$R_d = \frac{\sigma^2}{d \sigma_0^2}$$

Plugging this in the equation $A(\hat{\theta}) = \hat{\theta}$ leads to

$$\sum_d P_d \Phi \left(\frac{\sigma}{\sigma_0^2 \sqrt{d}} \left(\hat{\theta} - \theta_0 \right) - \sqrt{1 + R_d} \Phi^{-1}(c) \right) = \hat{\theta} \tag{2.10}$$

An Equilibrium is a $D + 1 - tuple$ $\left(\hat{\theta}, (x_d^\star)_{d=1}^D \right)$ that solves equations (2.9) and (2.10).

Equilibrium Analysis

One first thing to notice is that, in equation (2.9), as $\sigma_o^2 \to \infty$, the equilibrium strategy for all degrees $x_d^\star$ tends to $\hat{\theta} - \frac{\sigma}{\sqrt{j}} \Phi^{-1}(c)$ (the solution to section 2.6) and that similarly $x_d^\star$ tends to $\hat{\theta}$ for sufficiently large degrees. This seems to suggest that as the public signal becomes more noisy and agents switch to the private signal, then the same effect as in diffuse prior starts kicking in. In other words the marginal effect of degree on the weights placed on each signal disappears. In a sense, we can think of the degree distribution as determining how strongly the population will value their private signal against the public signal.

Before we proceed to establish the effect of the degree distribution on the size and outcome of coordinated attacks, we shall establish the existence of a unique equilibrium. Unlike the previous section, the presence of a public signal means we can only guarantee uniqueness above a minimum public variance. The reason rests on arguments from Morris-Shin (2002). As the public signal variance diminishes, all players will shift their posterior beliefs towards said signal, in effect increasing the level of correlated beliefs and, in a sense, continuously increasing the level common knowledge. At a certain point the level of strategic uncertainty is sufficiently low to generate multiplicity. The following result establishes the lower bound on public variance to guarantee uniqueness.

Proposition 2.7.1. *A unique equilibrium exists if* $\sigma_0^2 > \frac{\sigma}{\sqrt{2\pi}} \sum_d \frac{p_d}{\sqrt{d}}$

Proof. Rewrite equation (2.10) as $F\left(\hat{\theta}; \theta_0, \sigma^2, \sigma_0^2\right) = 0$ where

$$F\left(\hat{\theta}; \theta_0, \sigma^2, \sigma_0^2\right) = \sum_d P_d \Phi\left(\frac{\sigma}{\sigma_0^2 \sqrt{d}}\left(\hat{\theta} - \theta_0\right) - \sqrt{1 + R_d}\,\Phi^{-1}(c)\right) - \hat{\theta}$$

Note that $F\left(\hat{\theta}; \cdot\right)$ is continuous and differentiable in $\hat{\theta} \in (0, 1)$ and that

$$F(0; \cdot) = \sum_d p_d \Phi\left(\frac{-\theta_0 \sigma}{\sigma_0^2 \sqrt{d}} - \sqrt{1 + R_d}\,\Phi^{-1}(c)\right) > 0$$

and

$$F(1; \cdot) = \sum_d p_d \Phi\left(\frac{\sigma}{\sigma_0^2 \sqrt{d}}(1 - \theta_0) - \sqrt{1 + R_d}\,\Phi^{-1}(c)\right) - 1 < 0$$

Then I need to show F is monotonically decreasing in $\hat{\theta}$. Notice that

$$\frac{\partial F}{\partial \hat{\theta}} = \sum_d p_d \phi\left(\frac{\sigma}{\sigma_0^2 \sqrt{d}}\left(\hat{\theta} - \theta_0\right) - \sqrt{1 + R_d}\,\Phi^{-1}(c)\right) \cdot \left(\frac{\sigma}{\sigma_0^2 \sqrt{d}}\right) - 1$$

and given that $\underset{\hat{\theta}}{max} \phi(\cdot) = \frac{1}{\sqrt{2\pi}}$ the condition $\frac{1}{\sqrt{2\pi}} \frac{\sigma}{\sigma_0^2} \sum \frac{p_d}{\sqrt{d}} < 1 \implies \sigma_0^2 > \frac{\sigma}{\sqrt{2\pi}} \sum \frac{p_d}{\sqrt{d}}$ is both necessary and sufficient for F to be monotonic in $\hat{\theta}$. □

This result suggests that introducing heterogeneity in the model retrieves uniqueness for a greater range of parameter values. This is to be expected since the presence of a degree distribution essentially introduces a mass of individuals with a lower private variance with respect to the homogeneous population scenario. The presence of more informed individuals strengthens the role of private information, making coordination more difficult. In other words, the presence of varying degrees in this model essentially translates into a convex combination of weights placed on the public signal. As more weight is placed on high degree individuals (who in turn pay less attention to the public signal) then we retain uniqueness for smaller public-signal variances than the previous models allowed.

There is no explicit analytical solution to equation (2.10) of the form $\hat{\theta}(c, \mathbf{p})$ that would describe the impact of $\mathbf{p}$ on the probability of regime change. However, we can use the implicit function theorem to say something about the comparative statics across various types of degree distributions. I will focus on the set of parameter values that guarantees uniqueness (i.e. $\sigma_0^2 > \sigma$). The first result establishes a surprising non-monotone comparative statics for a population composed of only 2 degrees and completely general class of degree distributions. The following result focuses on power-law distributions in order to extend the result to an arbitrary number D of degrees.

Proposition 2.7.2. *Let $\sigma_0^2 > \sigma$ and $D = 2$. Define two general degree distributions $\mathbf{p} = (p_1, (1 - p_1))$ and $\mathbf{p}' = (p_1', (1 - p_1'))$ such that $\mathbf{p}$ FOSD $\mathbf{p}'$ (i.e. such that $p_1 < p_1'$). Then, there exists a threshold $\hat{c} \in (0, 1)$ such that*

- *for all $c \in (0, \hat{c})$ the probability of regime change is lower under $\mathbf{p}$ than under $\mathbf{p}'$*
- *for $c \in (\hat{c}, 1)$ the probability of regime change is larger under $\mathbf{p}$ than under $\mathbf{p}'$.*

Proof. Rewrite equation (0.6) for $D = 2$

$$p_1\Phi\left(\frac{\sigma}{\sigma_0^2}\left(\hat{\theta} - \theta_0\right) - \sqrt{1 + R_1}\Phi^{-1}(c)\right) +$$

$$(1 - p_1)\Phi\left(\frac{\sigma}{\sigma_0^2\sqrt{2}}\left(\hat{\theta} - \theta_0\right) - \sqrt{1 + R_2}\Phi^{-1}(c)\right) = \hat{\theta}$$

define an implicit function $F\left(\hat{\theta}(p_1), p_1\right) = 0$. Applying the implicit function theorem as before we obtain

$$\frac{\partial\hat{\theta}}{\partial p_1} = -\frac{\frac{\partial F}{\partial p_1}}{\frac{\partial F}{\partial \hat{\theta}}} =$$

$$-\frac{\Phi\left(\frac{\sigma}{\sigma_0^2}\left(\hat{\theta} - \theta_0\right) - \sqrt{1 + R_1}\Phi^{-1}(c)\right) - \Phi\left(\frac{\sigma}{\sigma_0^2\sqrt{2}}\left(\hat{\theta} - \theta_0\right) - \sqrt{1 + R_2}\Phi^{-1}(c)\right)}{\sum_d p_d\phi\left(\frac{\sigma}{\sigma_0^2\sqrt{d}}\left(\hat{\theta} - \theta_0\right) - \sqrt{1 + R_d}\Phi^{-1}(c)\right)\left(\frac{\sigma}{\sigma_0^2\sqrt{d}}\right) - 1}$$

Whenever $\sigma_0^2 > \sigma$ the denominator is negative (see previous proof). As a result, the sign of the comparative static is determined entirely by the sign of the numerator. Notice that

$$\frac{\partial \hat{\theta}}{\partial p_1} = 0 \implies \frac{\sigma}{\sigma_0^2}\left(\hat{\theta} - \theta_0\right) - \sqrt{1+R_1}\Phi^{-1}(c) = \frac{\sigma}{\sigma_0^2\sqrt{2}}\left(\hat{\theta} - \theta_0\right) - \sqrt{1+R_2}\Phi^{-1}(c)$$

rearranging we get:

$$\Phi\left(\frac{\left(\hat{\theta} - \theta_0\right)\left(\frac{\sigma}{\sigma_0^2} - \frac{\sigma}{\sigma_0^2\sqrt{2}}\right)}{\sqrt{1+R_1} - \sqrt{1+R_2}}\right) = c$$

where $\hat{\theta}$ is endogenously determined in equilibrium and decreases with c (check equation (2.10)). Since $\Phi(\cdot)$ is a continuous, monotone function defined over the interval $[0, 1]$ and moves positively with $\hat{\theta}$, we can be sure there exists a unique $\hat{c}$ that solves,

$$\Phi\left(\frac{\left(\hat{\theta}(\hat{c}) - \theta_0\right)\left(\frac{\sigma}{\sigma_0^2} - \frac{\sigma}{\sigma_0^2\sqrt{2}}\right)}{\sqrt{1+R_1} - \sqrt{1+R_2}}\right) = \hat{c}$$

Finally, notice that for all $c < \hat{c}$ the left hand side of this equation is larger than the right hand side so that $\frac{\partial \hat{\theta}}{\partial p_1} > 0$ and for all $c > \hat{c}$ the left hand side is smaller than the right so that $\frac{\partial \hat{\theta}}{\partial p_1} < 0$, thus proving the result. $\square$

The non-monotonicity implied by proposition 2.7.2 is surprising. Indeed, increasing the average connectivity of the population *does not* increase the likelihood of success unambiguously. Instead, a low cost of failure increases the marginal propensity to attack of low connected individuals by far more than the corresponding increase experienced by highly connected individuals. As a result, low connected players choose to attack more often for a greater range of θ values, including the precise value $\hat{\theta}$ that determines the likelihood of success. It is still true (as in proposition 2.6.2) that all failed attacks will register greater participation by less connected (i.e. less informed) individuals. But if costs are low, some successful attacks also will contain greater shares of low connected players. Intuitively, the low risk involved in attacking the status quo gives less informed individuals an advantage by being more reckless (i.e counting more on rare tail events). Anticipating this behavior, the minorities of more informed individuals will respond by adhering more to this propensity of attack than their signals would normally prescribe (Equilibrium threshold strategies respond to $\hat{\theta}$ and this

Let $d' > d$

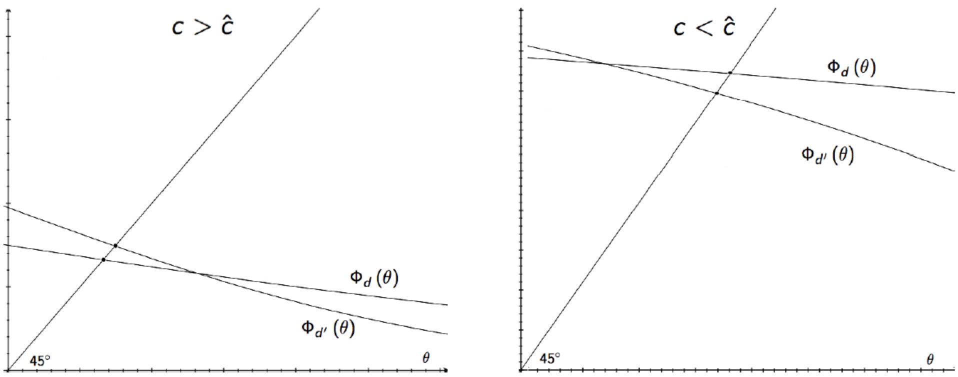

Figure 2.2: Size of Attack as a function of θ

captures the strategic response of players to aggregate behavior). Of course as more informed individuals cease to be a minority, the opposite will occur and less informed individuals will strategically respond to their expected aggregate behavior. As such, they will choose to be more cautious about attacking the regime than their dispersed, low-quality information would initially prescribe. Figure 2.2 plots the size of attacks as a function of θ for both low high costs. You can see in the right-hand panel that for costs sufficiently low, the mass of attackers of lower degree is in fact superior to the mass of attackers of high degree for all values of θ corresponding to failed attacks (as was the case in the previous section) but also for some values of θ corresponding to successful attacks. In other words, less connected players surpass more connected players in their shares of attack at a lower value of θ than was the case with no prior information. Then, if the shares of low connected individuals increase the value of $\hat{\theta}$ will increase and so will the probability of regime change.

But the question remains. Why are players with more dispersed posterior beliefs more willing to take action when the costs are low? A look at figure 2.3 reveals that this result comes from the symmetry of distributions. Less connected individuals attain a more dispersed posterior belief about the state of the world, θ, and so must center their distributions further away from the critical cutoff $\hat{\theta}$ in order to attain a probability of success equal to c. When $c > \frac{1}{2}$ (Right-hand panel) the cutoff will certainly lie to the right of every player's expected belief, and less connected players will therefore choose a lower equilibrium threshold. With a lower threshold there is no hope to attack in greater shares. But when $c < \frac{1}{2}$ (Left-

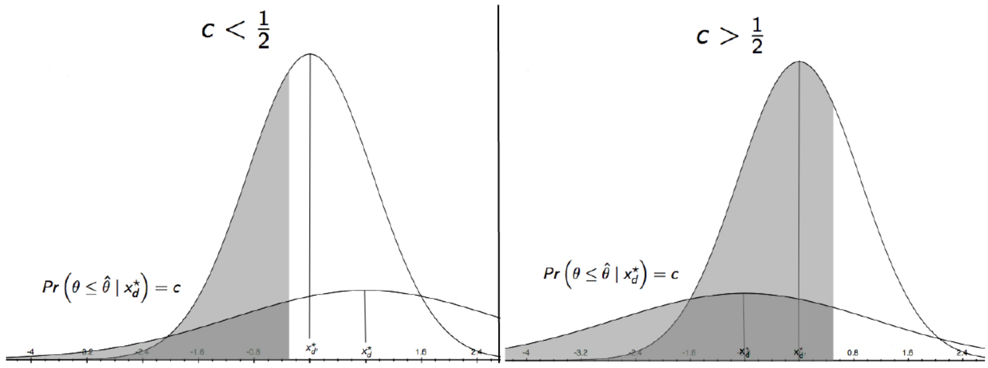

Figure 2.3: Less Informed Players choose larger $x_i^\star$ for c sufficiently low

hand panel) the cutoff will lie to the left of the distribution's center, whatever it turns out to be. In that case more dispersed distributions will center further to the right than less dispersed ones and thus low degree individuals will choose an equilibrium threshold that is larger. Moreover this effect increases with c and we can attain arbitrarily large distances between the threshold strategies of large and low degree players. If costs are sufficiently low, so that this distance is sufficiently large, then less connected individuals can attain such a larger threshold h that the probability of attack is larger at $\hat{\theta}$ (see figure 2.2).

Intuitively, the high expected gains and low expected losses that come with low values of c mean that players need not be very sure of the probability of success (remember at equilibrium players choose threshold such that probability of success equals c). In that case being less informed has a strategic advantage in so far as more weight is given to rare events, and these rare events are now enough to trigger action. Then as a low connected individual you will attack unsuccessfully more often than others, but you will also sometimes attack successfully more often than others. As a result, the status quo will need to be stronger to survive a population with your equilibrium behavior.

The previous result showed that the model exhibits non-monotone comparative statics with respect to the degree distribution for a general class of distributions and $D = 2$. Of course we would like to say something for a wider support of the degree distribution. The difficulty emerges in expressing a FOSD shift in one single parameter. We need this since our proofs rest on totally differentiating an implicit function in θ. Indeed for a great many number of distributions this cannot be done. Following a vast number of studies[8] that have documented the prevalence of scale-free characteristics across most types of large networks, we

[8]See, for instance, Barabasi & Reka (1999), Barabasi (2004),

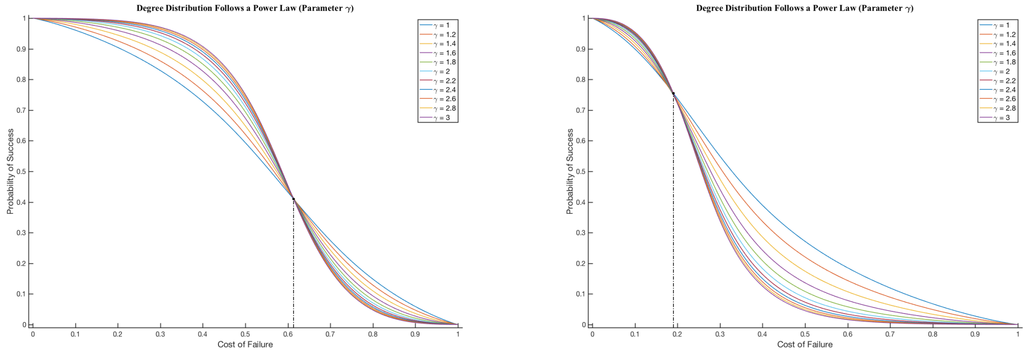

Figure 2.4: Probability of Success against Cost of Failure for model simulations under a powerlaw distributions and with parameters $D = 200$, $\sigma_0^2 = 4$, $\sigma^2 = 16$. Panel A: $\theta_0 = 0$. Panel B: $\theta_0 = 2$.

focus on degree distributions where the probability that a vertex is connected to k other vertices decays as a power law following $P(k) \propto k^{-\gamma}$. Then, we can consider shifts in γ as FOSD movements in the degree distributions. To see that this is equivalent, notice that lower values of γ imply that the probably of degrees decays more slowly, so that at least for large values of d, a power-law distribution with $\gamma' > \gamma$ is First Order Stochastically *dominated* with respect to a power-law distribution with parameter γ.

We solve the model numerically by first simulating power law degree distributions for $D = 200$ and a wide array of different $\gamma' s$. We then find the $\hat{\theta}$ that solves the equilibrium condition shown in equation (2.10). Recall that the value of $\hat{\theta}$ specifies the likelihood of successful regime change in equilibrium. From this exercise we therefore obtain the success probability as a function of c, for each different power law distribution (parametrized by γ). We plot the results in Figures 2.4 and 2.5 for different values of θ_0, σ_0, and σ. The two panels of Figure 2.4 capture the effect of changing θ_0, the mean of the prior beliefs about the strength of the status quo. The left panel corresponds to $\theta_0 = 0$ and the right panel corresponds to $\theta_0 = 2$. As found in Proposition 2.7.2 for the case of $D = 2$, raising the prior beliefs about θ lowers the value $\hat{c}$ at which the comparative statics are reversed. In any case, notice that there exists one unique value $\hat{c}\,(\sigma, \sigma_0, \theta_0) \in (0, 1)$ such that, to the left of $\hat{c}$ increasing γ increases the probability of regime change, and to the right of $\hat{c}$ the probability of success decreases with γ. The two panels of Figure 2.5 capture the effect of an increase in $\frac{\theta_0^2}{\theta}$. Again, it is clear that there exists one unique value $\hat{c}\,(\sigma, \sigma_0, \theta_0) \in (0, 1)$ where the ordering of the curves is reversed. However, in this case, raising $\frac{\theta_0^2}{\theta}$ in fact increases the value of $\hat{c}$. This was also true in Prop 2.7.2 for the case of $D = 2$. In general, while an analytical solution for

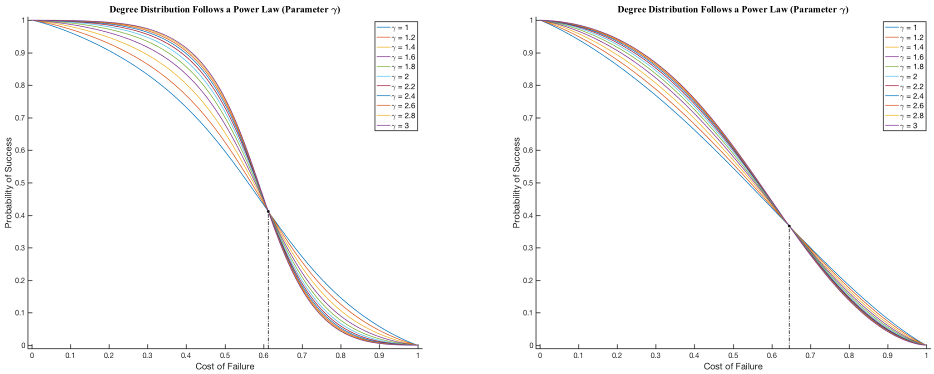

Figure 2.5: Probability of Success against Cost of Failure for model simulations under a powerlaw distributions and with parameters $D = 200$, $\theta_0 = 0$. Panel A: $\frac{\theta_0^2}{\theta} = 1$. Panel B: $\frac{\theta_0^2}{\theta} = 2$.

power law distributions ($P(k) \propto k^-\gamma$) is hard to come by, the numerical results presented in Figures 2.4 and 2.5 have been carried out for an extensive range of parameter values. [9] We can therefore state the following useful observation.

Observation: *Let $\sigma_0^2 > \sigma$. For the class of degree distributions following a Power Law (i.e. $P(d) \propto d^{-\gamma}$ for some $\gamma > 0$), there exists a unique threshold $\hat{c}(\sigma, \sigma_0, \theta_0) \in (0,1)$ such that for all $0 < c < \hat{c}$ the probability of regime change INCREASES with γ, and for all $1 > c > \hat{c}$ the probability of regime change DECREASES with γ. Moreover, $\hat{c}(\sigma, \sigma_0, \theta_0)$ decreases with θ_0 and increases with $\frac{\theta_0^2}{\theta}$.*

The results above suggests that, as with the case where $D = 2$, the success probability does not respond monotonically to a FOSD shift in the degree distribution, and instead identifies a threshold cost where the direction of comparative statics is reversed. The intuition corresponds to the arguments presented above and can be seen clearly in Figures 2.4 and 2.5.

2.8 Conclusion

Models of large-scale coordination with incomplete information have, by and large, neglected the role of communication, and in particular the role of connectivity in pooling information. In this paper, we propose a model of large-scale coordination within a network of truthful communication. We assume agents observe the private information held by their neighbors within a given social network.

[9] Only some examples are shown here. Code available upon request.

This generates a situation of locally public signals that correlate posterior beliefs according to the structure of social interactions. Moreover, a connectivity effect guarantees that the strength of posterior beliefs In this environment, we describe the equilibrium for two-action, two-outcome global games with large networks.

The main technical contribution of this paper provides an upper bound on network density as a function of size, such that, for any communication protocol, the correlation of posterior beliefs is sufficiently mild relative to the connectivity effect. This implies that, for large networks, the connectivity effect dominates and, in the limit, the problem reduces to independent posterior beliefs with strength proportional to connectivity. This result essentially transforms the problem into a set of mixing sequences and applies a weak form of the Law of Large Numbers under finite range dependence. This allows us to solve for an equilibrium, simply as a function of a network's degree distribution.

After characterizing the equilibrium, we perform comparative statics on the network by shifting the degree distribution. We show that these considerations are not innocent, and that the strategic impact of connectivity on equilibrium outcomes is far from obvious. Indeed, largely connected individuals, while they care little (a priori) for publicly observed information, must strategically respond to the behavior of less connected individuals, and therefore to the public signal indirectly. We show that, when prior beliefs are diffuse (i.e. publicly-held information is completely uninformative of the state of the world), then the probability of successful coordination does not depend at all on the degree distribution. However, the size of successful and unsuccessful attacks does vary with the degree distribution – more informed populations will correspond with smaller, unsuccessful attacks and larger successful attacks. On the other hand, when public information provides information, shifting the degree distribution affects the likelihood of coordination, and we show that the effect is non-monotone and depends crucially on the cost of mis-coordination. In particular, if the cost of failure is sufficiently small, then the probability of success increases as networks are, on average, less connected. The opposite is true for large failure costs. Intuitively, if payoffs don't fall by much when coordination fails then the probability of success is maximized by having less informed individuals that are less selective about when to attack. Indeed in this scenario failure is also more ubiquitous, but so are successful attacks, increasing the total probability.

We have considered here equilibrium and comparative statics results in the world where information's noise does not vanish. Part of the literature of global games has focused attention in what happens when noise tends to zero. In that sense, the work of Sakovics and Steiner (2013) is complementary to our approach: in their paper, different infinte groups can receive information with different probability distributions and they provide a closed form expression of the common threshold for all groups in the limiting case where noise vanishes.

A more thorough investigation of these type of communication protocols on coordination games is warranted. Particularly, the solution for finite populations introduces complicated correlation effects. It would be worthwhile to provide more nuanced predictions on the impact of social structure on equilibrium actions. As mentioned above, if posterior beliefs are correlated across nearby players, particularly popular signals will provide information on the state of the world (as always), but at the same time they will also provide information on others' equilibrium actions. Equilibrium considerations therefore will depend on a more detailed description of the network structure than that provided solely by the degree distribution. Secondly, while this paper obtains predictions as to the type of network that maximizes the probability of success (in terms of degree distributions), it does not, for the time being, provide predictions as to the most efficient network. These need not coincide if those networks that maximize the probability of success also increase the size of unsuccessful attacks. These considerations are left for future research.

2.9 Appendix (Sparseness Condition for Approximating Correlated Networks)

2.9.1 Background

The following discussion proves we can impose conditions on the network architecture that guarantee correlations across players are sufficiently local and can be disregarded when dealing with large populations. Recall that players optimally choose a threshold strategy following an updated belief about the share of the population that will choose to attack the status quo. In particular, players must form beliefs about the value of A defined as the share of the population that choose $a_i = 1$

$$A = \frac{1}{n}\sum_{d=1}^{D}\sum_{j\in N_d} 1_{\{x_j\leq x_d^*\}} = \sum_{d=1}^{D}\frac{n_d}{n}\left(\frac{1}{n_d}\sum_{j\in N_d} 1_{\{x_j\leq x_d^*\}}\right)$$

With a finite population, A is a binomial random variable and calculating the equilibrium translates into a grueling combinatorics exercise that requires calculating, for each player, the level of correlations with every other individual in order to form high-level posterior beliefs. For networks exceeding 5 players the calculations become highly intractable and provide little insight. Instead, we propose approximating a large network with an infinite population in order to rid the model of correlation effects and focus instead on the relation between connectivity and informativeness. Formally, we provide conditions on the network architecture that guarantees that,

$$\lim_{n\to\infty}\left(\frac{1}{n_d}\sum_{j\in N_d} 1_{\{x_j\leq x_d^*\}}\right) = Pr\left(x_d \leq x_d^\star\right)$$

Then we have that

$$A \longrightarrow \sum_{d=1}^{D} p_d \cdot Pr\left(x_d \leq x_d^\star\right)$$

where p_d is the share of the population with degree d. Not only do we gain in tractability, but we are now able to express the equilibrium as a fixed prediction of success/failure and size/composition of attacks for each value of θ- the alternative would provide for each state of the world θ a probability of success and failure and a distribution of possible sizes and compositions of attacks.

In short, the network here introduces two effects: local correlation in private information, and precision stemming from the connectivity of each individual. The following is a methodological contribution for ridding the model of the former effect in order to exploit the impact of the latter effect on the equilibrium.

2.9.2 Finite-Range Dependence and Strong Mixing Sequences

Lemma 2.9.1. *If a sequence* $\{x_i\}_{i=1}^{\infty}$ *of random variables (with the same mean) exhibits finite-range dependence- i.e. there exists an* I *such that if* $\mid i - i' \mid \geq I$ *then* x_i *and* $x_{i'}$ *are independent- and if the random variables emerge from some finite-moment generating function, then the LLN applies.*

Proof. Construct a new sequence $\{y_i\}_{i=1}^{\infty}$ in the following way: for $i = 1, 2, \ldots, I$ $y_i = x_i$, $y_{I+i} = \frac{1}{2}(x_i + x_{I+i})$, $y_{2I+i} = \frac{1}{3}(x_i + x_{I+i} + x_{2I+i})$ and so on for all integer multiples of I . In general for $n \in Z_+$ and for $i = 1, 2, \ldots, I$, we have that $y_{nI+i} = \frac{1}{n+1}\sum_{k=0}^{n} y_{kI+i}$. We have constructed a new sequence from sums of the original sequence in a way that guarantees that the finite-range dependence is preserved. Now we can talk about limits of the sequence and use the well known fact that if the partitions of a sequence all converge to the same limit, then the original sequence must also converge to this limit.

Consider partitioning the sequence $\{y_i\}_{i=1}^{\infty}$ into I sub-sequences in the following sense: first take $\{y_1, y_{I+1}, y_{2I+1}, \ldots\}$, then $\{y_2, y_{I+2}, y_{2I+2}, \ldots\}$, and so on. Then we can be sure from the way we have constructed the sequence y_n that each of these sub-sequences contains independent and identically distributed random variables so that SLLN applies- i.e. such that each of these sequences converges to the same limit $\mathbb{E}[x]$. This implies that the entire sequence y_n converges to $\mathbb{E}[x]$ - There is a theorem: if a set of sub-sequences that covers the original sequence converge *to the same point*, then the sequence also converges to that point. But if y_n converges, then so must the sequence composed from summing together blocks of I elements in the sequence y_n. This is clear from the definition of convergence (pick a larger N but still should get close to the limit). Formally define the sequence $z_n = \frac{1}{I}\sum_{k=(n-1)I+1}^{nI} y_n$. Then we have that $z_1 = \frac{1}{I}\sum_{i=1}^{I} x_i$, $z_2 = \frac{1}{2I}\sum_{i=1}^{2I} x_i$ and so on. Because this sequence converges, we get that $\lim_{n\to\infty} \frac{1}{nI}\sum_{i=1}^{nI} x_i = \mathbb{E}[x]$ so that the LLN applies over the original sequence x_n. $\square$

The above proof is mine, but the Lemma is certainly true since it refers to the limiting properties of a particular case of weakly dependent random sequences called "Mixing Sequences". What follows is an attempt to establish a direct translation between this lemma and the theorems that prove SLLN for strong-mixing sequences in the mathematics and statistics literature (in particular Theorem A of Li & Zhang (2010)).

We start by defining a strong-mixing sequence.

Definition. Let $\langle X_n \rangle = \{x_1, x_2, \ldots\}$ be a sequence of random variables defined in a probability space (Ω, F, P), and define a function $\alpha(s)$, called the strong mixing coefficient, as

$$\alpha\left(s\right)=sup\left\{\left|P\left(A\cap B\right)-P\left(A\right)P\left(B\right)\right|:A\in F_{-\infty}^{j},\;\;B\in F_{j+s}^{\infty},\;\;and-\infty<j<\infty\right\}$$

where $F_{a}^{b}\subset F$ denotes the subset of the sigma algebra generated by $\{x_a, x_{a+1}, \ldots x_b\}$. The process $\langle X_n\rangle$ is **Strong Mixing** if $\alpha(s)\to 0$ as $s\to\infty$.

Next we need to show that if a random sequence satisfies finite-range dependence, then it must necessarily satisfy the strong-mixing (or α-mixing) property.

Lemma 2.9.2. *If a sequence satisfies Finite Range Dependence then it also satisfies the Strong Mixing Property.*

Proof. Our definition of finite range dependence in lemma 0.1. can be expressed in terms of sigma-algebras as saying that for all $j\in\mathbb{Z}$, there exists an I such that for every $s>I$ we have $\hat{\alpha}(s,j)=0$, where,

$$\hat{\alpha}\left(s,j\right)=sup\left\{\left|P\left(A\cap B\right)-P\left(A\right)P\left(B\right)\right|:A\in F_{j}^{j},\;\;B\in F_{j+s}^{j+s},\right\}$$

But notice that, by finite range dependence, if x_j is independent of x_{j+s} then it is also independent of x_k for all $k>j+s$. Similarly all x_m with $m<j$ are also independent of x_{j+s} (and consequently also independent of all x_k with $k>j+s$). As a result we can redefine $\hat{\alpha}(s,j)$ as,

$$\hat{\alpha}\left(s,j\right)=sup\left\{\left|P\left(A\cap B\right)-P\left(A\right)P\left(B\right)\right|:A\in F_{-\infty}^{j},\;\;B\in F_{j+s}^{\infty},\right\}$$

In this case it is easy to see that for $\hat{j}\in argmax\left(\hat{\alpha}\left(s,j\right)\right)$ we can establish that $\hat{\alpha}\left(s,\hat{j}\right)=\alpha(s)$. And finally since we have that $\hat{\alpha}\left(s,\hat{j}\right)=0$ for all $s>I$ then this implies that $\alpha(s)\to 0$ as $s\to\infty$. This completes the proof. $\square$

Finally by Theorem A of Li 6 Zhang (2010) that establishes SLLN for strong mixing sequences, we have that Lemma 8.1 is proven.

2.9.3 A Naming Algorithm

Once the above Lemma is shown to hold, we can construct an algorithm for assigning indexes to the players in the network such that the resulting sequence has finite-range dependence. In that case, by the above lemma, the LLN applies to sub-sequences corresponding to players of the same degree, and thus equation (0.1) holds.

Algorithm. Start from any node in the network, call it 1. Assign consecutive indexes to each of $1's$ neighbors. Next, starting from the neighbor with the lowest index, assign consecutive indexes to the neighbors of $1's$ neighbors. If a node is already named, do not rename it. Continue in this way.

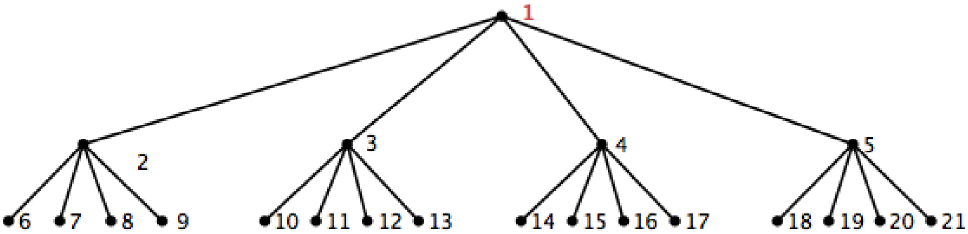

Figure 2.6: The Naming Algorithm for a Tree network with $n = 20$ and $d_1 = 4$, $d_2 = 4$ (i.e. $I = 20$). Notice that for all i and j with $|i - j| > 20$ will necessarily lie more than two links away.

Recall that D represents both the maximal degree and the set containing the degree of each player in the network. Now define $d_1 = max\,(D)$ and $d_2 = max\,(D \setminus \{d_1\})$ Given this naming algorithm, we can find a value of $I < n$ such that the sequence of all players has finite-range dependence. Specifically, we have that when

$$I = d_1\,(1 + d_2)$$

there is finite-range dependence for the entire sequence of nodes. To see this notice that any two players i and j with $|i - j| > d_1\,(1 + d_2)$ must necessarily lie more than two links away from each other. Given our information aggregation procedure, this guarantees that they are not correlated. Of course this is not true for the complete network, but in that case $I = (n - 1)\,n$ which is greater than n for $n > 2$. Clearly the network must be sufficiently sparse such that $I < n$[10]. Although the algorithm gives some freedom as to the precise labeling of the nodes it guarantees that any two nodes with labels I units away will necessarily lie more than two links away from each other. Figure 4 illustrates the algorithm and shows how the value $I = d_1\,(1 + d_2)$ guarantees two degrees of separation for a tree network of 21 players. The reason for using the tree is that, because no neighbor of 1's neighbors is also 1's neighbor, it constitutes the starkest example imaginable.

The only concern now is that not all nodes in the network have the same expected value (different degrees have different success probabilities in their Bernoulli random variables). In other words, the above argument constructs a sequence of Bernoulli random variables with finite-range dependence, but does not guarantee that they correspond to players of the same degree- i.e. with the same success probability $Pr\,(x_d \leq x_d^\star)$. But this is not a problem. Consider constructing D sub-sequences corresponding to players of the same degree. These sub-sequences

[10]In fact, the next section imposes additional conditions on the behavior of I as a function of the total population n. That is, $I < n$ is a necessary, but not sufficient condition for convergence of (0.1).

are also finite-range dependent,[11] and because they only include players of a given degree all necessarily have the same expected value- recall that all players of the same degree choose the same threshold strategy $x_d^{\star}$. Then by Lemma 0.1, the LLN holds for each sub-sequence. Equation (0.1) follows.

The argument above has assumed that I corresponds to a fixed integer. It is easy to see that fixing the maximum degree while increasing the total population increases network sparseness. It turns out, however, that weaker conditions exist. Specifically we can establish the same result for values of I that grow with n, provided the growth rate is sufficiently slow. The following section formalizes this result.

2.9.4 Conditions on the Growth Rate of Degrees

In this section we specify sufficient conditions on the behavior of the largest degrees in the network, $d_1 > d_2 > \ldots$ in order to guarantee that we can construct sequences of nodes with finite-range dependence, and hence the law of large numbers applies. As mentioned above, we need LLN to hold so that we can approximate the network with an infinite population and gain tractability. The algorithm above guarantees that we can construct $I = d_1(1 + d_2)$ sequences of independent random variables, each with $\frac{n}{I}$ elements. Of course, we need that as n tends to infinity, the value of I does not grow too fast so that we can be sure that $\frac{n}{I}$ also tends to infinity. Otherwise the sequences would only contain a finite number of terms (which is not possible). We could just impose that $d_1, d_2, \ldots$ are fixed to some constant, but we are interested in finding weaker conditions. So we need that

$$\frac{n}{I(n)} \underset{n \to \infty}{\longrightarrow} \infty$$

this implies that $I'(n) \longrightarrow 0$.

In general for $I = d_1 + d_1 d_2$ we have that $I'(n) = d_1'(n) + d_1'(n) d_2(n) + d_1(n) d_2'(n)$. We know that in general for any $i \in E$, $d_i'(n) \geq 0$ (otherwise nothing to prove) and so the condition that $I'(n) \to 0$ implies that all three terms go to zero. The first condition simply implies concavity- i.e. $d_1''(n) < 0$. This makes sense, it says that as n grows, the maximum degree should grow at a slower rate. The next two conditions however impose additional conditions on the concavity of these functions. So far I only have results for the case where degrees follow a power function of the entire population.

[11] It is easy to see that if a sequence is finite-range dependent, then so are all of its sub-sequences. The same value of I in fact works.

2.9.5 Power Functions ($a + b < 1$)

let $d_1 = n^a$ and $d_2 = n^b$ the concavity condition implies that $a, b < 1$. Now we seek additional restrictions on a and b in order to guarantee convergence. Notice that $d_1' (n) d_2 (n) = an^{a-1+b}$ so in order to guarantee $d_1' (n) d_2 (n) \to 0$ we need that $a + b < 1$. The same condition results from imposing $d_1 (n) d_2' (n) \to 0$. This means that given our aggregating procedure that defines I and given that we impose that degrees follow a power function of n, the condition for finite-range dependence is $a + b < 1$. So in the case of a regular network (where $d_1 = d_2$) the condition becomes $a < \frac{1}{2}$.

In general we can think of a number of communication protocols that generate all sorts of local correlations. We have assumed in this model that correlations are present up to 2 links of separation. But there is no reason why this should be the case. In general, if correlations emerged at k degrees of separation, then we would need to redefine our I. In this case

$$I = d_1 + d_1 d_2 + d_1 d_2 d_3 + \cdots + d_1 d_2 d_3 \ldots d_k = \sum_{j=1}^{k} \prod_{l=1}^{j} d_l$$

Then if we let $d_1 = n^a$, $d_2 = n^b$, $\ldots$,$d_k = n^k$, following the same argument as before, we can show that the new condition implies that all of the exponents sum to less than 1. That is we need $a + b + c + \cdots + k < 1$. So it should be clear that as the aggregation procedure generates correlations that stretch farther across the network, then the restrictions on the concavity of the degrees becomes stronger in the sense that it imposes structure on the shape of smaller degrees. What is interesting however, is that it does not impose a slower growth on the largest degree d_1 directly. It just requires that the sum of exponents be less than 1. So in fact, it is possible to maintain the same growth rate of d_1 as before, provided that the lower degrees grow slower (or not at all). In fact if all degrees lower than d_2 were fixed to some constant integer, then the condition would be the same for any aggregating procedure.

2.9.6 A General Result

We now provide a general characterization that guarantees $\frac{I(n)}{n} \to 0$. Notice that for any general k the convergence rate of $I(n)$ is determined by the last term in the sum in 2.9.5, so that the necessary condition becomes $\frac{\prod_{j=1}^{k} d_j}{n} \to 0$ or, what is the same, that $\prod_{j=1}^{k} d_j \in o(n)$.

Chapter 3

LEARNING, SORTING, AND TURNOVER IN UNSTABLE ENVIRONMENTS

3.1 Introduction

Complementarities in production incentivizes workers and firms to sort according to their marginal productivity. This result, known as Positive Assortative Matching (PAM), was introduced by Becker (1973) in the context of the marriage market, and has since appeared in a variety of different settings, most importantly in labour market models. For instance, Sattinger (1980) shows that assortativity can explain why the distribution of worker earnings is skewed to the right relative to the distribution of their measured skills. Kremer (1993) uses complementarities in production to explain the wage differences between developing and developed countries that cannot be accounted for by their differences in levels of physical or human capital. More recently the notion of assortative matching has also been applied to frictional search and matching models by Shimer and Smith (2000) and others. However, any reasonable labor market sorting model must account for job turnover as well, and the majority of existing models do not incorporate any form of worker mobility.

More recently, Eeckhout & Weng (2009) analyze a model of learning and sorting that captures job turnover and other relevant worker characteristics over the life cycle. In that model, workers and firms learn continuously about the worker's productivity type, and workers face the possibility of switching to a more suitable match at any given moment.[1] A match therefore provides value to the worker both in terms of competitive wages received, and in the learning experience associated with that match. Indeed, different firms can provide different learning opportunities, and the authors find that, under complementarities in production, PAM obtains even when learning in the more productive firm is slower.[2] A crucial element of their model is that worker types are fixed to either "high" or "low". All market participants, therefore, hold decreasing degrees of uncertainty (on average) along the worker's life-cycle. In fact, after enough time, uncertainty vanishes and workers' types are fully revealed. This implies that the value a particular worker perceives in a given match is determined entirely by the firm and worker types only. In other words, workers are summarized by commonly-held posterior beliefs about their underlying type.

Instead, I consider the possibility that the strength of beliefs is endogenous and related to workers' decisions about switching jobs. I do this by considering an alternative setting in which worker's productivity is not fixed, but instead follows a continuous-time stochastic process independent of output. Productivity can evolve randomly for a number of reasons, ranging from environments sub-

[1] The model can be thought of as bringing together the classical matching framework of Becker with the learning model of Jovanovic (1979) that first accounted for job turnover.

[2] A 3-year job at an investment fund admittedly reveals information differently from a 3-year job at the cheesecake factory.

ject to important technological changes, to settings in which workers suffer non-negligible productivity shocks that tend to disappear on average as the time horizon expands. For instance, a worker's productivity may vary if unforeseen events force him to be assigned to a different task at which his productivity changes, or because tasks itself evolve due to technological progress. These changes also affect an employer's inference process about a worker's ability, as current performance can become a poor predictor of future one.

Uncertainty in skills plays an important role in the evolution of the market's beliefs. While firms can influence an employee's productivity through tailored programs such as compensation schemes, on-the-job training or learning-by-doing, exogenous forces that affect the work environment can also have an important impact on performance. In settings where wages are based on perceived skills, the degree of randomness of the environment is thus expected to influence the strategic behavior of a worker whose ultimate goal is to affect his future income stream by building a good reputation.

Skills evolving randomly implies that market participants now learn over a changing and uncertain environment. As a result, uncertainty no longer vanishes, but instead converges deterministically to a long-run residual strength of beliefs that depends on firm characteristics. This setting has important implications on workers' decisions to switch firms that are absent in the stable environment of Eeckhout and Weng (2009). In particular, it is now possible for the level of uncertainty about a worker's productivity to increase deterministically after switching from one firm to the next. This process of *unlearning* comes from the fact that, in this context, bayesian updating "chases a moving target", and that some firms provide better learning opportunities than others. In the stable environment of Eeckhout and Weng (2009), even as some firms provide better learning opportunities than others, what has been learned so far cannot be unlearned – types are fixed, so that all signals accrued from output reveal additional information about the underlying type. In the current setting, however, the underlying type is itself random, and switching to a firm with slower learning means that output realizations are now less precise in pinpointing the new, random productivity values. As a result uncertainty increases. Taken together, these arguments imply that the option value of learning that comes from switching jobs differs wildly from Eeckhout and Weng (2009).

This paper asks when PAM can be expected as the unique equilibrium configuration in non-stationary environments like this one. I show that PAM obtains if the underlying skills process of workers follows a martingale, and if workers are risk-neutral. The intuition driving these results draws on the continuity of sample paths of posterior beliefs, which rules out "large surprises". This allows us to conclude that, even as workers never cease to learn (contrary to the framework of Eeckhout and Weng (2009)), they nonetheless can hold sufficiently extreme beliefs such that they will remain in the current firm under any reasonable time frame. Moreover, if the underlying skills process satisfies the martingale property (and workers are risk-neutral), expected discounted payoffs are unaffected by the time-dependent path of posterior variance. Indeed, while the strength of beliefs matters for how additional information feeds into future worker types, conditional expectations remain unaffected. This implies that under these assumptions the value function depends only on worker-firm types, and not on the current period directly. PAM follows by arguments similar to Eeckhout and Weng (2009).

After establishing PAM as the unique equilibrium, I extend this environment to allow the possibility that workers can exert a private level of effort that affects the trend of the output process. I follow recent work by Cisternas (2012) that models the classical career concerns model of Holmstrom (1999) in continuous time. Under a deterministic equilibrium, effort levels are found to depend on the market's posterior strength of beliefs. This implies that workers' payoffs now directly depend on the learning process, and the martingale property of skills is not sufficient to retain stationarity. In other words, introducing career concerns leads to non-stationary payoff streams and to value functions that are time-dependent. I find that the no-deviation condition of Eeckhout and Weng (2015), which rules out off-equilibrium deviations, can no longer be framed simply as the second-derivative version of the smooth pasting condition. In this scenario, there is no unique threshold that organizes workers across firm types for all periods. Instead, I characterize the equilibrium in terms of a distribution of threshold types and provide a new characterization of PAM for unstable environments.

3.1.1 Literature

This paper investigates a labor market model of learning similar to others in the literature (Jovanovic (1979, 1984), Moscarini (2005) or Eeckhout and Weng (2009)). However, it also relates very closely to the experimentation literature (Bolton and Harris (1999)) and the literature on continuous time games (Sannikov (2007)), including principal-agent models of information extraction (Sannikov (2007, 2008) and Cisternas (2012)), and mutli-armed bandit problems (Eeckhout and Weng (2015)) .

The model builds on the framework in Eeckhout and Weng (2009) that analyzes learning and turnover in a stationary environment with a continuum of agents, learning in all states, and a competitive spot wage. They specialize to a world where all abilities are simply “high” or “low”, whereas I allow for the underlying skills process of workers to follow a stochastic process. I show this has important implications for the types of learning paths workers can expect upon switching firms. Anderson and Smith (2010) were the first to explore the possibility that workers in a matching model have evolving characteristics. In the employment version of their model, they show PAM can fail because the transition function that maps current types to future types is not a martingale. In line with these results I show that non-stationary environments can retrieve PAM if beliefs evolve as a martingale. I show that if payoffs are directly affected by time-dependent properties of the learning process, non-stationarity alters the no-deviation condition of Eeckhout and Weng (2015).

Section 3.2 introduces the baseline model without effort provision, but with random skills. I show that the evolution of market beliefs is defined by two dynamic processes that track the posterior mean and variance. I show that the martingale property of beliefs allows us to retain stationarity and to pin down the value function for extreme types. This allows us to establish PAM as the unique equilibrium configuration in this setting with persistent learning. In Section 3.3 I introduce the possibility that workers can exert a private level of effort that affects the mean of the output process. These career concern incentives distort payoffs and generate a direct time dependence on the value function. I show that the no-deviation condition looks very different in this setting, and I discuss the implications this has for establishing PAM as an equilibrium allocation. Section 3.4 concludes with a discussion on future paths of research.

3.2 The Model

3.2.1 No Career Concerns

The economy is populated by a unit measure of workers and a unit measure of firms, both infinitely lived. Both workers and firms are ex ante heterogeneous. The firm's type $y \in \{H, L\}$ determines the volatility of output and the marginal productivity given workers' abilities, and is observable to all agents. The economy contains a fraction π of H type firms and $(1-\pi)$ of L type firms. The worker ability θ_t is unobservable to both worker and firms and we assumes evolve according to a stochastic differential equation (SDE). I assume a common prior of θ_0 and the discount rate is $r > 0$. A worker-firm pair (θ, y) produces stochastic output in continuous time and at firm-specific levels of volatility. More concretely, let us define the evolution of output for a match between a firm of type $y \in \{H, L\}$ and a worker of type $\theta \in \mathbb{R}$ as

$$d\xi_{t,y} = A_y \theta_t dt + \sigma_{\xi_y} dZ_t^{\xi} \tag{3.1}$$

where $A_y = \alpha \mathbf{1}_{\{y=H\}} + \beta \mathbf{1}_{\{y=L\}}$ determines the different productivity parameters for high and low firms, θ_t corresponds to a worker's underlying (and unknown) productivity at time t, $Z^{\xi} := \left(Z^{\xi}\right)_{t \geq 0}$ is a one-dimensional Brownian motion and

$\sigma_{\xi_y} > 0$ represents the volatility of the signal's noise component. We assume strict super modularity and worker monotonicity by imposing $\alpha > \beta > 0$.

The evolution of worker productivity (or skills) follows a similar, but independent, stochastic process defined by,

$$d\theta_t = \kappa\theta_t + \sigma_\theta dZ_t^\theta \tag{3.2}$$

where $\sigma_\theta > 0$ represents the volatility of productivity shocks for the worker. The parameter $k \in \mathbb{R}$ will be referred to as the slope of the skills process. For simplicity I assume in most of the analysis that skills evolve as a martingale (i.e. $\kappa = 0$), representing rapidly changing environments and/or workers who easily adjust to new scenarios. However, most of the results hold for $\kappa \neq 0$ since we only require the martingale property of posterior beliefs, which holds even for skills processes that are themselves not a martingale.

Since the market does not observe θ, it must form beliefs about the worker's skills based on observation of ξ. Given equations 3.1 and 3.2 we can formulate the evolution of posterior beliefs held by firm and worker. Notice that since information is symmetric, the following description represents the evolution of beliefs for all market participants. Following standard results from Lipster and Shiryaev (1977) the conditional distribution of θ_t given all available public information $\mathcal{F}_t := \sigma\left(\left(\xi_{s,y}\right)_{s\leq t}\right)$ retains the gaussian structure at all $t \geq 0$. The posterior mean $m_t = \mathbb{E}\left[\theta_t \mid \mathcal{F}_t\right]$ and posterior variance $\gamma_t = \mathbb{E}\left[(\theta - m_t)^2 \mid \mathcal{F}_t\right]$ evolve respectively as,

$$dm_{t,y} = \frac{A_y\gamma_t}{\sigma_\xi}dZ_t \tag{3.3}$$

$$\dot{\gamma}_{t,y} = \sigma_\theta^2 - \left(\frac{A_y\gamma_t}{\sigma_{\xi_y}}\right)^2 \tag{3.4}$$

where $dZ_t = \frac{1}{\sigma_\xi}\left(d\xi_t - A_y\mathbb{E}\left[\theta_t \mid \mathcal{F}_t\right]dt\right)$ is a diffusion process measurable with respect to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$ and captures unexpected movements in output. Equation 3.3 essentially describes the well-known Kalman-Bucy filter for a continuous-time state-space model.

A couple of interesting features are worth noting. First, the evolution of the posterior mean, $m_{t,y}$, preserves the stochastic structure of skills. Second, the posterior mean's response to unexpected innovations increases with the posterior variance, $\gamma_{t,y}$, and with the firm's signal-to-noise ratio, $\frac{A_y}{\sigma_{\xi_y}}$. This means beliefs react more strongly to new information whenever beliefs are less precise or signals are more accurate. Notice that by assuming $\kappa = 0$ we obtain that the evolution of the posterior mean is a martingale, $\mathbb{E}\left[dm_t \mid \mathcal{F}_t\right] = 0$; there is as much good news as bad news to be expected in the future. Most importantly, the posterior variance evolves *deterministically* with time; as opposed to the stable environment in

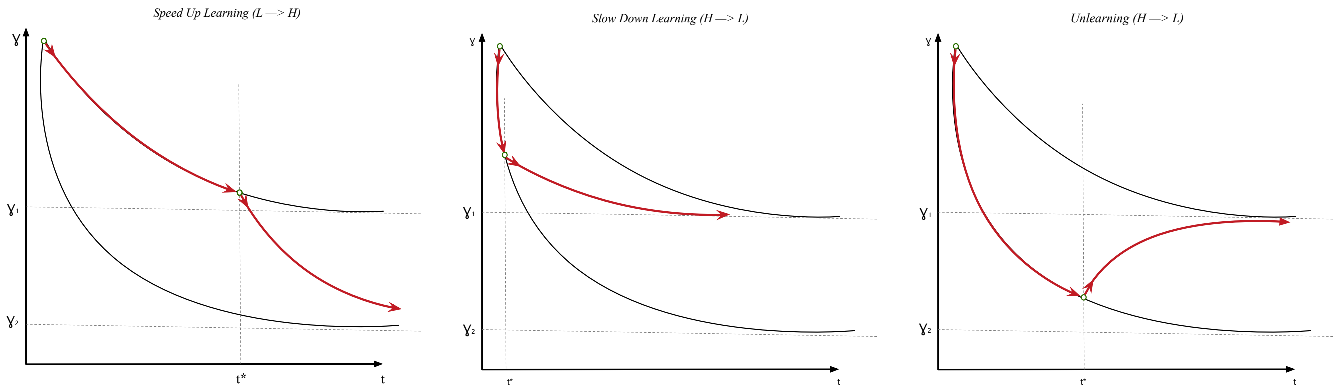

Figure 3.1: Strength of Beliefs with Job Turnover: The first panel shows a situation in which

Eeckhout and Weng (2009), all output realizations here are equally informative. This means the entire trajectory of γ_t is perfectly anticipated by all market participants whenever γ_0 is common knowledge. However, notice that γ_t evolves at different speeds according to firms' signal to noise ratio $\frac{A_y}{\sigma_{\xi_y}}$, which implies, as in Eeckhout and Weng (2009), that switching jobs affects the speed of learning. This setup nonetheless represents a drastic departure from Eeckhout and Weng (2009) because our volatility parameter is a function of time, while theirs depends on the current state of beliefs. Said differently, the evolution of beliefs in Eeckhout and Weng (2009) corresponds to a geometric brownian motion, while the posterior mean here is defined by a diffusion process with time-varying parameters. As explained below, this means that the value function now depends explicitly on time, as well as on currently held beliefs.

As long as skills evolve randomly (i.e. $\sigma_\theta \neq 0$) the model exhibits long-run *residual uncertainty*. The unique, steady-state level of residual uncertainty can be found by setting equation 3.4 to zero. In that case we have that,

$$\gamma_y^\star = \frac{\sigma_\theta \sigma_{\xi_y}}{A_y} \tag{3.5}$$

This long-run measure of uncertainty varies across firms according to the signal-to-noise ratio, and can lead to varying dynamics in the evolution of uncertainty. For instance, imagine that $\frac{\sigma_{\xi_H}}{\alpha} > \frac{\sigma_{\xi_L}}{\beta}$. This means that learning in H-type firms is slower (and converges to a higher level of residual uncertainty) than L-type firms, which exhibit faster learning. The first panel of Figure 1 shows what occurs when a worker switches from an H-type firm to an L-type firm at time $t^\star$: learning accelerates because the posterior variance decreases faster after switching. Conversely, a worker that switches from an L-type firm to an H-type firm experiences a slow-down in learning, or even a period of *unlearning*, depending on the time, $t^{\star}$, where the switch takes place. These two scenarios are shown in the last two panels of figure 1. If the worker switches from an L firm to an H firm early on – before the level of uncertainty falls below the long-run level of the "slower" firm – then learning slows down. This is shown in the second panel of Figure 1. If, on the other hand, the worker switches firms later on – when beliefs are more precise than they could ever be in the "slower" firm – then uncertainty will increase, and beliefs will become more dispersed over time until they converge to the higher level of residual uncertainty. This is shown on the third panel of Figure 1. An immediate consequence of these dynamics is that the time $t^{\star}$ at which workers switch firms determines the future evolution of beliefs about a worker's productivity, by affecting the dynamics of the posterior variance. This phenomenon is entirely absent in the model of Eeckhout and Weng (2009), where worker types are fixed. In that case, all information accrued thus far impacts posterior beliefs equally and no process of unlearning can ever occur. Notice also that, to the extent workers transition across firms in equilibrium, this model exhibits an ever evolving level of uncertainty across the life cycle of the worker.

We may now define an equilibrium as in Eeckhout & Weng (2009). Consider a competitive equilibrium that defines a wage schedule $w_y(m)$ specifying wages for each belief m about a worker's skill. Let $\mathcal{V}_y$ represent the discounted competitive profits of firm y and $r\mathcal{V}_y$ the flow profits. Then we define an equilibrium as follows:

Definition. *In a competitive equilibrium there is a wage schedule $w_y(m) = A_y m - \mathcal{V}_y$ and worker m chooses the firm y with the highest discounted present value. The market clears such that a measure $(1-\pi)$ of workers are employed in the L firm and a measure π in the H firm.*

A couple of remarks are in order. As Eeckhout & Weng (2009) rightly point out, identical types will obtain identical payoffs. More importantly, the definition places restrictions on off-equilibrium prices, requiring that a worker not employed at a firm y on equilibrium must expect a wage $w_y(m)$, in order to guarantee the firm cannot do better if employment suddenly happened. Lastly, wages are spot prices and therefore must be self-enforcing. In this sense, firms define the productive and learning possibilities of workers but themselves are price takers. This essentially boils down to a decision problem for the worker about when to switch firms.

In order to solve the model we need to write down the expected discounted payoff for each worker. Toward this end, let us formulate the worker's problem as follows: a worker that starts off in firm y at time t obtains the following expected present value of payoffs,

$$V_y(m,t) = \mathbb{E}\left[\int\limits_0^{\tau(\bar{m})} e^{-r(s-t)}\left(m_{s,y} - \mathcal{V}_y\right)ds + e^{-r\tau(\bar{m})}V_{-y}(m) \mid \mathcal{F}_t\right]$$

subject to equations 3.3 and 3.4 for all $s > t$. In the equation above, $\bar{m}$ represents a critical value at which workers decide to switch firms, and $\tau(\bar{m}) = inf\{t \geq 0 \mid m_t = \bar{m}\}$ is a stopping time marking the first passage through $\bar{m}$. From this expression we can arrive at a differential equation of the value function using Feynman-Kac's Formula and Itô's Lemma as follows,

$$rV_y(m,t) = w_y(m) + \frac{1}{2}\left(\frac{A_y\gamma_t}{\sigma_\xi}\right)^2 V_y''(m,t) + \frac{\partial V_y(m,t)}{\partial t} \tag{3.6}$$

The arguments above suggest that the deterministic behavior of posterior variance in 3.4 forces the value function to depend on time, t, as well as on worker type, m_t. Indeed, a worker of type m at two different times– t_1 and t_2 – faces two different paths for γ and, as a result, two different paths for how future innovations in output affect beliefs (recall that γ_t affects the stochastic evolution of m_t, as per equation 3.3).[3] Moreover, these paths can look very different, as shown in Figure 1, and can even consist of periods of *unlearning*. The value function captures the total flow payoff to a worker m in firm y. This is equal to the wage plus a second term that captures the option value of learning, similar to Eeckhout and Weng (2009). The third term (absent in Eeckhout and Weng (2009)) accounts for the fact that time now functions as a state variable: holding fixed the worker type, m, the expected future payoff of workers depends on calendar time, by affecting the response to innovations. I show that in the current setting, where skills evolve as a martingale (i.e. $\kappa = 0$) and workers are risk-neutral, the future evolution of γ does not affect current expectations and therefore keeps discounted payoffs constant. As a result, we can establish that only the worker's type, m, functions as a state variable in this particular setting. In the next section I show how relaxing these assumptions can alter results and, in some instances, overturn PAM.

Lemma 3.2.1. *If skills evolve as a martingale (i.e. $\kappa = 0$) and workers are risk neutral, the value function does not explicitly depend on time and equation 3.6 can be simplified as,*

$$rV_y(m) = w_y(m) + \frac{1}{2}\left(\frac{A_y\gamma_t}{\sigma_\xi}\right)^2 V_y''(m) \tag{3.7}$$

[3] Said differently, because the diffusion process of beliefs depends on time-varying coefficients, the expected present value depends on time too.

Proof. Consider the expected present value at time t of a worker that remains in firm y forever. We can write this formally as,

$$V_y(m,t) = \mathbb{E}\left[\int_0^{\infty} e^{-r(s-t)} f(m_s)\, ds \mid \mathcal{F}_t\right]$$

where $f(m_s) = m_{s,y} - \mathcal{V}_y$ corresponds to the flow payoffs, and $\mathcal{F}_t$ is the sigma-algebra generated from all information held at t. Consider the effect of a change in time, holding everything else constant. Because of risk-neutrality on the worker side, $f(m_s)$ is linear in m_s so we can rewrite the above equation as

$$V_y(m,t) = \int_0^{\infty} e^{-r(s-t)} \left(\mathbb{E}[m_s \mid \mathcal{F}_t] - \mathcal{V}_y\right) ds$$

Because skills evolve as a martingale (i.e. $\kappa = 0$) then the evolution of posterior beliefs are themselves a martingale (i.e. $\mathbb{E}[m_{t+h} \mid \mathcal{F}_t] = m_t$ for any $h > 0$), as seen from equation 3.3. This implies that,

$$V_y(m,t) = \int_0^{t} e^{r(t-s)} (m_s - \mathcal{V}_y)\, ds + \frac{1}{r}(m_t - \mathcal{V}_y)$$

Notice that, for any arbitrary time period t, the future expected payoffs for workers only depend on r, m, and $\mathcal{V}_y$. This is because the future looks identical to a risk-neutral worker holding beliefs with martingale property. This implies that, $\frac{\partial V(m,t)}{\partial t} = \lim_{h \to 0} \frac{V(m,t+h) - V(m,t)}{h} = 0$, which proves the result. □

Essentially, Lemma 3.2.1 argues that, while the posterior variance evolves deterministically over time, the expected, discounted future payoffs of a risk-neutral worker only depends on the posterior beliefs, m, and not on the beliefs' dispersion. Indeed, if skills fluctuate randomly with no trend, then the expected behavior of beliefs (and therefore wages) can best be predicted with currently held beliefs, which implies that the current (and future) level of uncertainty does not affect expected payoffs. It follows that the three different types of learning behavior shown in Figure 1 do not affect the value function, and therefore do not affect workers' decisions of where to switch jobs in this particular setting. Of course if workers were risk-averse, or if beliefs themselves exhibited persistence, then time would continue to function as a state variable. The next couple of sections explore similar environments, but in which calendar time remains a crucial determinant of future expected payoffs.

Although learning is a persistent phenomenon in this model, the value of learning can certainly be zero for some workers. It is the option value of switching to a more suitable match that generates the value of learning. Indeed, workers with sufficiently extreme types never change jobs, even as they continually update their type. This implies that, while learning is a persistent phenomenon in this model, we can nonetheless obtain accurate expressions for the value function of some extreme types. This is the content of the next result.

Lemma 3.2.2. *Workers with sufficiently extreme beliefs never switch jobs and obtain no option value from learning. As a result, their value function equals the present value of flow payoffs. Formally,*

$$\underset{m\to\pm\infty}{lim} V_y(m) = \frac{w_y(m)}{r}$$

Proof. Firs we obtain an explicit solution to the Bellman equation (an ordinary differential equation with non-constant coefficients) by standard techniques as:

$$V_y(m) = \frac{w_y(m)}{r} + Be^{\Psi_{y,t}m} + Ce^{-\Psi_{y,t}m} \tag{3.8}$$

where B and C are constants to be determined by the cutoff $\bar{m}$, and where

$$\Psi_{y,t} = \frac{\sigma_\xi\sqrt{2r}}{Y\gamma_t}$$

Now following Dixit (1993), if m is not restricted on the lower side, but instead has an upper barrier at some point $\bar{m}$ then starting from very negative values of m it is very unlikely that we reach $\bar{m}$ at any reasonable future time. Then, the present value of flow payoffs should be a good approximation for the value function. But with $\Psi_1 > 0$, $e^{-\Psi_1 m}$ goes to ∞ as $m \to -\infty$. This would ruin the desired approximation unless $C = 0$. We can make a similar argument for why $B = 0$ as $m \to +\infty$. Taken together, these arguments allow us to conclude that at the extremes the value function must equal the discounted flow payoffs of the worker, or more formally that,

$$V_y(m) \underset{m\to\pm\infty}{\to} \frac{w_y(m)}{r}$$

Plugging the above result into 3.6 we see that $V_y''(m) \underset{m\to\pm\infty}{\to} 0$. Workers with limiting types never switch jobs and obtain no option value from learning. □

Contrary to Eeckhout & Weng (2009), extreme values of m here do not represent a situation where learning never happens. In their model, the response of beliefs to new information vanishes as the underlying skill set of the worker is ultimately revealed. If workers don't learn, their value functions are fully characterized as the discounted flow payoffs. Instead, the speed of learning in this setting is deterministic and converges to a situation with long-run residual uncertainty. Learning here is a persistent phenomenon and new output realizations always shift posterior beliefs, regardless of what those beliefs are. We can be sure, however, that for extreme values of m workers receive their discounted flow payoffs almost surely. After all, sample paths are continuous and no positive probability is placed on reaching a bounded threshold $\bar{m}$ in finite time.

The current environment with randomly fluctuating skills introduces two main differences with respect to the setting in Eeckhout and Weng (2009). Lemma 2.1 and Lemma 2.2 demonstrate that under certain mild conditions, these differences are innocuous, and a similar analysis to Eeckhout and Weng (2009) applies. The following subsection shows how the analysis in Eeckhout and Weng (2009) applies to this setting, and therefore establishes assortative matching as the unique equilibrium in this environment. The next sections explore other type of environments where the previous two lemma's no longer hold and PAM is not guaranteed as the unique equilibrium outcome.

3.2.2 Equilibrium Analysis

As with any model of optimal control we can identify some well-known smoothness properties on the value function that rules out kinks and discontinuities at the barriers where workers decide to switch. These are commonly known as value-matching and smooth-pasting conditions and require that at any equilibrium cutoff the worker must receive the same value across jobs and that the marginal value must also be identical. Formally we can express these conditions as $V_y(\bar{m}) = V_{-y}(\bar{m})$ and $V_y'(\bar{m}) = V_{-y}'(\bar{m})$ respectively. Following Eeckhout and Weng (2015) I show that an equilibrium requires an additional no-deviation condition that rules out deviations from the equilibrium prescriptions.

Next I go on to show some properties of the value function that will allow us to arrive at the equilibrium characterization.

Lemma 3.2.3. *The Equilibrium value functions V_y are strictly convex for $m \in \mathbb{R}$*

Proof. As in Eeckhout & Weng (2009) we can argue that $V_y(m) > \frac{w_y^{a^\star}(m)}{r}$ for all m finite since otherwise all the workers would stay in one firm y forever and markets would not clear. Then, from Lemma 3.2.1 it must be the case that $\frac{1}{2}\left(\frac{Y\gamma_t}{\sigma_\xi}\right)^2 V_y''(m) > 0$ which is only true if V_y is convex. □

Intuitively market clearing requires some workers to work for high firms and some with low firms. This means some worker must change jobs at some point so the value of learning must be positive to make the switch profitable. But the value of learning can only be positive if the value function is convex. Now, since we know what the value function looks like for some extreme beliefs (shown in Lemma 2.2), then we can use convexity and the smooth pasting condition to acquire additional properties of the value function.

Lemma 3.2.4. *The Equilibrium value functions V_y are strictly increasing.*

Proof. Imagine workers with $m < \bar{m}$ work for firm y. It is straightforward that $\underset{m \to -\infty}{lim} V_y'(m) = \frac{A_y}{r} > 0$ and since V_y is strictly convex, $V_y'(m) > 0$ for all $m \in (-\infty, \bar{m})$. At $\bar{m}$ the worker will switch to firm $-y$ but smooth pasting implies $V_{-y}'(\bar{m}) = V_y'(\bar{m}) > 0$. Then strict convexity again ensures $V_{-y}'(m) > 0$ and so on in the case of multiple switching points. This guarantees the value function is increasing over the entire domain. □

With this we can prove a crucial implication of super modularity which allows us to hone in on an equilibrium sorting result.

Lemma 3.2.5. *Under super modularity, in any equilibrium, the limiting "bad" worker $m \to -\infty$ matches with L firm while the limiting "good" worker $m \to \infty$ matches with the H firm. The opposite under strict submodularity. Moreover,*

$$\frac{min(\alpha, \beta)}{r} < V_y'(m) < \frac{max(\alpha, \beta)}{r}$$

where $\alpha = \underset{h \to 0}{lim} \frac{\alpha(m+h) - \alpha m}{h}$ is the infinitesimal change in expected output of a firm y for an infinitesimal increase in the worker type (similarly for β).

Proof. Imagine workers with $m \in (-\infty, \bar{m})$ are employed in firm y and workers with $m \in (\hat{m}, \infty)$ are employed in firm $-y$. We know that $\underset{m \to -\infty}{lim} V_y'(m) = \frac{A_y}{r} < \frac{A_{-y}}{r} = \underset{m \to \infty}{lim} V_{-y}'(m)$ by convexity, so under supermodularity it must be that $\underset{m \to -\infty}{lim} V_y'(m) = \frac{\beta}{r} < \frac{\alpha}{r} = \underset{m \to \infty}{lim} V_{-y}'(m)$, which means workers with extremely low types go to L firm (with productivity parameter β) and the opposite for high types. □

So far we have focused on equilibrium path behavior. In order to properly characterize the equilibrium we need to make sure we are not allowing for any profitable deviation off-equilibrium. As in Eeckhout and Weng (2009) we consider the equivalent in continuous time of a one-shot deviation. That is, we make sure no worker wants to deviate from their equilibrium strategy for a time interval $[t, t + dt)$ and take the limit as $dt \to 0$. We find that in this modified setting we obtain the same condition on the marginal value of learning as Eeckhout and Weng (2009). This result is crucial to determine PAM as the unique equilibrium.

Lemma 3.2.6. *To deter possible deviations, a necessary condition is:*

$$V_H''(\bar{m}) = V_L''(\bar{m})$$

for any cutoff $\bar{m}$*.*

Proof. Let $\Sigma_{t,L} = \frac{1}{2}\left(\frac{\beta\gamma_t}{\sigma_{\xi_L}}\right)^2$ and $\Sigma_{t,H} = \frac{1}{2}\left(\frac{\alpha\gamma_t}{\sigma_{\xi_H}}\right)^2$. Without loss of generality, consider that on equilibrium workers with $m > \bar{m}$ work on H firms (L firms for $m < \bar{m}$). Consider a one-shot deviation from a worker in a high type firm that switches to a low firm for dt at time $\hat{t}$ and then goes back to equilibrium behavior. In this case, the value function is defined as

$$\tilde{V}_L(m) = w_L(m)\,dt + e^{-rdt}E\left[V_H(m+dm)\right]$$

where $dm = \frac{Y\gamma_t}{\sigma_{\xi_L}}dZ_t^{a^\star}$.

Apply Ito's Lemma and get,

$$\tilde{V}_L(m) = w_L(m)\,dt + e^{-rdt}\left[V_H(m,t) + \Sigma_{\hat{t},L}V_H''(m,t)\,dt\right]$$

this implies

$$\lim_{dt\to 0}\frac{\tilde{V}_L(m) - V_H(m)}{dt} = w_L(m) - w_H(m) + \left(\Sigma_{\hat{t},L} - \Sigma_{\hat{t},H}\right)V_H''(m)$$

which must be less than zero for any $m > \bar{m}$. In particular we have for $m \to \bar{m}$ that,

$$w_L(\bar{m}) - w_H(\bar{m}) + \left(\Sigma_{\hat{t},L} - \Sigma_{\hat{t},H}\right)V_H''(\bar{m}) \leq 0$$

$$\implies w_L(\bar{m}) + \Sigma_{\hat{t},L}V_L''(\bar{m}) - w_H(\bar{m}) - \Sigma_{\hat{t},H}V_H''(\bar{m}) + \Sigma_{\hat{t},L}\left(V_H''(\bar{m}) - V_L''(\bar{m})\right) \leq 0$$

by value matching at $\bar{m}$ this implies that,

$$V_H''(\bar{m}) \leq V_L''(\bar{m})$$

Similarly we can consider a one shot deviation from a $m < \bar{m}$ worker moving to a high type firm for dt before switching back to equilibrium behavior. A similar argument allows us to conclude that in this scenario delivers the following condition

$$V_L''(\bar{m}) \leq V_H''(\bar{m})$$

Together these two conditions imply that in order to prevent any one-shot deviation from equilibrium behavior it must be the case that $V_H''(\bar{m}) = V_L''(\bar{m})$. $\square$

With all this we can now proceed to characterize the equilibrium allocation. We show that, under strict supermodularity, it is impossible to have two cutoffs $m_1 < m_2$ such that workers with $m < m_1$ match with low-type firms, workers with $m \in [m_1, m_2]$ match in high-type firms, and workers with $m > m_2$ match in low-type firms.

Theorem 3.2.7. *PAM is the unique stationary competitive equilibrium allocation under supermodularity.*

Given Lemma 3.2.5, workers with sufficiently low $m's$ will accept low type offers and workers with high $m's$ will accept high-type offers. Then, all we need to prove is that, under supermodularity, it is impossible to have a worker first accept low type offers, then accept high-type offers and finally accept low-type offers again. If this is ruled out, there must exist a unique cutoff $\bar{m}$, such that $m < \bar{m}$ will accept low offers and $m > \bar{m}$ will accept high type offers. This corresponds to the PAM allocation. In other words, all we need to show in order to prove the above theorem is the following claim,

Claim 3.2.8. Under strict supermodularity it is impossible to have $m_1 < m_2$ and equilibrium value function V_H V_{L1} and V_{L2} such that

$$V_H(m_1) = V_{L1}(m_1) \quad \textit{and} \quad V_H''(m_1) = V_{L1}''(m_1)$$

$$V_H(m_2) = V_{L2}(m_2) \quad \textit{and} \quad V_H''(m_2) = V_{L2}''(m_2)$$

are satisfied simultaneously.

Proof. By contradiction suppose claim 0.6 is true. Then, by the value matching condition we have

$$w_H(m_1) + \Sigma_H V_H''(m_1) = w_L(m_1) + \Sigma_L V_{L1}''(m_1)$$

and

$$w_H(m_2) + \Sigma_H V_H''(m_2) = w_L(m_2) + \Sigma_L V_{L2}''(m_2)$$

Let $s_H = \frac{\alpha}{\sigma_{\xi_H}}$ and $s_L = \frac{\beta}{\sigma_{\xi_L}}$ then we can invoke the no-deviation condition and rewrite the above equations as:

$$\frac{s_H^2 - s_L^2}{s_H^2} r V_H(m_1) = w_L(m_1) - \frac{s_L^2}{s_H^2} w_H(m_1)$$

and

$$\frac{s_H^2 - s_L^2}{s_H^2} r V_H(m_2) = w_L(m_2) - \frac{s_L^2}{s_H^2} w_H(m_2)$$

Taken together we have that:

$$\frac{s_H^2 - s_L^2}{s_H^2} r \left[V_H\left(m_2\right) - V_H\left(m_1\right)\right] = \left[w_L\left(m_2\right) - w_L\left(m_1\right)\right] - \frac{s_L^2}{s_H^2}\left[w_H\left(m_2\right) - w_H\left(m_1\right)\right]$$

Now since V_H is convex and $V_H'\left(m_1\right) > \frac{\beta}{r}$ by Lemma 0.3, then we have the following

$$\frac{s_H^2 - s_L^2}{s_H^2}\beta\left(m_2 - m_1\right) < \beta\left(m_2 - m_1\right) - \frac{s_L^2}{s_H^2}\alpha\left(m_2 - m_1\right)$$

[4]. This implies $\alpha < \beta$, a contradiction to supermodularity! □

Theorem 3.2.7 extends the results of Eeckhout and Weng (2009) into unstable environments characterized by residual uncertainty and very diverse learning environments. It establishes that under supermodularity, even if workers' skills are subject to some form of random fluctuation, workers with better posteriors about their ability sort into more productive jobs. In other words, productivity considerations dominate the learning advantages in this setup because competitive wages adjust and offset the difference in learning speeds. The arguments from Eeckhout and Weng (2009) apply in this environment given lemmas 2.1 and 2.2. In the next section, I show what happens in setting where Lemma 2.1 fails to hold.

3.3 Introducing Career Concerns

In this section I consider the possibility that workers can exert a hidden effort level that affects the trend of the output process. In doing so, workers can attempt to manipulate the inference process of firms, in order to command higher wages.[5] Formally, we modify the description of output in equation 3.1 for the following expression,

$$d\xi_{t,y} = \left(A_y\theta_t + a_t\right)dt + \sigma_{\xi_y} dZ_t^\xi \tag{3.9}$$

where, as before, $A_y = \alpha\mathbf{1}_{\{y=H\}} + \beta\mathbf{1}_{\{y=L\}}$ determines the different productivity parameters for high and low firms, θ_t corresponds to a worker's productivity level at time t, $Z^\xi := \left(Z^\xi\right)_{t\geq 0}$ is a one-dimensional Brownian motion, and $\sigma_{\xi_y} > 0$ represents the volatility of the signal's noise component. The additional term, a_t,

[4]Here as in Eeckhout and Weng, we are using that $r\mathcal{V}_{L1} = r\mathcal{V}_{L2}$ which holds by the assumption of perfect competition.

[5]As with the majority of these career concerns models, along the equilibrium path the firm perfectly anticipates the effort level of the worker, and is therefore not fooled. However the worker is trapped into exerting effort given the market's expectations. That is, effort provision is sustained purely by off-equilibrium beliefs.

corresponds to the worker's effort choice, and I assume it enters additively in the trend component of output. Notice that workers' effort profile will affect the type of information that the market can obtain about the underlying skills process. I again impose strict super modularity and worker monotonicity by assuming that $\alpha > \beta > 0$. The evolution of skills remains unchanged, and is described in equation 3.2.

Given equations 3.9 and 3.2 and a conjectured effort level by the firm, $a^{\star}$, we can formulate the evolution of posterior beliefs by the firm. Following standard results from Lipster and Shiryaev (1977) the conditional distribution of θ_t given all available public information $\mathcal{F}_t$ retains the gaussian structure at all $t \geq 0$ and the posterior mean $m_t^{\star} = \mathbb{E}^{a^{\star}}\left[\theta_t \mid \mathcal{F}_t\right]$ and posterior variance $\gamma_t^{\star} = \mathbb{E}^{a^{\star}}\left[(\theta - m_t^{\star})^2 \mid \mathcal{F}_t\right]$ evolve respectively as,

$$dm_{t,y}^{\star} = \frac{A_y \gamma_t}{\sigma_{\xi}} dZ_t^{a^{\star}} \tag{3.10}$$

$$d\gamma_{t,y}^{\star} = \left(\sigma_{\theta}^2 - \left(\frac{A_y \gamma_t}{\sigma_{\xi_y}}\right)^2\right) dt \tag{3.11}$$

where $dZ_t^{a^{\star}} = \frac{1}{\sigma_{\xi}}\left(d\xi_t - \left(a_t^{\star} + A_y m_t^{\star}\right) dt\right)$ is a diffusion process measurable with respect to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$ and captures the unexpected movement in output. Notice that the evolution of the posterior mean and variance are very similar to the baseline model with no career concerns. This implies that all the arguments above about residual uncertainty and learning behavior extends to this setting as well. However, I will show that with career concerns these elements severely transform the problem at hand.

Indeed we could allow for the worker to have additional sources of information other than the past history of output realizations. We will define $\mathbb{F}^w \equiv (\mathcal{F}_t^w)_{t\geq 0}$ as the worker's information structure, containing, in particular, the public information generated by output $\mathbb{F} \equiv (\mathcal{F}_t)_{t\geq 0}$. Given a strategy $a \in \mathcal{A}$ the worker's posterior mean evolves as $dm_{t,y} = \sigma_{t,y} dZ_t^w$ where Z_t is an $\mathbb{F}^w$-progressively measurable Brownian motion and σ_{ty} is any non-negative process. If instead we impose that the worker has no additional information other than the effort level, then the posterior mean would evolve as $dm_{t,y} = \frac{Y \gamma_t}{\sigma_{\xi}} dZ_t^a$ and, since along the equilibrium path it must be that $a = a^{\star}$, then we have symmetric learning across worker and firm. We assume the latter for the time being.

I follow Cisternas (2012) in solving for an equilibrium profile of effort $a^{\star}$ that only depends on the evolution of the market's posterior variance. The reason for focusing on equilibria that are deterministic is, first, that forcing the solution to depend on output directly is technically difficult, and, more importantly, that workers' incentives would otherwise be distorted away from pure reputational considerations. Holmstrom (1999) also focuses on what he calls non-contingent equilibria in his seminal work on career concerns. Indeed, the scope of this paper is to determine the sorting effects from introducing career concern incentives in the classical way of Holmstrom.

Definition. *A public equilibrium effort level, $a^\star$, is deterministic if it depends only on the evolution of the market's posterior variance.*

As before, we assume a competitive spot wage that pays workers' expected productivity minus firms' profits. In this case, we can define the wage as,

$$w_y^{a^\star}(m,t) := \lim_{h\to 0}\frac{\mathbb{E}^{a^\star}\left[\xi_{t+h}\mid\mathcal{F}_t\right]-\xi_t}{h} - \mathcal{V}_y = a_t^\star + m_t^\star - \mathcal{V}_y \tag{3.12}$$

where $\mathcal{V}_y$ is firm $y's$ profits. Notice that, unlike in the previous section, the deterministic evolution of the posterior variance now directly affects payoffs. This is because the effort level $a^\star$, which enters the wage directly, responds to γ only by the assumption of deterministic public equilibria. This implies that Lemma 2.1 is not applicable in this setting, and the value function therefore depends explicitly on time, as shown below.

The worker's problem can be defined as follows: given a conjecture $a^\star$, a worker that starts off in firm y solves at all periods $t\geq 0$ and for any history $\mathcal{F}_t$ the following problem:

$$\max_{a\in\mathcal{A}}\quad V_y(m,t) = \mathbb{E}^a\left[\int_0^{\tau(\bar{m})} e^{-r(s-t)}\left(w_y^{a^\star}(m,s)-g(a_s)\right)ds + e^{-r\tau(\bar{m})}V_{-y}(m)\mid\mathcal{F}_t\right] \tag{3.13}$$

subject to (for all $s>t$)

$$dm_s^\star = -\delta_{s,y}m_s^\star dt + \delta_{s,y}\left(d\xi_{s,y}-a_s^\star ds\right)$$

$$d\xi_{s,y} = \left(a_s + A_y m_s\right)ds + \sigma_\xi dZ_s^a$$

$$m_s = m_0 + \frac{A_y}{\sigma_\xi}\int_0^s \gamma_s dZ_s$$

where we define $\delta_{s,y}=\frac{A_y\gamma_t}{\sigma_\xi^2}$ and where $\tau(\bar{m}) = inf\left\{t\geq 0\mid m_t=\bar{m}\right\}$ is a stopping time marking the first passage through $\bar{m}$. Notice the incentives for signal jamming that arise by affecting the evolution of output, in turn affecting the evolution of the firm's posterior mean. The worker is aware that by deviating from $a^\star$, she can affect the evolution of $m^\star$. In other words, the worker may induce a distribution over outcome paths that differs from the one anticipated by the market ($\mathbb{E}^a[\cdot]$ operator). This in turn affects the market's beliefs about how skilled the worker is. As a consequence, an increase in effort today generates, on average, a boost in the reputational component of future wages.

It is clear from the problem above that the optimal strategy $a \in \mathcal{A}$ should in principle depend on the cutoff $\bar{m}$ at which the worker decides to switch firms. In other words, a comprehensive model should have $a(\tau(\bar{m}))$. However this would mean that actions depend the stochastic history which we rule out since we only deal with non-contingent strategies a la Holmstrom. As such, we solve the model as if actions were taken sequentially: that is, choose an action profile for a given match and then given these equilibrium actions fixed, we choose an optimal switching cutoff.

Proposition 3.3.1. *The unique Equilibrium in deterministic and adapted strategies is characterized by the first order condition*

$$g'\left(a^{\star}_{t,y}\right) = \frac{A_y^2 \gamma_{t,y}}{\sigma^2_{\xi_y}} \int_t^{\infty} e^{-\int_t^s \left(r + \frac{A_y^2 \gamma_{u,y}}{\sigma^2_{\xi_y}}\right) du} ds \tag{3.14}$$

and $\frac{da^{\star}_{t,y}}{dt} \leq 0 \Longleftrightarrow \frac{d\gamma_{t,y}}{dt} \leq 0.$

Proof. See Cisternas (2012) - Prop.10 □

A couple of important remarks are in order. First, recall that as γ_t decreases over time, the sensitivity to new information decreases. This in turn decreases workers' benefits from signal jamming, since beliefs become less uncertain, and thus less flexible to new information. This goes in line with the traditional idea that career concerns motives generate higher returns under more uncertainty. However, it can be shown that as γ decreases past beliefs are held longer in employers' estimations of a particular worker. As explained extensively by Cisternas (2012), the second result in Proposition 3.3.1 argues that, in a standard career concerns environment a la Holmstrom, short-term losses losses outweigh any long-term benefits from persistent distortions in beliefs. Secondly, Proposition 3.3.1 clarifies the form in which effort moves with γ. As γ decreases over time, so does the equilibrium effort $a^{\star}_t$. Indeed equilibrium effort profiles decrease deterministically from some initial point and similarly converge to a long-run steady state level of effort.

In order to make progress we need to know more about the behavior of $a^{\star}$. In particular, it is important to determine whether equilibrium effort not only decreases together with γ, but whether its rate of change is also proportional to that of γ. Knowing this and the steady-state level of effort allows us to rank equilibrium effort levels across firms at any point in time according to their signal-to-noise ratios . This is the content of the following lemma.

Lemma 3.3.2. *The equilibrium level of effort satisfies* $\frac{d^2 a^\star_{t,y}}{dt^2} \geq 0 \iff \frac{d^2 \gamma_{t,y}}{dt^2} \geq 0$. *As a result, if* $\frac{A_L}{\sigma^2_{\xi_L}} > \frac{A_L}{\sigma^2_{\xi_L}}$ *then* $a^\star_{t,L} \geq a^\star_{t,H}$ *for all* $t \geq 0$.

Proof. To prove the first statement let $p_{t,y} = \frac{A_y^2 \gamma_{t,y}}{\sigma^2_{\xi_y}}$, $\lambda_{t,y} = \int_t^\infty e^{-\int_t^s r + p_{u,y} du} ds$, and $l_{t,y} = p_{t,y}\lambda_{t,y}$. One can show that,

$$\frac{d^2 log\left(l_{t,y}\right)}{dt^2} = \frac{\ddot{\gamma}_t}{\gamma_t} + \frac{\dot{\gamma}_t}{\gamma_t}\left(p_{t,y} - \frac{\dot{\gamma}_t}{\gamma_t}\right) + \frac{1}{\lambda_{t,y}}\left(r + p_{t,y} - \frac{1}{\lambda_{t,y}}\right)$$

and I drop the firm-type index, y, on the $\gamma_t' s$ where possible for convenience. From the ODE that governs γ_t, (see equation 3.4) it can be shown that $\frac{\dot{\gamma}_{t,y}}{\gamma_{t,y}} = \frac{\sigma_\theta}{\gamma_{t,y}} - p_{t,y}$ and that $\frac{\ddot{\gamma}_{t,y}}{\dot{\gamma}_{t,y}} = -2p_{t,y}$. This implies that

$$\frac{d^2 log\left(l_{t,y}\right)}{dt^2} = -\frac{\sigma_\theta \dot{\gamma}_t}{\gamma_t^2} + \frac{1}{\lambda_{t,y}}\left(r + p_{t,y} - \frac{1}{\lambda_{t,y}}\right)$$

Now suppose $\gamma_t > \gamma^\star$ which occurs if and only if $\dot{\gamma}_t \leq 0$ and $\ddot{\gamma}_t \geq 0$ (from the definition of γ in 3.4). Then,

$$\lambda_{t,y} < \frac{1}{r + \frac{A_y^2 \gamma^\star}{\sigma^2_{\xi_y}}} := \frac{1}{r + p^\star_y}$$

implying that,

$$\frac{d^2 log\left(l_{t,y}\right)}{dt^2} > -\frac{\sigma_\theta \dot{\gamma}_t}{\gamma_t^2} + \frac{1}{r + p^\star_y}\left(r + p_{t,y} - \frac{1}{r + p^\star_y}\right) > 0$$

To prove the second statement, notice that at steady state equilibrium effort levels can be ranked. In particular we can substitute the steady state level of γ into the first order condition for effort and obtain the following expression.

$$g'\left(a^\star_y\right) = \frac{\sigma_\theta A_y}{r\sigma_{\xi_y} + \sigma_\theta A_y}$$

It is clear from this that if $\frac{A_L}{\sigma^2_{\xi_L}} > \frac{A_L}{\sigma^2_{\xi_L}}$ and $g\left(\cdot\right)$ is convex, then $a^\star_L > a^\star_L$ in steady state. Now, if $\frac{A_L}{\sigma^2_{\xi_L}} > \frac{A_L}{\sigma^2_{\xi_L}}$ then for all $t \geq 0$ it must be the case that $\dot{\gamma}_{t,L} < \dot{\gamma}_{t,H} < 0$. Together this implies that $a^\star_{t,L} \geq a^\star_{t,H}$ for all $t \geq 0$. □

Lemma 3.3.2 states that 'signal jamming' incentives move in tandem with the posterior mean-squared error at a speed proportional to γ. This implies that the various learning experiences outlined in Figure 1 can be thought to operate similarly in the effort choices of workers, and, therefore, that effort levels can increase during the episodes of *unlearning,* as described above. The main difference now is that the behavior of γ directly affects payoffs through its impact on $a^{\star}$, whereas in the previous section γ only imposed second-order effects on beliefs, which did not affect current expectations. Moreover, Lemma 3.3.2 establishes that firms with faster learning experience a more rapid decrease in signal-jamming incentives while converging to a higher steady-state level of effort than slower learning firms. This is explained by the fact that firms with faster learning indeed obtain a lower long-run level of uncertainty, which decreases the short-term incentives to distort market expectations, but increase the long-run benefits from past distortions.

We can arrive at an expression for the value function by first simplifying the constraints in (3.13) to get a unique expression for $m_s^{\star}$ as a function of parameters. The first and second constraint, together with the fact that on equilibrium firms perfectly anticipate the optimal effort level of workers (i.e. $a = a^{\star}$), can be expressed together as,

$$dm_s^{\star} = \delta_{s,y}\left(m_s - m_s^{\star}\right) ds + \delta_{s,y}\sigma_{\xi} dZ_s^{a^{\star}} \tag{3.15}$$

where it is clear that asymmetric learning generates beliefs that are not a martingale and drift according to the difference in beliefs across market participants. However, we assume for now that the filtration from which beliefs are updated is identical across all agents.[6] This implies the third constraint is in fact identical to equation 3.10 which means that $m_s = m_s^{\star}$ for all s on the equilibrium path. The three constraints, therefore, are expressed together as in equation 3.10. Intuitively, in a career concerns model the firm is never fooled on equilibrium, and if we assume the only additional information for the worker is his effort level, then learning must be symmetric on equilibrium.

Now we are ready to express the value function using Feynman-Kac as,

$$rV_y\left(m,t\right) = \tilde{w}_y^{a^{\star}}\left(m,t\right) + \frac{1}{2}\left(\frac{Y\gamma_t}{\sigma_{\xi}}\right)^2 V_y''\left(m,t\right) + \frac{\partial V_y\left(m,t\right)}{\partial t} \tag{3.16}$$

where $\tilde{w}_y^{a^{\star}}\left(m\right) = a_t^{\star} + m_t^{\star} - g\left(a_t^{\star}\right) - \mathcal{V}_y$ is the competitive spot wage paid to the worker net of effort costs (i.e. $\tilde{w}_y^{a^{\star}} = w_y^{a^{\star}} - g\left(a_t^{\star}\right)$). The solution to this second

[6]In other words, the only additional information workers have with respect to firms is their private effort provision. But this is correctly anticipated by firms in equilibrium, so that learning remains symmetric.

order partial differential equation is complicated to find (if it exists analytically), but we can make progress by working out some properties of the last term in equation 3.16. Recall from Lemma Lemma 3.2.1 that the direct effect from an infinitesimal change in time on expected future beliefs is zero. Therefore, we only need to consider the direct effect on payoffs, given that equilibrium effort depends directly on the time evolution of γ. In other words, we have the following relationship,

$$\frac{\partial V_y\left(m,t\right)}{\partial t}=\frac{\partial V_y\left(m,t\right)}{\partial a_t}\frac{\partial a_t}{\partial t}$$

I show that this derivative is negative in the martingale setting (i.e. $\kappa=0$) for all t.

Lemma 3.3.3. *If skills evolve as a martingale (i.e.* $\kappa=0$*) and workers exert hidden effort as in* (3.14)*, the value function depends explicitly on time and the derivative is negative,*

$$\frac{\partial V_y\left(m,t\right)}{\partial t}<0$$

Proof. From Lemma 3.2.1 we know that when beliefs are a martingale time has no first-order impact on posterior expectations and, if payoffs only depend on market beliefs, the value function is unaffected. We therefore only need to consider the direct effect of an infinitesimal change in time on payoffs. Given the direct relationship between effort a_t and the time evolution of γ_t, we have that

$$\frac{\partial V_y\left(m,t\right)}{\partial t}=\frac{\partial V_y\left(m,t\right)}{\partial a_t}\frac{\partial a_t}{\partial t}$$

Now, the first term on the right-hand-side of this equation corresponds to the direct effect on flow payoffs at time t of a marginal increase in a_t. Given the definition of $\tilde{w}_y^{a^\star}\left(m\right)$ in equation 3.16 above, we have that $\mathbb{E}^{a^\star}\left[\frac{\partial V_y(m,t)}{\partial a_t}\right]=1+\mathbb{E}^{a^\star}\left[\frac{\partial m_t}{\partial a_t}\mid\mathcal{F}_t\right]-g'\left(a_t^\star\right)=1-g'\left(a_t^\star\right)$, where the last equality comes from the martingale property of m_t.

I claim that, given $\kappa=0$, then $g'\left(a_t\right)<1$ for all $r>0$. To see this, substitute the steady state value of γ in equation (3.5) into the optimality condition for effort in equation (3.14). In steady state, effort is constant, say to $a^\star$, and characterized by the first order condition,

$$g'\left(a^\star\right)=\frac{\sigma_\theta A_y/\sigma_\xi}{r+\sigma_\theta A_y/\sigma_\xi}$$

This implies that $g'\left(a^\star\right)\nearrow 1$ as $r\to 0$. Then, for all $r>0$, we have that $g'\left(a^\star\right)<1$. This implies that $\frac{\partial V_y(m,t)}{\partial a_t}>0$. Finally notice that, by Proposition

3.3.1 and equation 3.4, the following is true: $\frac{\partial a_t}{\partial t} \iff \frac{\partial \gamma_t}{\partial t} < 0$ Together this implies that $\frac{\partial V_y(m,t)}{\partial a_t}\frac{\partial a_t}{\partial t} < 0$, which proves the desired result. $\square$

Lemma 3.3.3 allows us to sign the derivative of the value function with respect to time by noticing that equilibrium effort is persistently bellow the efficient benchmark $g'(a^\star) = 1$. I show below that we can establish convexity as before and, as a result, extend some of the properties of the value function that we established in the previous section. However, I also show that the no-deviation condition looks quite different from before and I argue that this might work against establishing PAM as the unique equilibrium configuration.

3.3.1 Equilibrium Analysis

The arguments above help us to derive some properties of the value function that will allow us to arrive at the equilibrium configuration. I show that we can use Lemma 3.3.3 to establish convexity, and to pinpoint the value function for extreme types. However, I also show that with career concerns the no-deviation condition can no longer be framed as the second derivative version of the smooth-pasting condition.

Lemma 3.3.4. *The Equilibrium value functions V_y are strictly convex for $m \in \mathbb{R}$*

Proof. As in Eeckhout & Weng (2009) we can argue that $V_y(m) > \frac{w_y^{a^\star}(m)}{r}$ for all m finite since otherwise all the workers would stay in one firm y forever and markets would not clear. Then, from Lemma 3.3.3 $\frac{\partial V_y(m,t)}{\partial t} < 0$, and from equation 3.16 it must be the case that $\frac{1}{2}\left(\frac{Y\gamma_t}{\sigma_\xi}\right)^2 V_y''(m) > 0$ which is only true if V_y is convex. $\square$

After we sign the movement of the value function over time in Lemma 3.3.3, convexity follows quite naturally. More importantly, once we have established that the value function is convex, we use the smooth pasting condition to conclude that it must also be increasing and, using the arguments for extreme types in Lemma 3.2.2, we can also establish that supermodularity pushes extremely bad workers to work with L type firms and extremely good workers to work with the H type firm. In other words, Lemma 3.2.4 and Lemma 3.2.5 follow immediately as before.

So far, the basic properties of the value function that existed in the baseline model without career concerns also hold in this environment. Indeed, since we can establish convexity, and since we can be sure that extreme types stay in the same firm forever, the other lemmas follow through as before. However, payoffs now directly depend on the current strength of market beliefs, via its impact on the equilibrium profile of effort. [7] As shown in Figure 3.2.1, the strength of beliefs evolves in very distinct paths according to the time and type of switching that takes place. This has important implications on the conditions that must hold in order to rule out possible deviations from equilibrium allocations. As before, we are after a condition on the value function that deters workers from one-shot deviations away from equilibrium. However, the following lemma makes clear how non-stationary environments such as this one completely alter the standard no-deviation condition of Eeckhout and Weng (2015).

Lemma 3.3.5. *In the model with Career Concerns, a necessary condition to deter possible deviations is:*

$$\begin{cases} V_H''\left(\bar{m},t\right) \geq V_L''\left(\bar{m},t\right) & if\ \frac{A_H}{\sigma_{\xi_H}} < \frac{A_L}{\sigma_{\xi_L}} \\ V_H''\left(\bar{m},t\right) \leq V_L''\left(\bar{m},t\right) & if\ \frac{A_H}{\sigma_{\xi_H}} > \frac{A_L}{\sigma_{\xi_L}} \end{cases}$$

for any cutoff $\bar{m}$ and for all periods $t \geq 0$.

Proof. I consider the first case only (the second case follows an identical argument). Let $\Sigma_{t,L} = \frac{1}{2}\left(\frac{\beta\gamma_t}{\sigma_{\xi_L}}\right)^2$ and $\Sigma_{t,H} = \frac{1}{2}\left(\frac{\alpha\gamma_t}{\sigma_{\xi_H}}\right)^2$. Without loss of generality, consider that on equilibrium worker with $m > \bar{m}$ work on H firms (L firms for $m < \bar{m}$). Consider a one-shot deviation from a worker in a high type firm that switches to a low firm at time $\hat{t}$ for dt and then goes back to equilibrium behavior. In this case, the value function is defined as

$$\tilde{V}_L\left(m,\hat{t}\right) = w_L^{a^\star}\left(m,\hat{t}\right)dt + e^{-rdt}E\left[V_H\left(m+dm,\hat{t}+dt\right)\right]$$

where $dm = \frac{A_y\gamma_t}{\sigma_{\xi_L}}dZ_t^{a^\star}$. [8] Apply Ito's Lemma and get,

$$\tilde{V}_L\left(m,\hat{t}\right) = w_L^{a^\star}\left(m\right)dt + e^{-rdt}\left[V_H\left(m,\hat{t}\right) + \Sigma_{t,L}V_H''\left(m,\hat{t}\right)dt + \frac{\partial V_H\left(m,\hat{t}\right)}{\partial t}dt\right]$$

Define $\frac{\partial a_{\hat{t}}^{H\to L}}{\partial t}$ as the derivative of $a^\star$ for a worker that switches from H to L at time $\hat{t}$, and define $\frac{\partial a_{\hat{t}}^{H}}{\partial t}$ as the derivative of $a^\star$ for a worker that, at time $\hat{t}$ still works for H and has not switched firms. With this notation we can obtain the expected gains from deviating as follows,

$$\lim_{dt\to 0}\frac{\tilde{V}_L\left(m,\hat{t}\right) - V_H\left(m,\hat{t}\right)}{dt} =$$

[7] As in Holmstrom's classical model of career concerns, the equilibrium provision of effort relates directly to the firm's certainty about a worker's type.

[8] I assume that during the deviation from equilibrium firms still perfectly anticipate the equilibrium effort level of the worker. This is common procedure, as in Cisternas and Holmsotrom.

$$w_L^{a^\star}\left(m,\hat{t}\right)-w_H^{a^\star}\left(m,\hat{t}\right)+\left(\Sigma_L-\Sigma_H\right)V_H''\left(m,\hat{t}\right)+\frac{\partial V_H\left(\bar{m},\hat{t}\right)}{\partial a_t}\left(\frac{\partial a_{\hat{t}}^{H\to L}}{\partial t}-\frac{\partial a_{\hat{t}}^{H}}{\partial t}\right)$$

This must be less than zero for any $m>\bar{m}$. In particular we have for $m\to\bar{m}$, at $\hat{t}$, that,

$$w_L^{a^\star}\left(\bar{m},\hat{t}\right)+\Sigma_{L,\hat{t}}V_L''\left(\bar{m},\hat{t}\right)+\frac{\partial V_L\left(\bar{m},\hat{t}\right)}{\partial t}$$
$$-w_H^{a^\star}\left(\bar{m}\right)-\Sigma_{H,\hat{t}}V_H''\left(\bar{m}\right)-\frac{\partial V_H\left(\bar{m},\hat{t}\right)}{\partial t}+\Sigma_{L,\hat{t}}\left(V_H''\left(\bar{m}\right)-V_L''\left(\bar{m}\right)\right)$$
$$+\frac{\partial V_H\left(\bar{m},\hat{t}\right)}{\partial a_t}\frac{\partial a_{\hat{t}}^{H\to L}}{\partial t}-\frac{\partial V_L\left(\bar{m},\hat{t}\right)}{\partial a_t}\frac{\partial a_{\hat{t}}^{L}}{\partial t}\leq 0$$

by value matching at $\bar{m}$ and $\hat{t}$ this implies that,

$$\Sigma_{L,\hat{t}}\left(V_H''\left(\bar{m}\right)-V_L''\left(\bar{m}\right)\right)+\frac{\partial V_H\left(\bar{m},\hat{t}\right)}{\partial a_t}\frac{\partial a_{\hat{t}}^{H\to L}}{\partial t}-\frac{\partial V_L\left(\bar{m},\hat{t}\right)}{\partial a_t}\frac{\partial a_{\hat{t}}^{L}}{\partial t}\leq 0$$

Now, from Lemma 3.3.2 we know that $\frac{\partial V_H\left(\bar{m},\hat{t}\right)}{\partial a_t}>\frac{\partial V_L\left(\bar{m},\hat{t}\right)}{\partial a_t}>0$ and that $\frac{\partial a_{\hat{t}}^{H\to L}}{\partial t}<\frac{\partial a_{\hat{t}}^{L}}{\partial t}<0$. This implies that indeed it is possible for $V_H''\left(\bar{m}\right)\geq V_L''\left(\bar{m}\right)$. Similarly we can consider a one shot deviation from a $m<\bar{m}$ worker moving to a high type firm for dt before switching back to equilibrium behavior. A similar argument allows us to conclude that in this scenario delivers the following condition,

$$\Sigma_{H,\hat{t}}\left(V_L''\left(\bar{m}\right)-V_H''\left(\bar{m}\right)\right)+\frac{\partial V_L\left(\bar{m},\hat{t}\right)}{\partial a_t}\frac{\partial a_{\hat{t}}^{L\to H}}{\partial t}-\frac{\partial V_H\left(\bar{m},\hat{t}\right)}{\partial a_t}\frac{\partial a_{\hat{t}}^{H}}{\partial t}\leq 0$$

Again, from Lemma 3.3.2 we know that $\frac{\partial V_H\left(\bar{m},\hat{t}\right)}{\partial a_t}>\frac{\partial V_L\left(\bar{m},\hat{t}\right)}{\partial a_t}>0$ and that either $\frac{\partial a_{\hat{t}}^{L\to H}}{\partial t}>0>\frac{\partial a_{\hat{t}}^{H}}{\partial t}$ or $\frac{\partial a_{\hat{t}}^{L\to H}}{\partial t}>\frac{\partial a_{\hat{t}}^{H}}{\partial t}>0$. This implies that $V_H''\left(\bar{m}\right)\geq V_L''\left(\bar{m}\right)$. Together these two conditions imply that in order to prevent any one-shot deviation from equilibrium behavior it must be the case that $V_H''\left(\bar{m}\right)\geq V_L''\left(\bar{m}\right)$. The second case follows similarly. □

The no-deviation condition in this environment looks very different from the stationary environment of the previous section. Indeed, the evolution of equilibrium effort now matters for payoffs and drives a wedge between the option value of learning workers perceive across jobs at the cutoff. As shown in Figure 3.2.1, the particular slopes of $a^\star$ upon switching will depend on when this switching takes place, and the type of switch in question. The result above states that we can nonetheless obtain a general pattern that deters deviations and thus pinpoints the relative curvatures of the value function in equilibrium. In particular, to deter possible deviations, the option value of learning must compensate workers from entering firms where learning is quicker and additional rents can be extracted. This result requires certain properties of the equilibrium effort schedule. In particular, it must be the case that effort provision in equilibrium drops at a speed proportional to the drop in posterior variance. This is what allows us to pinpoint the direction in which the value functions must differ in order to deter deviations from the equilibrium threshold strategy.

The next natural step is to consider whether PAM can still be expected in this environment. While I have not yet been able to prove this analytically, a number of informal arguments can be made. First of all, it seems unlikely that PAM can be defined as in the stationary environment of the previous section. The direct impact of γ_t on payoffs implies that two workers who reach a candidate threshold $\bar{m}$ at two different times t_1 and t_2 expect different future payoff streams. In this context it seems unreasonable to expect a unique threshold $\bar{m}$ that orders workers across firms independent of the time $\hat{t}$ at which this threshold is reached. In any case, Lemma 3.3.5 can characterize the no-deviation condition of workers independent of the time at which they contemplate deviating from their prescribed strategies. This is quite remarkable considering the above arguments, and it provides some evidence that even in this unstable environment, certain orderings can be made to pinpoint equilibrium allocations. However, even with a time-independent no-deviation condition, it seems intuitive that we cannot rule out the existence of multiple cutoffs (as in the proof of Theorem 3.2.7) since the optimal effort $a^\star$ need not be equal across the two cutoffs. This intuition needs to be formalized into a general result of sorting in unstable environments that I leave for future research.

3.4 Conclusion

This paper considers a model of turnover in the labor market where workers' decisions to switch jobs are subject to the learning experience of each match. The main objective is to determine whether equilibrium strategies predict sorting of workers across firms under complementarities in production. Building on previous work by Eeckhout and Weng (2012), the current framework considers that learning occurs in unstable environments, where the unobserved skill set of workers evolves randomly over time. I show that modeling the underlying state as a diffusion process has important implications for the type of learning experiences workers can expect upon switching firms. In particular, I show that episodes of *unlearning,* in which the strength of beliefs decays, is now possible. As a result, the period at which workers switch becomes just as relevant as the beliefs about their underlying productivity. I show that under certain conditions (risk neutrality and skills evolving as a martingale), these learning experiences are a second-order effect that do not effect future expected payoffs, and PAM can be sustained as in Eeckhout and Weng (2009). The argument requires that posterior beliefs have infinite support, so that the continuity of sample paths allows us to pin point the value function at the extremes. I also show that relaxing these conditions can generates equilibria with no unique sorting result.

I then extend the basic setup to allow workers to exert a level of effort that is unobservable to firms and which can affect the output process from which the market forms beliefs about the worker's type. These "signal-jamming" incentives generate spot wages that depend on equilibrium effort provisions directly, so that payoffs become time dependent. As a result, the no-deviation condition in Eeckhout and Weng (2015) looks very different and PAM breaks down. As it pertains to previous results in the literature, it is interesting to note that the career concerns version of the model still sustains bayesian learning processes. Payoffs are non-stationary because effort not only affects the value of learning, it also directly affects wages.

These results open the door to a number of additional queries on the robustness of PAM as an equilibrium configuration in non-stationary environments. For instance, it would be worthwhile to explore other assumptions on the behavior of the underlying skills process. A particularly relevant approach would allow for the possibility that effort directly affects the skills process, instead of output. These type of human capital accumulation stories are analyzed in the work of Cisternas (2012) and they represent very different learning environments from the career concerns motives studied here. In particular, if the skills process is unobservable, the market cannot fully anticipate the equilibrium investments, and learning is no longer symmetric across market participants. It would be interesting to know the equilibrium implications of this environment. Eeckhout and Weng (2009) explore a human capital accumulation story, but these fluctuations are exogenous and are not strategically modeled as an investment decision. Secondly, While PAM obtains in the case with no career concerns, the learning process is quite different from the work of Eeckhout and Weng (2009) so we should therefore expect that the equilibrium distribution should look quite differently. Indeed we should proceed and characterize the stationary distribution in equilibrium. All this is left for future research.

Bibliography

Ambrus, A., Chandrasekhar, A., and Elliot, M. (2015). Social investments, informal risk sharing, and inequality. Working Paper 20669, NBER.

Ambrus, A., Mobius, M., and Szeidl, A. (2014). Consumption risk-sharing in social networks. *American Economic Review*, 104(1):149–182.

Anderson, A. and Smith, L. (2010). Dynamic matching and evolving reputations. *The Review of Economic Studies*, 77(1):3–29.

Angeletos, G., Hellwig, C., and Pavan, A. (2006). Signaling in a global game: Coordination and policy traps. *Journal of Political Economy*, 114(3):452–484.

Angeletos, G., Hellwig, C., and Pavan, A. (2007). Dynamic global games of regime change: Learning, multiplicity, and the timing of attacks. *Econometrica*, 75(3):711–756.

Angeletos, G. and Werning, I. (2006). Crises and prices: Information aggregation, multiplicity, and volatility. *American Economic Review*, 96(5):1720–1736.

Angelucci, M., de Giorgi, G., and Rasul, I. (2015). Resource pooling within family networks: Insurance and investment. Working Paper 2649, eSocialSciences.

Attanasio, O., Barr, A., Camilo Cardenas, J., Genicot, G., and Meghir, C. (2012). Consumption risk-sharing in social networks. *American Economic Journal: Applied Economics*, 4(2):134–167.

Attanasio, O., Meghir, C., and Mommaerts, C. (2015). Insurance in extended family networks. Working Paper 21059, NBER.

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.

Baker, G. Gibbs, M. and Holmstrom, B. (1994). The wage policy of a firm. *The Quarterly Journal of Economics*, 109(4):881–919.

Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). Who's who in networks. wanted: The key player. *Econometrica*, 74(5):1403–1417.

Banerjee, A., Chandrasekhar, A., Duflo, E., and Jackson, M. (2014). Insurance in extended family networks. Working paper, MIT.

Becker, G. (1973). A theory of marriage: Part i. *Journal of Political Economy*, 81(4):813–846.

Billand, P., Bravard, C., and Sarangi, S. (2015). Mutual insurance networks and unequal resource sharing in communities. Working paper, DIW Berlin.

Bloch, F. and Dutta, B. (2009). Communication networks with endogenous link strength. *Games and Economic Behavior*, 66(1):39–56.

Bloch, F., Genicot, G., and Ray, D. (2008). Informal insurance in social networks. *Journal of Economic Theory*, 143(1):36–58.

Blum, J., Hanson, D., and Koopmans, L. (1963). On the strong law of large numbers for a class of stochastic processes. *Z. Wahrsch. verw. Gebiete*, 2:1–11.

Blundell, R., Pistaferri, L., and Preston, I. (2008). Consumption inequality and partial insurance. *The American Economic Review*, 98(5):1887–1921.

Bolton, P. and Harris, C. (1999). Strategic experimentation. *Econometrica*, 67(2):349–374.

Bourlès, R., Bramoullé, Y., and Perez-Richet, E. (2015). Altruism in networks. Working paper, Aix Marseille.

Bramoullé, Y. and Kranton, R. (2007a). Public goods in networks. *Journal of Economic Theory*, 135(1):478–494.

Bramoullé, Y. and Kranton, R. (2007b). Risk-sharing networks. *Journal of Economic Behavior and Organization*, 64(3-4):275–294.

Bueno de Mesquita, E. (2010). Regime change and revolutionary entrepeneurs. *Amer- ican Political Science Review*, 104(3):446–466.

C., K. and Townsend, R. (2012). Kinship and financial networks, formal financialaccess, and risk reduction. *American Economic Review: Papers and Proceedings*, 102(3):289–293.

Calvó-Armengol, A. and de Martí Beltran, J. (2007). Communication networks: Knowledge and decisions. *American Economic Review, Papers and Proceedings,*, 97(2):86–91.

Calvó-Armengol, A. and de Martí Beltran, J. (2009). Information gathering in organizations: Equilibrium, welfare, and optimal network structure. *Journal of the European Economic Association*, 7(1):116–161.

Carlsson, H. and van Damme, E. (1993). Global games and equilibrium selection. *Econometrica*, 61(5):989–1018.

Chade, H. and Eeckhout, J. (2013). Stochastic sorting. Working paper, Society for Economic Dynamics.

Chanrasekhar, A. G., Kinnan, C., and Larreguy, H. (2015). Social networks as contract enforcement: Evidence from a lab experiment in the field. Working Paper 20259, NBER.

Chwe, M. (1999). Structure and strategy in collective action. *American Journal of Sociology*, 105:128–156.

Chwe, M. (2000). Communication and coordination in social networks. *Review of Economic Studies*, 67:1–16.

Cisterns, G. (2012). Shock persistence, endogenous skills and career concerns. Working paper, MIT.

Cochrane, J. (1991). A simple test of consumption insurance. *Journal of Political Economy*, 99(5):957–976.

Cripps, M. W., Ely, J., Mailath, G., and Samuelson, L. (2008). Common learning. *Econometrica*, 76(4):909–933.

Dahleh, M., Tahbaz-Salehi, A., Tsitsiklis, J., and Zoumpoulis, S. (2012). Structure and strategy in collective action. Working paper, MIT Lab.

Dewatripont, M., Jewitt, I., and Tirole, J. (1999). The economics of career concerns, part i: Comparing information structures. *Review of Economic Studies*, 66(1):199–217.

DeWeerdt, J. and Dercon, S. (2006). Risk-sharing networks and insurance against illness. *Journal of Development Economics*, 81(2):337–356.

Dixit, A. (1993). *The Art of Smooth Pasting*. Routledge, New York.

Edmond, C. (2013). Information manipulation, coordination, and regime change. *Review of Economic Studies*, 80(4):1422–1458.

Eeckhout, J. and Jovanovic, B. (2011). Occupational choice and development. *Journal of Economic Theory*, 147(2):657–683.

Eeckhout, J. and Weng, X. (2009). Assortative learning. Working paper.

Eeckhout, J. and Weng, X. (2015). Common value experimentation. *Journal of Economic Theory*, 160:317–339.

Fafchamps, M. and Gubert, F. (2007). The formation of risk sharing networks. *Journal of Development Economics*, 83(2):326–350.

Fafchamps, M. and Lund, S. (2003). Risk-sharing networks in rural philippines. *Journal of Development Economics*, 71(2):261–287.

Faingold, E. and Sannikov, Y. (2007). Reputation effects and equilibrium degeneracy in continuous-time games. Discussion Paper 1624, Cowles Foundation.

Farber, H. and Gibbons, R. (1996). Learning and wage dynamics. *The Quarterly Journal of Economics*, 111(4):1007–1047.

Foster, A. and Rosenzweig, M. (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.

Galeotti, A. and Goyal, S. (2010). The law of the few. *American Economic Review*, 100(4):1468–1492.

Gibbons, R. and Murphy, K. (1992). Optimal incentive contracts in the presence of career concerns: Theory and evidence. *The Journal of Political Economy*, 100(3):468–505.

Guimaraes, B. and Morris, S. (2007). The law of the few. *Journal of Monetary Economics*, 54(8):2205–2230.

Gurven, M., Kaplan, H., and Zelada Supa, A. (2007). Mortality experience of tsimane amerindians of bolivia: Regional variation and temporal trends. *American Journal of Human Biology*, 19:376–398.

Hagenbach, J. and Koessler, F. (2010). Strategic communication networks. *Review of Economic Studies*, 77(3):1072–1099.

Hall, R. (1978). Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence. *Journal of Political Economy*, 86(6):971–987.

Harris, M. and Holmstrom, B. (1982). A theory of wage dynamics. *The Review of Economic Studies*, 49(3):315–333.

Hassanpour, N. (2010). Dynamic models of mobilization in political networks. Working paper, Duke Univeristy.

Hayashi, F., Altonji, J., and Kotlikoff, L. (1996). Risk-sharing between and within families. *Econometrica*, 64(2):261–294.

Holmstrom, B. (1979). Moral hazard and observability. *The Bell Journal of Economics*, 10(2):74–91.

Holmstrom, B. (1999). Managerial incentive problems: A dynamic perspective. *The Review of Economic Studies*, 66(1):169–182.

Hooper, P. (2011). The structure of energy production and redistribution among tsimane' forager-horticulturalists. Phd dissertation, University of New Mexico.

Hörner, J. and Samuelson, L. (2013). Incentives for experimenting agents. *The RAND Journal of Economics*, 44(4):632–663.

Jackson, M., Rodriguez-Barraquer, T., and Tan, X. (2012). Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*, 102(5):1857–1897.

Jovanovic, B. (1979). Job matching and the theory of turnover. *The Journal of Political Economy*, 87(5):972–990.

Karlan, D., Mobius, M., Rosenblat, T., and Szeidl, A. (2009). Trust and social collateral. *Quarterly Journal of Economics*, 124(4):1307–1361.

Keller, G., Rady, S., and Cripps, M. (2005). Strategic experimentation with exponential bandits. *Econometrica*, 73(1):39–68.

Kinnan, C. (2014). Distinguishing barriers to insurance in thai villages. Working paper, Northwestern University.

Leicht, E., Holme, P., and Newman, M. (2006). Vertex similarity in networks. *Physical Review*, 73(2):026120.

Li, B. and Zhang, K. (2010). Strong law of large numbers for *- mixing sequence. *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering*, 4(8):1098–1100.

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

Ligon, E., Thomas, J., and Worrall, T. (2000). Mutual insurance, individual savings and limited commitment. *Review of Economic Dynamics*, 3(2):216–246.

Liptser, R. and Shiryaev, A. (2001). *Statistics of Random Processes I and II*. Springer-Verlag, New York.

Mace, B. (1991). Full insurance in the presence of aggregate uncertainty. *Journal of Political Economy*, 99(5):928–956.

Martin, M., Lassek, W., Gaulin, S., Evans, R., Woo, J., Geraghty, S., Davidson, B., Morrow, A., Kaplan, H., and Gurven, M. (2012). Fatty acid composition in the mature milk of bolivian forager-horticulturalists: Controlled comparisons with a us sample. *Maternal and Child Nutrition*, 8(3):404–418.

Mazzocco, M. and Saini, S. (2012). Testing efficient risk sharing with heterogeneous risk preferences. *American Economic Review*, 102(1):428–68.

Morris, S. and Shin, H. S. (2003). Global games: Theory and applications. In Dewatripont, M., Hansen, L., and Turnovsky, S., editors, *Advances in Economics and Econometrics*, chapter 3, pages 56–114. Cambridge University Press.

Moscarini, G. (2005). Job matching and the wage distribution. *Econometrica*, 73(2):481–516.

Munshi, K. and Rosenzweig, M. (2009). Why is mobility in india so low? social insurance, inequality, and growth. Working Paper 14850, NBER.

Munshi, K. and Rosenzweig, M. (2016). Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, 106(1):46–98.

Nakata, T. and Tonetti, C. (2015). Small sample properties of bayesian estimators of labor income processes. *Journal of Applied Economics*, 18(1):121–148.

Rochet, J.-C. and Vives, X. (2004). Coordination failures and the lender of last resort: Was bagehot right after all? *Journal of the European Economic Association*, 2(6):1116–1147.

Sákovics, J. and Steiner, J. (2012). Who matters in coordination problems? *American Economic Review*, 102(7):3439–3461.

Sannikov, Y. (2007). Games with imperfectly observable actions in continuous time. *Econometrica*, 75(5):1285–1329.

Sannikov, Y. (2008). A continuous-time version of the principal-agent model. *Review of Economic Studies*, 75(3):957–984.

Schulhofer-Wohl, S. (2011). Heterogeneity and tests of risk sharing. *Journal of Political Economy*, 119(5):925–958.

Stokey, N. (2009). *The Economics of Inaction: Stochastic Control Models with Fixed Costs*. Princeton University Press, New Jersey.

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. *Review of Economic Studies*, 61(3):495–526.

Van Zandt, T. and Vives, X. (2007). Monotone equilibria in bayesian games of strategic complementarities. *Journal of Economic Theory*, 134(1):339–360.

Vives, X. (2005). Complementarities and games: New developments. *Journal of Economic Literature*, 43(2):437–479.