

NBER WORKING PAPER SERIES

CONTINGENT LINEAR FINANCIAL NETWORKS

Bomin Jiang
Roberto Rigobon
Munther A. Dahleh

Working Paper 26814
<http://www.nber.org/papers/w26814>

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue
Cambridge, MA 02138
March 2020, Revised November 2020

We thank Mila Getmansky, Andrew Lo, Loriana Pelizzon, and the participants of the USC Dornsife Institute for New Economic Thinking seminar for their comments. For communication contact us at bominj@mit.edu or rigobon@mit.edu. All remaining errors are ours. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by Bomin Jiang, Roberto Rigobon, and Munther A. Dahleh. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Contingent Linear Financial Networks
Bomin Jiang, Roberto Rigobon, and Munther A. Dahleh
NBER Working Paper No. 26814
March 2020, Revised November 2020
JEL No. E0,E44,G1,G21

ABSTRACT

In this paper, we develop a methodology to estimate hidden linear networks when only an aggregate outcome is observed. The aggregate observable variable is a linear mixture of the different networks and it is assumed that each network corresponds to the transmission mechanism of different shocks. We implement the methodology to estimate financial networks among US financial institutions. Credit Default Swap rates are the observable variable and we show that more than one network is needed to understand the dynamic behavior exhibited in the data.

Bomin Jiang
Institute for Data, Systems, and Society
Massachusetts Institute of Technology
77 Massachusetts Ave, 32D-758
Cambridge, MA 02139
bominj@mit.edu

Munther A. Dahleh
Dept. of Electrical Engineering
and Computer Science
Massachusetts Institute of Technology
77 Massachusetts Ave, 32D-616
Cambridge, MA 02139
alinam@mit.edu

Roberto Rigobon
MIT Sloan School of Management
100 Main Street, E62-516
Cambridge, MA 02142
and NBER
rigobon@mit.edu

Contingent Linear Financial Networks

Abstract

In this paper, we develop a methodology to estimate hidden linear networks when only an aggregate outcome is observed. The aggregate observable variable is a linear mixture of the different networks and it is assumed that each network corresponds to the transmission mechanism of different shocks. We implement the methodology to estimate financial networks among US financial institutions. Credit Default Swap rates are the observable variable and we show that more than one network is needed to understand the dynamic behavior exhibited in the data.

Keywords: Financial Network, Identification through Heteroskedasticity, Mixture Model, EM Algorithm

JEL Codes: C13, C61, E00, G21, G28

1 Introduction

The 2008 world financial crisis started in what was supposed to be a relatively small, and segmented market. At that time the outstanding value of Sub-Prime debt was a tad below a trillion dollars, while the market value of credits in both the formal and shadow US financial sectors was higher than 23 trillion dollars. So, a default on 4 percent of an isolated market (the high-risk mortgages) was expected to have an equal minuscule effect. Not surprisingly, the Obama administration only requested 800+ billion dollars for the rescue package. It is obvious, today, that the systemic consequences of the sub-prime crisis were unexpectedly larger. The shock propagated to other financial sectors and countries. The world total losses reached several trillion. This has spurred research on systemic risk, and in particular on the

estimation of the underlying financial networks governing the propagation of shocks - or as it is sometimes known as financial contagion.

The estimation of systemic risk has been closely linked to the estimation of financial networks. Banks can be interconnected through many different channels. One type of link is related to the exposure banks have to similar microeconomic or industry shocks. For example, two bank's balance sheets can be interconnected because both are lending to the same firm or sector and it suffers a shock. The second type of channel is related to interbank contracts. Two banks can be interrelated because one bank lends to the other, or they hold each other's liabilities. Therefore, a deterioration in the balance sheet of the borrowing bank affects the quality of the assets of the lending bank. Finally, banks can be exposed to similar macroeconomic shocks such as exchange rate, inflation rate, interest rates, economic activity, real estate, etc. In sum, there are many possible ways in which banks are linked to each other.¹ The objective of the empirical literature has been to estimate the "average" linkage among banks, and to determine the contribution each bank has to the overall risk in the economy — a measure of systemic risk.

Observing and understanding the contracts underlying the links between banks is sometimes challenging, however. Direct lending from one bank to the other of course is simple to document. The relationship across other contracts — sometimes complex and through indirect channels — implies that the information required to estimate the financial networks is daunting. First, some trivial aspects, such as the currency of denomination of a contract, is unclear. For example, a financial contract can specify all its flows and payments in US

¹See Allen and Gale (2000), Freixas et al. (2000) for earlier contributions, and Acemoglu et al. (2015), Allen et al. (2012), Caballero and Simsek (2013), Cabrales et al. (2014), Elliott et al. (2014), Gai and Kapadia (2010), and Gai et al. (2011) for recent theoretical papers. Empirical papers that estimate financial networks include Billio et al. (2010), Merton et al. (2013), Adrian and Brunnermeier (2016), and Girardi and Ergün (2013).

Dollars, but its return can be perfectly correlated with the Euro-Dollar rate. So, even though this is a contract denominated in the local currency, it is actually a de-facto foreign exchange contract. Contracts such as default swaps with tranches, or contracts that are indexed to macroeconomic variables are complex and fully understanding the type of connection that they generate across two financial institutions might be impossible. Second, most of the times the contagion path is outside the regulatory or country jurisprudence. For example, interconnections can occur through firms that might not even reside in the country of analysis. For instance, assume a bank lends to Intel, who sells to a South Korean firm, that manufactures a monitor sold to a California firm, who gets a loan from a different bank. In this setting, the two banks are related through the South Korean firm. However, the requirements on reporting to be able to uncover such relationship are impossible — i.e. the US regulator can't force the South Korean company to reveal its clients and suppliers. The complexity, number, and variety of contracts that can be used among financial institutions have several implications in the estimation of networks. In principle, if every contract is fully disclosed and documented the relation between banks could be described precisely. In practice, however, the granularity and detail of the required information are unfeasible, even to regulators and central banks. For this reason, the financial network literature tends to simplify the information demands by approximating the underlying network using aggregate data.

In the literature, two main approaches have been taken. The first one uses information-theoretic principles to fill in the blanks of pairwise data using aggregate data.² For example, one can read the total inter-bank debt from Form 10-Ks. After that, the worst-case scenario can be evaluated by inferring the pair-wise debt structure using a maximum entropy principle. Those methods, while used in some stress tests, only give a rough estimate of the debt

²See Upper (2011); Elsinger et al. (2013) for maximum entropy and Anand et al. (2015) for minimal density

structures and do not fully utilize the information from time-varying data.

The second approach uses the correlation of prices of financial contracts (Onnela et al., 2004) to construct interbank financial networks. The estimates based on correlations can be biased in the presence of heteroskedastic shocks.³ Furthermore, both of these approaches, do not take into account the possibility that the network between banks is asymmetric, non-linear, and that there might be more than one network at play.

First, for any given network, it is likely that the transmission of shocks between banks is asymmetric and non-linear. In other words, the impact on the balance sheet of two banks depends on the size of the shock — a small shock might imply a small propagation, but a large shock might produce a large propagation. The reason is that larger shocks might trigger clauses in the financial contracts that smaller shocks do not cause.⁴ More importantly, regulation of financial institutions is inherently non-linear — bankruptcy for instance, where the bank is allowed to operate freely within a capital requirement range, but quite restrictively outside of it. This implies that the nature of the network might be different depending on the size of the shock. In fact, the average correlation might severely underestimate the transmission of shocks when a larger shock hits the system.

Second, and equally importantly, it is likely that there is more than a single network describing the system. Each type of shock is likely to be transmitted through different contracts and the propagation would be conditional on the shock. For example, imagine that there are three banks, and two are negatively exposed to real estate movements while the third one is positively exposed to real estate, but all three banks are negatively exposed to interest rate increases. If in the economy there is an increase in interest rates, all three banks move together, but if there is a decline in the price of real estate, only two are affected. The correlation structure, and therefore the network, changes with the shock.

³See Forbes and Rigobon (2002) for the bias in assessing financial contagion when using correlations.

⁴As in Acemoglu et al. (2015)

Third, both of the previous mechanisms imply that the propagation of shocks between banks is asymmetric. For example, assume there are two banks, a large national bank (such as CitiBank) and a small regional bank such as (Watertown Savings Bank). A shock that decreases the value of the large bank's assets by 10 percent might have an impact on the small one that is very different from a shock that reduces the small bank's assets by the same proportion. The shock to the large bank might be more meaningful to the small one than vice-versa. When the measurement of systemic risk uses either stock market prices, or credit default swaps, these interactions are contemporaneous. Therefore, the estimation of even a single network is complicated by endogenous variable biases. A correlation is incapable of capturing these differences.

Consequently, the estimation of systemic risk in the financial system is complicated by asymmetries, contemporaneous endogenous biases, and parameter instability. Our methodology makes several assumptions and simplifications to be able to solve the identification problem. First, the underlying networks are unobservable, but the outcome of their interactions is observable. In other words, the changes in bank balance sheets and financial conditions are the outcomes that can be observed and it is the result of many interactions acting jointly. Therefore, the estimation of the contingent financial networks is not only an identification problem but also a computational challenge.

The second assumption is the realization that non-linearity and multi-networks are empirically intertwined. In fact, the non-linearity is tantamount to the network being contingent to the size of shock: larger shocks imply different transmission mechanisms than smaller ones. There are a number of papers modeling financial networks demonstrating the cause of potential shifts in the transmission mechanism. Some insist on the change of risk transmission mechanism is induced by monetary policies (Altunbas et al., 2010). That is because monetary policy has a significant effect on company financial decisions, which in turn could change the structure of financial contracts and result in discontinuities in risk transmission

mechanisms. Others believe market volatility could result in a phase transition of non-linear financial networks, which incurs new risk transmissions mechanism (Acemoglu et al., 2015). Multiple networks can be a piecewise approximation of the non-linearity underlying these papers. Therefore, we make two important assumptions about the structure of the network: First, we estimate a collection of linear networks. In other words, in the case of non-linearity, we approximate a non-linear network as a set of linear ones. In the case of different networks for each type of shock, we approximate each one as a linear network. Second, we assume that the overall outcome is a mixture of multiple underlying unobservable networks and the shocks.

Our model specification is flexible enough that allows for two possible interpretations. The first interpretation is one in which idiosyncratic shocks hit each node, and propagate through the different networks. This interpretation is akin to a structure where the shocks are independent and specific to each firm, and the propagation occurs throughout the interrelationships among them. The second interpretation assumes that all the shocks that affect the network are aggregate shocks, but the effect on each firm is idiosyncratic. In other words, firms are exposed to aggregate shocks in different degrees. Once the aggregate shock occurs, all firms are affected. In this interpretation, we still allow the network to play a role in the transmission of the shocks. There is an initial correlation, however, across firms that comes from the exposure to the aggregate shock. This interpretation captures macroeconomic circumstances where firms are exposed to interest rates, exchange rates, real estate, inflation, etc. Their exposure to each of the shocks might be different. Once the shock occurs, it will be propagated by a different network. The underlying assumption is that the linkages between firms once an inflation shock occurs, might be different from the linkages that exist once a real estate shock occurs. Our flexible interpretation, however, cannot distinguish between these two interpretations. We can estimate the underlying structure, but which interpretation is the relevant one depends on the application. In our model, they are

isomorphic.

The identification problem is solved using the heteroskedasticity observed in the data. The estimation imposes a maximum number of parameters that can be recovered from the data; therefore, we limit the number of networks that can be estimated. This implies that it is likely that in many circumstances we are still not capturing the full complexity in the financial system. In this case, however, our estimation procedure captures the most “dissimilar” networks describing the data. Furthermore, we can evaluate and discuss the informational value gained by allowing an additional network. We implement the estimation procedure through an intuitive EM algorithm. We derive identification assumptions for each one of the network interpretations, either purely idiosyncratic or purely aggregate shocks.

The main contribution of the paper is the estimation of a mixture model of multiple endogenous asymmetric linear networks.⁵ The intuition behind the estimation comes from the fact that in a linear network the correlation structure across banks is the result of a linear mixture of the covariance implied by each network - where the weights are the relative variances of each shock. When the variance of those shocks changes through the sample, the weights change, leading to variation in the covariances. From that variation of the second moment, the underlying structure of the networks can be recovered. Therefore, the identification requires finding periods where the shocks experience different volatilities. There are two ways to do so: first, statistical identification where the changes in the variance in the observed data determine the regimes. This procedure is very similar to the one developed by Sentana and Fiorentini (2001). The advantage of this methodology is that it allows estimating the asymmetry of large versus small shocks - where the statistical identification truly captures the differences in the propagation mechanisms. The second alternative is to find periods where macroeconomic shocks experience different variances and use those as heteroskedastic sources. The advantage of this procedure is that the identification is closer

⁵The procedure follows Rigobon (2003).

to economic events rather than statistical events. So, if the relationships are contingent to the shocks this procedure allows us to estimate the networks that closely capture such linkages.

We apply our methodology to estimate the US financial network among the 10 largest financial institutions in the country. The data collected are their credit default swaps (CDSs). In this paper, a first pass to the problem, we concentrate on estimating a small network.

The CDSs have been used widely in the literature studying risk contagion. Most notably, the CDSs of sovereign debt has been examined by various papers to understand international risk propagation. E.g., Kalbaska and Gałkowski (2012) investigates the eurozone contagion via a regression on CDS spread changes. In general, if a risk contagion mechanism exists among certain countries, their CDS spread will co-move. From the trend of CDS spread, the author reaches a conclusion that Sovereign debt risk is mainly limited to EU countries. Similarly, Caporin et al. (2018) also studies European debt crisis by examines the CDS spread. Through a Bayesian quantile regression that incorporates shock heteroskedasticity, they conclude that the increases in the correlation of CDS come from heteroskedasticity instead of structural changes of risk propagation mechanism.

The CDSs of banks have also been widely used in studying interbank risk propagation. As pointed out by Eichengreen et al. (2012), the CDS spreads of major banks co-move and reflects market economic prospects. Furthermore, during the Subprime Crisis and the following crisis of Lehman Brothers, the common factor is more dominant, i.e., the absorption ratio (as defined in (Kritzman et al., 2010)) is higher. The fact that major financial crises are reflected by CDS regime changes is the exact characteristic we want to study. In our contingent linear financial network model, a significant regime shift is required for unique identification. As shown in (Eichengreen et al., 2012), the common factor accounted for 62% of the variance of major bank CDSs before the 2007 breakout of Subprime Crisis and raised

to 77% during the crisis, signaling a major regime shift. Apart from stronger co-movements, we also want to discover whether the change of regime is a sole result of heteroskedasticity of shocks, or a result of both the heteroskedasticity of shocks and the structural change of risk propagation mechanisms.

We study 9 different macroeconomic variables that try to capture different aggregate shocks that typically afflict economies: (1) commodity price shocks (WTI Oil Prices), (2) Inflation (PriceStats daily inflation index), (3) Economic Activity (Non-farm payroll), (4) Stock Market Prices (S&P), (5) Risk (VIX), (6) Housing Prices (Case & Shiller index), (7) liquidity provision (short term interest rates), (8) exchange rates (nominal trade-weighted exchange rate), and (9) the yield curve (the difference between the long and short interest rates). Of course in a finite sample, it is not clear that all shocks occur - for instance high inflation has not happened in a long time in the US. This implies that not all networks can be identified.

In the case of a single network, the solution is established through GMM. On the other hand, in the case of multiple networks, the expectation-maximization (EM) algorithm, which is commonly used for estimating mixture models, is adopted. The EM algorithm alternates between an expectation (E) step, which updates the probability of a network dominating a regime, and a maximization (M) step, which estimate each network through heteroskedasticity based on the probability updated in the E step. A Wishart distribution is assumed for sample variance-covairance matrices so that the log-likelihood function used in the EM algorithm is well-defined.

The paper is organized as follows: In Section 2, a problem of the financial network is formulated and the problem of identification is discussed. The section starts with a well-studied two-bank case to provide some intuition of the identification problem, and then provides a proof of solution uniqueness in the single network case. It also discusses the identification

problem when the data generating process is explained by multiple networks. Section 3 discusses the case when a single network is estimated. We present both the statistical and the macroeconomic strategies of identification. Section 4 presents a test for the number of networks. In the 10 banks case, 3 networks is required to explain the data. The results of the estimated networks is then discussed in this section. Finally, Section 5 concludes.

2 Modeling Financial Networks

The estimation of financial networks is not new. A significant proportion of papers, however, concentrate on the estimation of symmetric single networks — non-directional, and non-contingent graphs. In this section, we present first, some preliminary evidence based on correlations. We then introduce a linear financial network model, and show: 1. how the network model can be uniquely identified with variance-covariance data; 2. together with heteroskedastic shocks, how it can explain the volatility in correlation models, and 3. how the heteroskedasticity of shocks can be identified through macroeconomic indices.

2.1 The Time-Varying Correlations

It is desired to find out how and why banks are related to each other, and which banks are “crucial” in the risk transmission mechanisms, and which ones are less relevant. One natural measure for those purposes is the correlation coefficients. In modern portfolio theory (MPT), correlation coefficients also play an essential role. MPT quantitatively formalizes the concept of diversification via the statistical notion of covariance, or correlation.

However, most of those applications of linear correlation models make an oversimplified assumption, that the correlation coefficients of any two given financial instruments are time-

invariant, or at least time-invariant in the period of analysis.⁶ In practice, there are numerous examples where such assumption does not apply. In this paper, we are concerned with the systemic financial risks which primarily propagate through large banks, so we use the Credit Default Swap (CDS) data of the top 10 US banks as an example.^{7 8}

⁶Although most of the literature estimates the strength of the network transmission by using simple correlations, there are notable exceptions that are worth highlighting. Adrian and Brunnermeier (2016) proposes a new measure of comovement called CoVar — which is defined as the value at risk conditional on the bank being in distress. Girardi and Ergün (2013) extend that measurement to expand the definition of distress. These types of measures are consistent with networks being contingent.

⁷It is important to highlight that CDS might exhibit excessive comovement due to the presence of a government guarantee. (Merton et al., 2013) uses a different approach to measure the credit risk of the banks. They use contingent claims analysis instead of CDS. The CDS are partially guaranteed by government policy — for instance, deposit insurance. Future research should evaluate the robustness of the results presented in this paper when different measures of financial performance are used.

⁸See Appendix for details of the CDS data used.

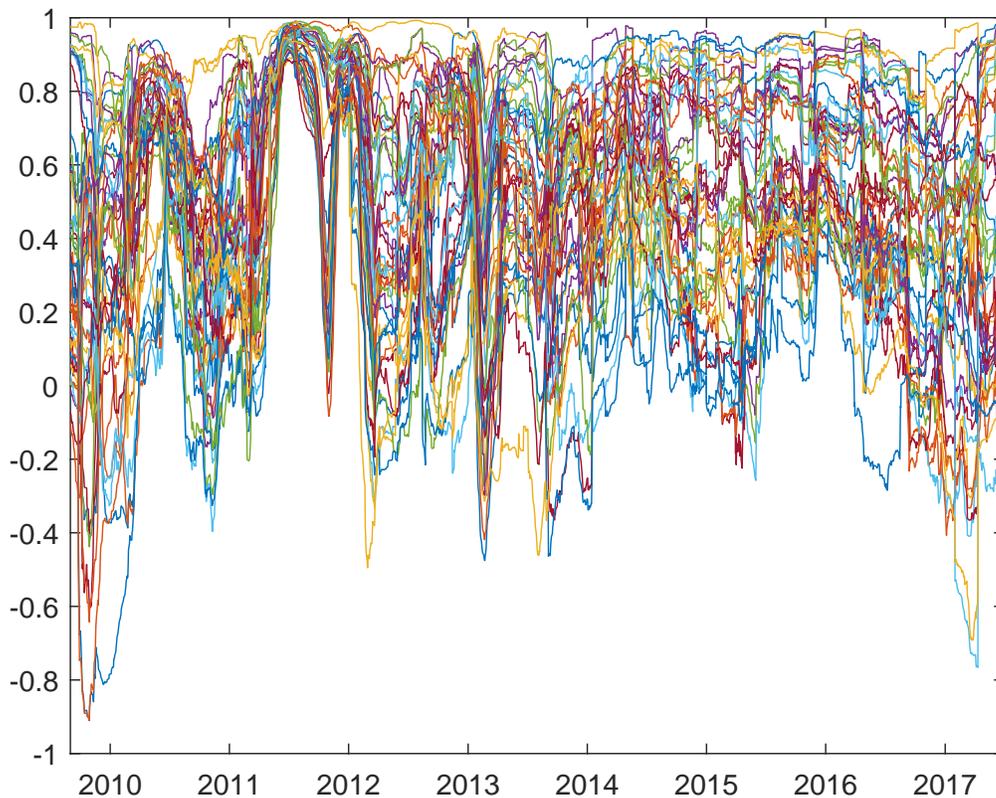


Figure 1: The 45 pairwise correlation coefficients of CDS of the top 10 US banks. The stock exchange id numbers of the 10 banks are JPM, BAC, WFC, C, GS, MS, COF, HSBC, AXP, and CSGN.

The change of volatility could be due to either time-varying linear regression coefficients (a.k.a. linear financial networks, as we will define later in the paper), or heteroskedastic shocks. Figure 1 shows the 45 pairwise correlation coefficients of CDS of the top 10 US banks calculated using a 200 days moving window. It is immediately notable that the correlation cannot be considered time-invariant. At an extreme, the correlation coefficient rises from -0.1 to 0.8 within 100 days. Up to this point, it is unclear whether the time-varying correlation is a

result of heteroskedasticity of shocks, or a result of the structural change of risk propagation mechanisms, or both. The two-banks illustrative example we give in Section 2.3.1 shows that all three cases are possible. To distinguish the three cases, we need to develop a linear network model.

2.2 A Linear Network Model

Suppose there are N financial institutions (indexed by $n = 1, 2, \dots, N$) in a contingent financial network with M possible networks, which denoted by the directed graphs $\mathcal{G}_m = \{\mathcal{V}, \mathcal{E}_m\}$ for $m \in \{1, 2, \dots, M\}$. \mathcal{V} is the set of common nodes in all the graphs and each node $v_n \in \mathcal{V}$ corresponds to a financial institution n . Furthermore, each edge $e_{ijm} \in \mathcal{E}_m$ denotes the risk spillover through m th network between node i and node j .

There are two possible — simple — assumptions on how to implement the contingent networks. One in which the shocks are idiosyncratic and hit each bank individually and then they are propagated in the network, or the second one where the shocks are hitting the system as a whole. The first assumption is one in which the systemic risk of an individual shock affects is determined by the propagation between one bank and the other, while in the second assumption the relative importance of the aggregate shocks is what makes them systemic.

We assume that the system is affected by shocks denoted as ϵ_t , a N -by-1 vector representing N shocks at time t . A simple framework when a market is hit by a shock as follows: Assume that at each time instance t , each node in the contingent financial network receives a shock. When a shock ϵ_t transmits through network m , it has an impact on financial institutions in the network, both directly and indirectly.

The direct impact is due to the fact that institution i has some direct exposure in this

market, and it is denoted by $\theta_m \epsilon_t$. Here θ_m is a N -by- N matrix, where each entry i, j represents the exposure of bank i to shock j in network m . We assume that θ_m are invertible. Note the structure of the exposure matrix θ_m reflects on the nature of the shock: idiosyncratic or aggregate. We will discuss those two cases in detail later this section.

Assume that the impact of shocks propagating through network m is not observable but the total risk measure, defined as the mixture of impact of different networks

$$X_t = \sum_{m=1}^M w_{mh} F_{m,t}$$

is observable, where $F_{m,t}$ is the impact of shocks propagating through network m , and w_{mt} is an indicator random variable. $w_{mt} = 1$ if network m dominates at time t , and $w_{mt} = 0$ otherwise. Let h indicate different heteroskydastic regimes. Assume that the indicator random variable w_{mh} is 1 with probability p_m . Note this p_m is a fixed prior distribution in the mixture model that does not depend on h .

Suppose the impact of shocks propagating through network m satisfies

$$F_{m,t} = f_m(F_{m,t}) + \theta_m \epsilon_t. \quad (1)$$

where $f_m : \mathbb{R}^N \rightarrow \mathbb{R}^N$ is a vector valued function capturing the risk propagation mechanism. A first order Taylor expansion gives

$$F_{m,t} = \alpha_m + \Xi_m F_{m,t} + \theta_m \epsilon_t. \quad (2)$$

In equation (2) we can see the direct and indirect effects of the shock — regardless if the shock is aggregate or idiosyncratic. The indirect impact of ϵ_t is through Ξ_m ; which represents the network on how financial institutions have exposure through the balance sheet of the other financial institutions. We define Ξ_m , $m \in \{1, 2, \dots, M\}$ as the weighted-directed adjacency matrix of network m . We assume there cannot be self-loops in the network hence all diagonal entries of Ξ_m are zero.

For the rest of this paper, we assume that our data is demeaned, hence

$$F_{m,t} = \Xi_m F_{m,t} + \theta_m \epsilon_t. \quad (3)$$

This is a decomposition of any shock to the system as the exogenous part ($\theta_m \epsilon_t$) and the indigenous network part (Ξ_m). In this model, the time subscript is the same on both sides of the equation, assuming we are at the equilibrium. This is a recursive assumption that was first posed by Christiano et al. (1999).

Note that rearranging (3) we obtain

$$F_{m,t} = [I - \Xi_m]^{-1} \theta_m \epsilon_t. \quad (4)$$

Hence we observe

$$X_t = \sum_{m=1}^M w_{mt} [I - \Xi_m]^{-1} \theta_m \epsilon_t, \quad (5)$$

This describes the observed variable (X_t) as a linear mixture of different unobservable networks (Ξ_m) and different shocks ($\theta_m \epsilon_t$).

This representation allows for two different interpretations. One is where all the shocks affect every bank individually and then the effect is transmitted through the networks, and θ_m is identity matrix, while the second one assumes the shocks to be entirely aggregate and each bank has a different exposure, and θ_m is unconstrained. Column j of θ_m represent the exposure of banks to shock j .

The estimation problem is to uncover the unobserved networks from the moments of the observed variables.

2.2.1 Idiosyncratic Shocks' Network

This subsection first discusses the case where a network is hit only by idiosyncratic shock. In this interpretation any correlation in the risk factors is purely explained by the structure of the network. i.e., every financial institution is hit by an idiosyncratic shock, and all the correlation comes from the network.

Like before, we consider shocks propagate through a constant network (4). In this scenario we assume that the each shock n in the shock vector ϵ_t (an N -by-1 vector), is a idiosyncratic shock that only affects financial institutions n , and those idiosyncratic shocks, as the name suggested are independent of each other. In other words, the overall model can be described as follows:

$$F_{m,t} = (1 - \Xi_m)^{-1} \theta_m \epsilon_t \quad (6)$$

$$E[\epsilon_t[i] \epsilon_t[j]] = 0 \quad (7)$$

$$\theta_m = I \quad (8)$$

where i and j represent the index of shocks. In this case, θ_m is always an identity matrix, so we define observed networks

$$\Gamma_m = \Xi_m$$

2.2.2 Aggregate Shocks' Network

The second representation assumes that the shocks continue to be independent, but that they affect all financial institutions through a constant propagation. Here all those shocks are aggregate. The banks are all exposed to those shocks, and in addition, there are additional indirect effects through the network. For each network m , the representation is similar to

the previous one

$$F_{m,t} = (1 - \Xi_m)^{-1} \theta_m \epsilon_t \quad (9)$$

$$E[\epsilon_t[i] \epsilon_t[j]] = 0 \quad (10)$$

Define

$$\Gamma_m = I - ((I - \Xi_m)^{-1} \theta_m)^{-1} \quad (11)$$

Then

$$F_{m,t} = (I - \Gamma_m)^{-1} \epsilon_t \quad (12)$$

Here, the second interpretation has exactly the same observed network with the first one. The difference is that the observed network does not directly equals the model network, as it also encodes exposure information. However, understanding the observed network in the second representation is as important as the first one, because the effect of a large exposure to aggregate risk for banks with high centrality is more devastating.

In the identification process, we cannot differentiate between these two models. The two models are equivalent, that θ_m is an over parametrization and for each network Ξ hit by aggregate shocks, we can always find a corresponding Γ ⁹. Klein and Vella (2009) used the second interpretation, while ours uses the first interpretation. We use bank CDS data and assumes that shock are idiosyncratic. For details of the two above interpretations, see the summary in (Lewbel, 2012).

As will become clear in the next section, the model as currently specified cannot be estimated using standard methods. We rely on an method called identification through heteroskedasticity to solve the problem of estimation.¹⁰ Next subsection discusses the so-

⁹This is true as we assume that θ_m is invertable.

¹⁰See Rigobon (2003), Sentana and Fiorentini (2001). For the description of the original procedure see Wright (1928). For further developments see Lewbel (2012), Lewbel (2018a), and Lewbel (2018b). For applications on crises see Caporin et al. (2018) and for applications on monetary policy see Nakamura and Steinsson (2018), Rigobon and Sack (2003), Rigobon and Sack (2004), and Rigobon and Sack (2008).

lution to the identification problem. A crucial ingredient is the presence of heteroskedasticity. Therefore, we assume that there are regimes of economic environments, denoted by $h \in \{1, 2, \dots, H\}$. Let \mathcal{R}_h be the set of time instances t that belongs to heteroskedastic regime h . Furthermore, define $n_h = |\mathcal{R}_h|$ as the number of samples in regime h . If time $t \in \mathcal{R}_h$, then the shock at that time instance is distributed $\epsilon_t \sim \mathcal{N}(0, \Sigma_h)$.

2.3 The identification problem

The identification procedure we use in this paper is related to the identification through heteroskedasticity developed in Rigobon (2003). The intuition of the identification can be developed in a two by two endogenous system of equations, see Appendix A. In this section, we focus on the identification of a multi-bank financial network.

Suppose there are N banks and M types of shock (or networks). Suppose further that the networks $\Gamma_1, \Gamma_2, \dots, \Gamma_M$ are constant over time, but assume the variance of shock ϵ is time-varying, i.e., there is heteroskedasticity in the time series data. Let Σ_h , a diagonal matrix, be the variance of shock ϵ in regime $h \in \{1, 2, \dots, H\}$, be unknown constants.

The model parameters we need to identify are

$$\psi = \begin{bmatrix} \{\Gamma_m\}_{m=1}^M \\ \{\Sigma_h\}_{h=1}^H \end{bmatrix},$$

and the number of parameters for each network m is $N(N-1) + N \sum_h w_{mh}$. The first term comes from the networks. There are $N(N-1)$ elements in each network (diagonal are ones). The second term comes from the variance of the structural shocks. The shock affects N banks for each regime — and there are $\sum_h w_{mh}$ regimes in which the specific network dominates.

In this section, we will start by assuming w_{mh} is known. Later, we will show how it can

be estimated. Then, observed moments (moment constraints) are given by the variance-covariance matrix of X_i in each regime h . Without loss of generality, we can assume that $\forall m, \mathbb{E}[F_{m,t}] = 0$. The following constraints are obtained

$$E_{t \in \mathcal{R}_h} X_t X_t^\top - \sum_{m=1}^M w_{mh} \left((I - \Gamma_m)^{-1} \Sigma_h (I - \Gamma_m)^{-\top} \right) = 0,$$

where $E_{t \in \mathcal{R}_h} X_t X_t^\top$ denotes the expected value of the matrix $X_t X_t^\top$ given that t belongs to regime h . Because $w_{mh} \in \{0, 1\}$ and $\sum_m w_{mh} = 1$, the above constraints can be written as

$$E_{t \in \mathcal{R}_h} X_t X_t^\top - \left((I - \Gamma_m)^{-1} \Sigma_h (I - \Gamma_m)^{-\top} \right) = 0, \forall w_{mh} = 1. \quad (13)$$

In the above, each matrix $E_{t \in \mathcal{R}_h} X_t X_t^\top - \sum_{m=1}^M \left((I - \Gamma_m)^{-1} \Sigma_h (I - \Gamma_m)^{-\top} \right)$ is a N -by- N symmetric matrix, therefore each regime h provides $\frac{N(N+1)}{2}$ moment constraints. In total, $\sum_h w_{mh}$ regimes implies $\frac{N(N+1)}{2} \sum_h w_{mh}$ moment constraints.

The condition for an exact or over-identified model is

$$N(N-1) + N \sum_h w_{mh} \leq \frac{N(N+1)}{2} \sum_h w_{mh} \quad (14)$$

which implies that the data must have at least

$$\sum_h w_{mh} \geq 2 \quad (15)$$

for just-identification and

$$\sum_h w_{mh} \geq 3 \quad (16)$$

for over-identification. i.e., we need each network to dominate in at least 3 regimes to achieve over-identification. This identification is the order condition: how many equations are needed for the system to have less unknowns than knowns. In the single network case, this is simple: the single network will dominate all regimes, hence $H = \sum_h w_{mh}$. In the multiple network

case, we need to figure out how many different mixture of networks (M) present in the data, and which network dominate which regimes. Those two questions are fundamental questions for the identification of many mixture models, and we will discuss them in more detail later in Section 4

2.4 Solution Uniqueness

When there is a single network, parameters in the above model can be estimated via the Generalized Method of Moments (GMM). We define the score as a vector function

$$g(\Gamma, \Sigma_h) = \left\{ X_t X_t^\top - \left((I - \Gamma)^{-1} \Sigma_h (I - \Gamma)^{-\top} \right) \right\}_{t \in \mathcal{R}_h, h=1, \dots, H} \quad (17)$$

where $\{\cdot\}_{h=1, \dots, H}$ denote the H -column vector whose entry are evaluated at $h = 1, \dots, H$.

Estimation is achieved by solving the optimization problem¹¹

$$\begin{aligned} \min_{A, \Sigma_h} & E_{t \in \mathcal{R}_h} g(\Gamma, \Sigma_h)^\top V^{-1} E_{t \in \mathcal{R}_h} g(\Gamma, \Sigma_h) \\ \text{s.t. } & g(\Gamma, \Sigma_h) = \left\{ X_t X_t^\top - \left((I - \Gamma)^{-1} \Sigma_h (I - \Gamma)^{-\top} \right) \right\}_{t \in \mathcal{R}_h, h=1, \dots, H} \end{aligned} \quad (18)$$

Σ_h are diagonal.

GMM requires that the score function is zero iff the system parameters are correct. We can prove that this requirement is satisfied when there is only one network, i.e. $M = 1$. To prepare for the proof, we define $\Sigma = \left[\text{diag}(\Sigma_1) \quad \text{diag}(\Sigma_2) \quad \dots \quad \text{diag}(\Sigma_H) \right]$, where $\text{diag}(\cdot)$ denotes the diagonal entries of the matrix in the form of a column vector. In addition, define $\Omega_h = \mathbb{E}_{t \in \mathcal{R}_h} X_t X_t^\top$

Definition 1. Kruskal rank (Stegeman and Sidiropoulos, 2007) Kruskal rank k of a matrix A is the maximum value of k such that ANY k columns and rows of the matrix A are linearly independent.

¹¹ V is a weighting matrix in the GMM. The GMM is valid as long as V is positive definite, but an optimal V is proportional to the variance-covariance matrix of the score.

Lemma 1. *If (i). $M = 1$, (ii). $I - \Gamma$ has full rank and (iii). The Kruskal rank of Σ is 2 or higher, then $\Omega_h - \left((I - \Gamma)^{-1} \Sigma_h (I - \Gamma)^{-\top} \right) = 0$ has a unique solution.*

Note unlike regular matrix rank, which requires it exists K columns or rows that are linearly independent, the Kruskal rank requires ANY k columns and rows are linearly independent. In the context of our identification problem, this means that the variance of shocks need to present enough heteroskedasticity to obtain a unique solution.

Intuition of the proof:

The identification problem is equivalent to a tensor decomposition problem (for details of this equivalence, refer to the Appendix). According to the Kruskal's rank condition, if

$$Krank(a_r) + Krank(a_r) + Krank(\sigma_r) \geq 2R + 2$$

where $Krank(\cdot)$ stands for Kruskal rank, then the tensor decomposition problem has a unique solution. In the above equation, given assumption (ii), we have $Krank(a_r) = N$. Furthermore, R is the rank of the tensor in (36), which equals MN . Inserting the numbers into the Kruskal's rank condition and considering assumption (i) and (iii), we obtain the solution uniqueness.

3 Results: Single Network Estimation under Endogeneity

Throughout this paper we use Credit Default Swap (CDS) as a measure of the risk of bank bonds. The mechanism of CDS can partly justify the additivity assumption (5): suppose $F_{m,t}$ is a CDS that only covers one specific type of shocks (e.g. interest rate). Then to cover all types of shocks (interest rate, GDP, stock index, etc.), one has to purchase multiple CDS

for all those types of shocks, and the overall cost is the addition of each individual CDS, i.e.,
$$X_t = \sum_{m=1}^M F_{m,t}.$$

Suppose there is a set of shocks that we are interested in. A question arises whether or not we can measure them directly rather than model them as exogenous variables. E.g., suppose we are interested in the risk propagation caused by the fluctuation of interest rate, then we may use the time series of interest rates to fit the model. However, one should note that it is not the interest rate at each time instant that affects CDS, but rather, it is the expectation of interest rate over future time instances that does. Furthermore, it is the change of expectation, rather than the change of interest rate itself, that acts as a shock. Therefore, in the lack of a model for such expectation, it is better to model them as exogenous shocks and obtain their variance through the identification process.

Figure 2 shows the 100-day moving volatility of the 10 banks of interest. The first thing to notice is that they exhibit a significant regime shift of heteroskedasticity over time. Around the time of Nov. 2009, the volatilities of CDS of all 10 banks are high, while the volatilities are significantly lower in mid-2010. In addition, their volatility co-move most of the time, despite regime shifts. The fact that they co-move, allow us to identify a network that is well connected. Apart from similarities in the overall trend, the CDS volatilities exhibit a certain level of bank-specific movements. For example, in early-2016, the CDS of Capital One Financial Corporation has very high volatility that is comparable with that around late 2009. In contrast, the CDS of Morgen Stanley is at a relatively low historical level. That kind of bank-specific characteristic could allow as explore the different risk propagation mechanisms among different banks.

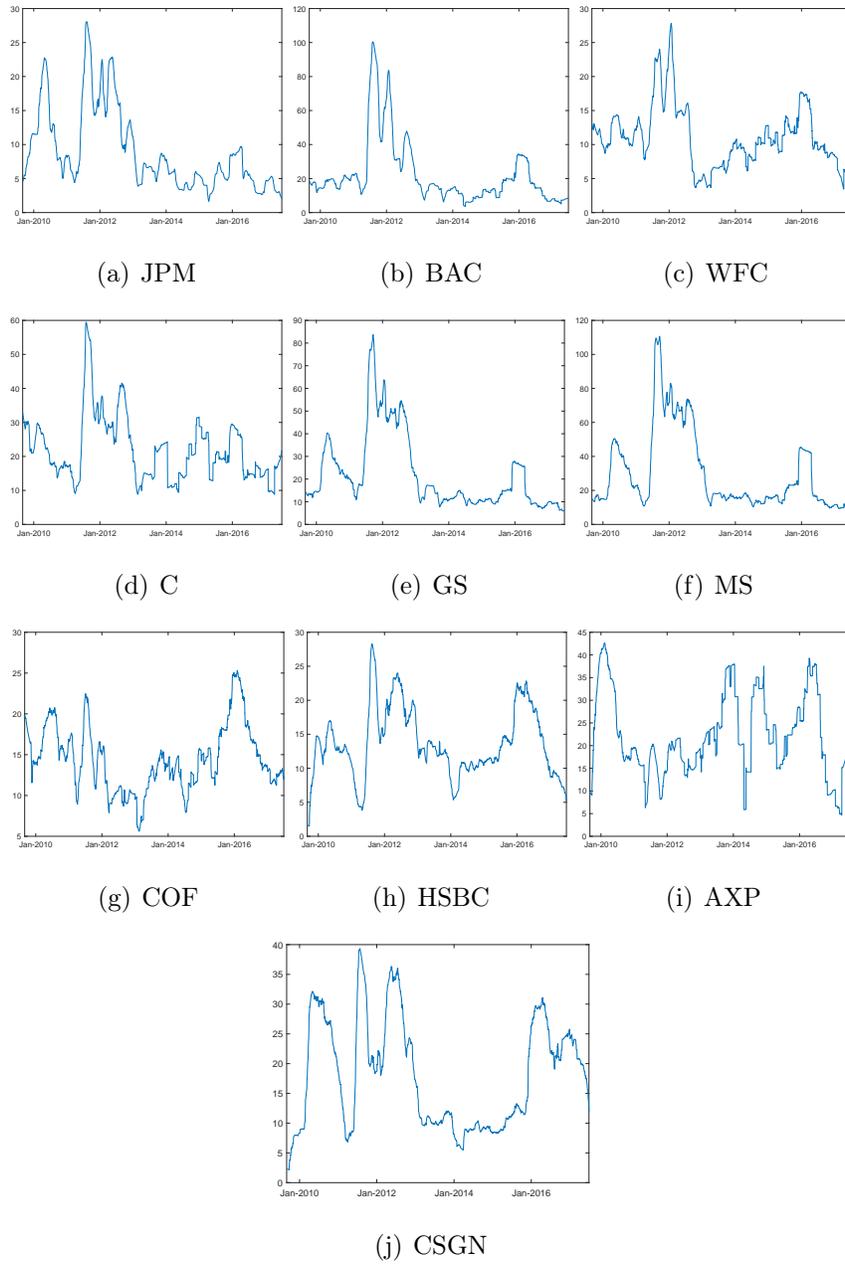


Figure 2: 100 days moving volatility of the 10 banks of interest. CDS data of 10 largest banks in US are collected from Sep. 1, 2009 to June. 20 2017. The stock exchange id tickers of the 10 banks are JPM, BAC, WFC, C, GS, MS, COF, HSBC, AXP, and CSGN.

3.1 Selection of Regimes

As mentioned in previous sections, our model is not uniquely identifiable if the variance of shock is constant. When the variance of shock is not constant, we can divide our data into different heteroskedastic regimes. Because the network parameters are constant over regimes, additional regimes could offer more additional constraints than additional unknowns. With enough heteroskedastic regimes, our model can be uniquely identified (Rigobon, 2003). In practice, however, dividing data into heteroskedastic regimes is not trivial. An effective regime division method should achieve heteroskedasticity among regimes and maintaining homogeneity within regimes. In addition, the divided regimes should have some reasonable interpretations in macroeconomics. There are two broad categories of methods to identify heteroskedastic regimes: regimes divided by statistical properties of the data itself, and regimes divided by other exogenous variables (in our case, it makes sense to use macroeconomic factors). An identification process using regimes defined by statistical properties is named statistical identification, and an identification process using regimes defined by macroeconomic factors is named macroeconomic identification. In this section, we will introduce those two categories of regime divisions in detail, and then apply both to our problem of identifying contingent linear financial networks.

The first category of regime dividing methods is by statistical properties of the data itself. Because we want to separate the variance of shock, it makes sense to look at the quantile level of CDS data volatility. To maximize the separation of unobserved networks, one could use the quantile level of volatility of CDS to define regimes. For example, if there are two banks A and B in the network, one can define four regimes: bank A's volatility is at top 20% quantile level while B is not; bank B's volatility is at top 20% quantile level while A is not; both banks' volatility is at top 20% quantile level, and neither banks' volatility is at top 20% quantile level. Regimes with insufficient number of samples are not used in the

identification process. It is clear that the overall financial network cannot be the same in different regimes.

However, regimes divided by a fixed quantile level is usually very unbalanced, i.e., the low volatility regimes have far more data points than the high volatility regimes. From an economic perspective, this is fine because exceptionally high volatility only occurs during crises. For identification purposes, however, this is not optimal, and many regimes have so few data points to be used in the identification. Alternatively, we could use unsupervised learning techniques, such as K-mean and Gaussian Mixture Model to divide data into groups according to their volatility levels. Intuitively, a clustering algorithm group data points at different time instances into subsets, and maximize the similarities of the volatility vector in each subset. Due to the inherent limitation of most clustering algorithms, the global optimal grouping is usually very difficult to find, and the algorithm is sometimes trapped at local optimal solutions. If the dimension of the volatility vector is very high, i.e., we are identifying the network for a large number of banks, then it is even harder for the algorithm to find the global optimal solutions. In this case, we can reduce the dimension of the volatility vector by applying PCA prior to clustering.

The second category of regime dividing methods is by macroeconomic factors. While the above selection of regimes could maximize the separation of unobserved networks, it does not have any economic interpretation. To draw a connection with the macroeconomic environment, one could use the quantile level of macroeconomic factors instead of quantile levels of CDS volatility. The macroeconomic factors we choose include inflation rate, oil/commodity index, security index, market volatility indicators, currency index, and interest rate/yield. Because the CDS data is available daily, ideally we would like all the above macroeconomic factors to be daily as well. Daily inflation data, in particular, is available through the MIT Billion Price Project (Cavallo and Rigobon, 2016) and the PriceStats data platform.

3.2 Estimates: Statistical Identification

	JPM	BAC	WFC	C	GS	MS	COF	HSBC	AXP	CSGN
JPM		-0.01 (0.03)	0.14 (0.05)	-0.08 (0.02)	0.26 (0.04)	0.03 (0.02)	-0.03 (0.02)	0.07 (0.04)	-0.02 (0.02)	0.04 (0.02)
BAC	0.36 (0.18)		0.99 (0.05)	-0.11 (0.04)	0.53 (0.1)	0.06 (0.07)	0.11 (0.06)	0.09 (0.1)	0.04 (0.03)	0.02 (0.07)
WFC	0.2 (0.05)	0.13 (0.02)		0.08 (0.03)	-0.01 (0.03)	0.01 (0.02)	0.37 (0.04)	-0.02 (0.03)	0.03 (0.01)	-0.09 (0.02)
C	0.03 (0.16)	0.17 (0.04)	0.22 (0.25)		0.11 (0.14)	0.18 (0.09)	0.09 (0.1)	0.1 (0.08)	0.04 (0.04)	-0.06 (0.03)
GS	0.61 (0.08)	0.05 (0.03)	-0.03 (0.08)	0.01 (0.03)		0.47 (0.04)	0.02 (0.04)	0.1 (0.04)	-0.06 (0.02)	-0.03 (0.03)
MS	0.25 (0.07)	0.04 (0.04)	-0.11 (0.08)	0.2 (0.03)	1.0 (0.01)		-0.03 (0.06)	0.07 (0.05)	0.09 (0.03)	0.13 (0.04)
COF	-0.02 (0.06)	-0.02 (0.02)	0.76 (0.06)	0.22 (0.03)	0.04 (0.05)	-0.14 (0.03)		0.01 (0.05)	0.01 (0.03)	0.16 (0.03)
HSBC	0.14 (0.07)	0.02 (0.02)	-0.05 (0.05)	-0.01 (0.02)	0.1 (0.04)	0.06 (0.02)	0.02 (0.03)		-0.12 (0.02)	0.23 (0.04)
AXP	0.27 (0.12)	-0.01 (0.03)	0.12 (0.1)	0.53 (0.05)	-0.45 (0.09)	0.08 (0.06)	0.61 (0.08)	-0.24 (0.11)		0.04 (0.06)
CSGN	0.3 (0.09)	-0.01 (0.06)	-0.34 (0.09)	-0.02 (0.03)	-0.1 (0.08)	0.14 (0.07)	0.21 (0.06)	0.8 (0.09)	0.17 (0.03)	

Table 1: Estimates of the network structure. Standard deviations (in brackets) are obtained by bootstrapping (2000 resamples) across regimes. In this case, regimes are decided by CDS quantile. There are $H = 20$ regimes.

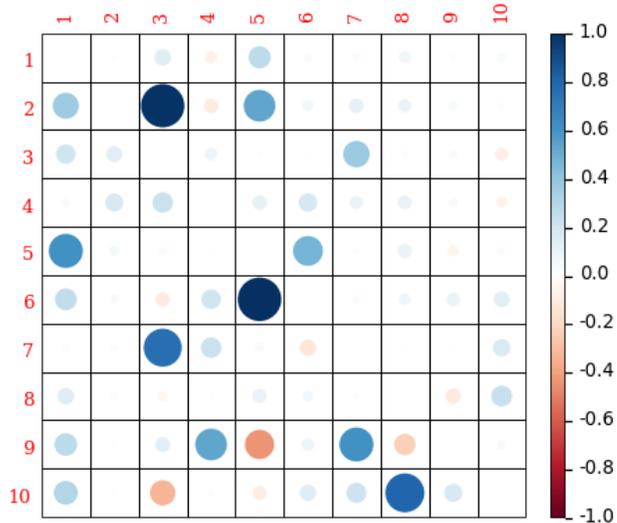


Figure 3: Visualization of the single network identified by statistical regimes

We now turn to the estimation results for our linear network model. As defined in Section 2, we estimate the linear shock propagation channel among the top 10 US banks. In our estimates, we assume that the linear structural parameters that we estimate is always between -1 and 1. i.e.,

$$\gamma_{i,j,m} \in [-1, 1], \forall i, j, m. \quad (19)$$

Mathematically, this constraint will remove any non-unique solutions due to columns permutations. Economically, this means that the bank which receives a shock directly is most affected. This approach of removing permutation solutions will be problematic if a lot of the estimated structural parameters are on the boundary (i.e., the constraint (19) is binding

for a lot of parameters). However, at least for the CDS spread dataset that we study, only 1 parameter is on the boundary in statistical identification, and 3 parameters are on the boundary in macroeconomic identification.

Table 1 shows the estimates of the 10-by-10 network structure using statistical identification.¹² Standard deviations (in brackets) are obtained by bootstrapping (2000 resamples) across regimes. In this case, as the name “statistical identification” suggested, regimes are decided by CDS quantile levels discussed in Section 3.1. The matrix in Table 1 can be regarded as a weighted directed graph. Each column shows where the shock is originated from, and each row shows where the shock propagates to. For example, the structural parameter in the 4th column (Citigroup) and the 1st row (JP Morgan) represents the channel where shock propagates from Citigroup to JP Morgan. In an earlier example in Section 2.1, we used the correlation between Citigroup to JP Morgan as an example to show that the correlation between banks can be very volatile. Here the structural parameter, on the other hand, is reliably estimated with bootstrapping standard deviation of only 0.02.

The estimated matrix of structural parameters is asymmetric in general. However, this does not mean any causal relationship between each pair of banks. In our original model, without (19), any column permutation of Γ will give a new solution and change the direction of edges of the weighted directional graph in Table 1. Now with constraint (19), the directions of the edges of the graph are pinned down by the constraint, but not by any inherent causality in the data.

Figure 3 gives a visualization of the same network. The values of structural parameters are represented by the color and size of the corresponding circles. Positive structural pa-

¹²We do not use Lasso in this small scale estimation problem, because Lasso may introduce bias that affects our statistical tests in the next section. Furthermore, in our identification using CDS data, a reasonable amount of Lasso penalty will not change the result much. However, in a larger scale estimate (e.g., a banking network with 1000+ nodes), Lasso will be very useful in obtaining sparse networks.

rameters are displayed in blue and negative structural parameters are displayed in red. In addition, color intensity and the area of the circles are proportional to the absolute values of the structural parameters. The visualization helps identify patterns in the risk transmission mechanism. For example, Wells Fargo is exposed to a number of different shocks, while Bank of America is more resilient to shocks transmitted from other banks.

3.3 Estimates: Macro Identification

	JPM	BAC	WFC	C	GS	MS	COF	HSBC	AXP	CSGN
JPM		-0.01 (0.02)	0.23 (0.06)	-0.06 (0.03)	0.19 (0.04)	0.04 (0.02)	-0.04 (0.03)	0.23 (0.07)	-0.02 (0.02)	0.05 (0.02)
BAC	0.55 (0.17)		1.0 (0.14)	-0.04 (0.12)	0.34 (0.12)	0.29 (0.08)	-0.36 (0.08)	-0.28 (0.19)	-0.08 (0.03)	0.12 (0.05)
WFC	0.25 (0.05)	0.07 (0.02)		0.03 (0.03)	0.0 (0.03)	0.01 (0.02)	0.19 (0.04)	0.13 (0.04)	0.03 (0.02)	-0.06 (0.02)
C	-1.0 (0.16)	0.25 (0.09)	0.24 (0.27)		0.3 (0.14)	0.35 (0.09)	0.16 (0.19)	-0.6 (0.25)	0.08 (0.04)	-0.18 (0.1)
GS	0.37 (0.07)	0.01 (0.02)	-0.07 (0.05)	0.06 (0.03)		0.52 (0.04)	0.0 (0.04)	0.22 (0.06)	-0.05 (0.02)	-0.13 (0.03)
MS	0.41 (0.13)	-0.0 (0.04)	0.14 (0.11)	0.17 (0.04)	1.0 (0.03)		-0.26 (0.06)	0.01 (0.1)	0.06 (0.04)	0.2 (0.04)
COF	-0.13 (0.07)	0.14 (0.02)	0.57 (0.09)	0.08 (0.05)	0.01 (0.05)	-0.14 (0.04)		0.03 (0.06)	0.01 (0.04)	0.19 (0.03)
HSBC	0.4 (0.1)	0.02 (0.03)	0.17 (0.08)	-0.02 (0.05)	0.15 (0.04)	-0.05 (0.03)	0.01 (0.04)		-0.04 (0.02)	0.26 (0.04)
AXP	-0.03 (0.12)	0.02 (0.04)	-0.04 (0.17)	0.49 (0.05)	-0.55 (0.14)	0.17 (0.1)	0.38 (0.13)	0.14 (0.1)		-0.14 (0.07)
CSGN	0.38 (0.1)	0.1 (0.03)	-0.08 (0.07)	-0.02 (0.03)	-0.24 (0.07)	0.17 (0.04)	-0.04 (0.05)	0.78 (0.08)	0.14 (0.02)	

Table 2: Estimates of the network structure. Standard deviations (in brackets) are obtained by bootstrapping (2000 resamples) across regimes. In this case, regimes are decided by macroeconomic indicators. There are $H = 20$ regimes.

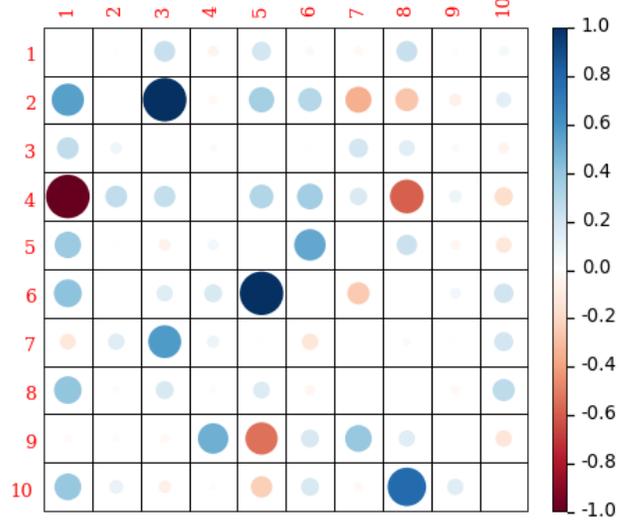


Figure 4: Visualization of the single network identified by macroeconomic regimes

If we select regimes by the quantile levels of macroeconomic factors rather than the quantile levels of CDS spreads, we obtain estimates of the network by macro identification. Table 1 and Figure 3 shows the result of macro identification. Similarly to the statistical identification, Table 2 shows the estimates of the 10-by-10 network structure. Standard deviations (in brackets) are obtained by bootstrapping (2000 resamples) within regimes. Furthermore, Figure 4 gives a visualization of the network. The values of structural parameters are represented by the color and size of the corresponding circles. Positive structural parameters are displayed in blue and negative structural parameters are displayed in red.

In macro identification, standard deviations of structural parameters are estimated by

bootstrapping within each regime, as opposed to bootstrapping across regimes in the entire dataset for statistical identification. This is because the regimes and networks are interchangeable. When we bootstrap the whole dataset, what constitutes network 1 changes. and in fact, the computer has no way to pick network 1 always, and this will exacerbate the standard errors. In the end, the estimates are the mixture of the networks.

Comparing Figure 3 with Figure 4 we notice that the two networks are almost identical. This is because the linear network identified using our approach is the same no matter how regimes are selected. In the last subsection and this subsection, we are using the same dataset, and the model we adopt is a linear model, hence the underlying linear financial network is the same. The bootstrapping standard deviations, on the other hand, are very different between statistical identification and macroeconomics identification. Both methods give consistent estimators but the estimators converge at different rates as the number of data points increases.

4 Multiple Contingent Network Estimation

This section presents a test for the multiple network assumption. In particular, we propose a testing procedure that compares the consistency of the estimates of structural variables in a single network case versus a multiple network case. In a point-wise test, one can obtain the distribution of differences and carry out the test without any assumptions on distributions. However, a point-wise test cannot provide a summary of the results. If we further assume that structural variables follow independent but not necessarily identical Gaussian distributions, we can test the sum of normalized residuals, which follows a Chi-squared distribution.

4.1 F-test for Network Contingency and the Rejection of the Single Network Hypothesis

In this subsection, we construct an intuitive F-test that is used to reject the single network hypothesis. Suppose one observes two sequences of data $\{X_t\}_{t \in D_1}$ and $\{X_t\}_{t \in D_2}$, and estimates structural parameters Γ_{D_1} and Γ_{D_2} . We want to know whether the two sets of structural parameters are consistent. Let $\gamma_{i,j,1}$ and $\gamma_{i,j,2}$ denote the i th row, j th column entry of the network estimated from data set D_1 and D_2 respectively.

We begin with a number of assumptions

Asymptotic Assumptions: (a) The parameter space Ψ of $\psi = \begin{bmatrix} \Gamma \\ \{\Sigma_h\}_{h=1}^H \end{bmatrix}$ is a compact subset of \mathbb{R}^d , and the true value ψ_0 lies in the interior of the parameter space Ψ . (b) The moment function $\psi \rightarrow g(\psi)$ defined in (17) identifies ψ_0 : $g(\psi) = 0$ iff $\psi = \psi_0$. (c) The empirical moment function $\psi \rightarrow \hat{g}(\psi)$ converges uniformly in probability to the moment function $\psi \rightarrow g(\psi)$, namely $\sup_{\psi \in \Psi} \|\hat{g}(\psi) - g(\psi)\| \rightarrow_p 0$ (d) The empirical Jacobian $\hat{G}(\psi) = \frac{\partial}{\partial \psi} \hat{g}(\psi)$ is continuous and is uniformly consistent for the Jacobian matrix, $G(\psi) = \frac{\partial}{\partial \psi} g(\psi)$, i.e., $\sup_{\psi \in \Psi} \|\hat{G}(\psi) - G(\psi)\| \rightarrow_p 0$ (e) The matrix $G(\psi_0)^\top G(\psi_0)$ is positive definite. (f) The empirical moment function evaluated at the true parameter value obeys a central limit theorem:

$$\sqrt{n} \hat{g}(\psi_0) \sim N(0, \Omega)$$

asymptotically, where n is the number of samples.

Note those assumptions are inherently the same with the assumptions in the original GMM paper by Hansen (1982). Note that the key to our single network test is the solution uniqueness assumption (b) of GMM summarized above. Failing to establish a result for solution uniqueness will cause incorrect estimates of parameter distribution. Generally

speaking, the bootstrap estimated parameter distribution will have a larger variance and therefore fails to reject the hypothesis even if the hypothesis is incorrect. Lemma 1 in our paper deals with this problem.

Under the null hypothesis that

$$H_0 : \gamma_{i,j,1} \text{ and } \gamma_{i,j,2} \text{ are the same}$$

their difference

$$\gamma_{i,j,1} - \gamma_{i,j,2} \tag{20}$$

should follow a distribution with zero mean. If the estimated value of $\hat{\gamma}_{i,j,1} - \hat{\gamma}_{i,j,2}$ lies in the 0.05 left or right quantiles of the bootstrapping distribution, we can reject the null hypothesis and claim that with 90% confidence

$$H_1 : \Gamma_{D_1} \text{ and } \Gamma_{D_2} \text{ are different}$$

The above point-wise test has the advantage of distribution-free. However, without a summarizing statistic, one cannot draw conclusions on the overall network. Suppose further that $\gamma_{i,j,1} - \gamma_{i,j,2}$ follows Gaussian distribution $\mathcal{N}(\bar{\gamma}_{i,j}, \sigma_{i,j}^2)$. Under the null hypothesis that

$$H_0 : \Gamma_{D_1} \text{ and } \Gamma_{D_2} \text{ are the same} \tag{21}$$

their squared difference $(\gamma_{i,j,1} - \gamma_{i,j,2})^2 / \sigma_{i,j}^2$ follows a Chi-squared difference with degree of freedom 1.

In addition, under the independence assumption, the sum of squared differences

$$\sum_{i \neq j} (\gamma_{i,j,1} - \gamma_{i,j,2})^2 / \sigma_{i,j}^2 \sim \mathcal{K}(N(N-1))$$

where \mathcal{K} denotes a Chi-squared distribution.

If the estimated value of $\sum_{i \neq j} (\gamma_{i,j,1} - \gamma_{i,j,2})^2 / \sigma_{i,j}^2$ lies in the 0.1 right quantiles of the bootstrapping distribution, we can reject the null hypothesis and claim that with 90% confidence

$$H_1 : \Gamma_{D_1} \text{ and } \Gamma_{D_2} \text{ are different}$$

4.2 Estimates: the Sensitivity of Network Structures to Macroeconomic Environments

Using the statistical test derived in the last subsection, we are able to compare the network contingency given any two sets of data. In this subsection, we divide our dataset according to quantile levels of macroeconomic factors and test the network contingency to those factors. This procedure is analogous to a sensitivity test of our model. For example, we first look at the network contingency to the inflation rate.

More specifically, let X_t , a 10-by-1 vector time series, denote the CDS spread of top 10 US banks, and S_t denote the time series of a macroeconomic factor. Similarly to Section 3, macroeconomic factors in this paper include Price Index, WTI, S&P Index, VIX Volatility, 3-month Interest Rate, US Dollar Index, and 10-year minutes 3-month Interest Yield Curve. The dataset is divided into two parts

$$D_1 = \{X_t | S_t \leq \text{median}[S_t]\}, D_2 = \{X_t | S_t > \text{median}[S_t]\}$$

and both the pointwise test and Chi-squared test in Section 4.2 is carried out on these two sub-datasets. The result of tests are summarized in Table 3. In this table, we do not list the corresponding p -values because the J-statistics is large enough to reject the null hypothesis with p -value less than 0.01 in all scenarios.

	Mean diff	Mean std	Num. rejections	J-stat
WTI	0.212803	0.131362	23	373.41
SP500	0.180177	0.138155	11	340.13
VIX	0.156272	0.119048	23	256.89
i_3M	0.158900	0.141999	15	315.09
USD	0.188294	0.136103	19	375.21
i_10Y-i_3M	0.185558	0.152019	20	279.12
Inflation	0.143617	0.141792	22	176.85

Table 3: Structural changes driven by macroeconomic indicators

Among the seven macroeconomic factors, network structural changes are most significant when asset prices are used to split the sample. Inflation and Volatility (as measured by the VIX) have the least rejections. Nevertheless, in all seven cases we reject the hypothesis of single network.

4.3 Estimating mixture models using EM-Algorithm

Although we build the model with a mixture of multiple networks, so far, the identification process has mostly been with a single network. In the case of a single network, the mixture random variable $w := \{w_{mh}\}_{m=1, h=1}^{M, H}$ is trivial, and the identification is through the GMM algorithm. Because the last subsection showed that a single network is not sufficient to describe the data, we move towards the model setup of a multi-network mixture.

Recall the main characteristics of our mixture model as follows: the data X_t in each regime $t \in \mathcal{R}_h$ is generated by

$$X_t = \sum_{m=1}^M w_{mh}(1 - \Gamma_m)^{-1}\epsilon_t \quad (22)$$

where

$$\epsilon_t \sim \mathcal{N}(0, \Sigma_h)$$

$\Gamma := \{\Gamma_m\}_{m=1}^M$ and $\Sigma := \{\Sigma_m\}_{h=1}^H$ are parameters to be identified. In addition, w_{mh} denote the indicator random variable that equals to 1 if network m dominates in regime h . Assume

$$w_{mh} = \begin{cases} 1 & \text{w.p. } p_m \\ 0 & \text{o.w.} \end{cases}. \text{ Because only one network dominates in each regime,}$$

$$\sum_{m=1}^M p_m = 1 \tag{23}$$

Our objectives is to estimate parameters Γ and Σ by observing X_t .

Similarly to the case of single network identification, we construct

$$\Omega_h = \frac{1}{n_h} \sum_{t \in \mathcal{R}_h} X_t X_t^\top$$

the maximum likelihood estimator (MLE) of population variance of X_t in each heteroskedasticity regimes h . As it will become clear later in this subsection, using Ω_h will allow us to separate the identification of Γ and Σ . Furthermore, it allows us to estimate each Γ_m separately.

The above mixture model is difficult to identify, because w_{mh} is a random variable which is not observed. In this case, the Expectation-Maximization (EM) Algorithm can be used to make the problem tractable. Note that $\Omega := \{\Omega_h\}_{h=1}^H$ given w is Wishart distributed, even though Ω itself is not. The EM Algorithm takes this advantage by iterating between the E step and M step. In the E step, it computes the discrete distribution of w_{mh} given current parameter estimates, and calculate the function

$$Q(\Gamma, \Sigma \mid \Gamma^{(current)}, \Sigma^{(current)}) = \mathbb{E}_{w \mid \Omega, \Gamma^{(current)}, \Sigma^{(current)}} [\log L(\Gamma, \Sigma; \Omega, w)]$$

where $L(\Gamma, \Sigma; \Omega, w)$ is the likelihood function assuming w is observable. In the M step, it computes the optimal parameters Γ and Σ to maximize the $Q(\cdot)$ function. It is a general

result of EM that improving $Q(\cdot)$ improves the likelihood function of the mixture model $L(\Gamma, \Sigma; \Omega)$, see Little and Rubin (2019).

A critical step of implementing the EM algorithm is to calculate the likelihood function of Ω given w . Given $w_{mh} = 1$, i.e. network m dominates in regime h , we have

$$\mathbf{v}_{mh} := n_h(1 - \Gamma_m)\Omega_h(1 - \Gamma_m)^\top = \sum_{t \in \mathcal{R}_h} \epsilon_t \epsilon_t^\top \sim \text{Wishart}(\Sigma_h, n_h) \quad (24)$$

The Wishart distributed \mathbf{v}_{mh} has a probability density function (PDF)

$$f_{\mathbf{v}}(\mathbf{v}_{mh}) = \frac{1}{2^{n_h N/2} |\Sigma_h|^{n_h/2} \Gamma_N\left(\frac{n_h}{2}\right)} |\mathbf{v}_{mh}|^{(n_h - N - 1)/2} e^{-(1/2) \text{tr}(\Sigma_h^{-1} \mathbf{v}_{mh})} \quad (25)$$

where n_h is the number of samples in regime h , N is the number of banks in the network, $|\cdot|$ is the determinant of a matrix, and $\Gamma_N(\cdot)$ is the multivariate Gamma function with dimension N .

Now we can write the likelihood function $L(\Gamma, \Sigma; \Omega, w)$ using the Wishart PDF

$$L(\Gamma, \Sigma; \Omega, w) = \prod_m \prod_h \left(f_{\mathbf{v}}(\mathbf{v}_{mh}) \right)^{\mathbb{I}_{w_{mh}=1}} \quad (26)$$

Recall that we can separate the estimation of the network matrices Γ with the shock variances Σ in the single network case. Similarly, we can also do that in EM, but for a different reason: the optimal diagonal Σ_h that maximize (25) equals to the diagonal entries of \mathbf{v}_{mh} . i.e.,

Lemma 2. *Let \mathbf{v}_{mh} be a given N -by- N positive semidefinite matrix. Also, define \mathbb{PD} as the set of positive semidefinite diagonal matrices. Then*

$$\Sigma_h^* := \arg \max_{\Sigma_h \in \mathbb{PD}} \frac{1}{2^{n_h N/2} |\Sigma_h|^{n_h/2} \Gamma_N\left(\frac{n_h}{2}\right)} |\mathbf{v}_{mh}|^{(n_h - N - 1)/2} e^{-(1/2) \text{tr}(\Sigma_h^{-1} \mathbf{v}_{mh})} = \frac{1}{n_h} \text{diag}(\mathbf{v}_{mh}) \quad (27)$$

The proof is given in Appendix C.

The above Lemma means that we do not have to identify the join of Γ and Σ . We can just identify Γ , and the optimal Σ can be obtained analytically. Now that we have separated the identification of the network Γ with Σ , we focus on a EM-algorithm that identifies Γ . At this point, we are looking at a mixture model with parameters Γ , hidden random variables w , and observed random variables Ω . We define the function Q as follows

$$\begin{aligned}
Q(\Gamma \mid \Gamma^{(current)}) &= \mathbb{E}_{w \mid \Omega, \Gamma^{(current)}} [\log L(\Gamma; \Omega, w)] \\
&= \sum_m \sum_h \mathbb{E}_{w \mid \Omega, \Gamma^{(current)}} \mathbb{I}_{w_{mh}=1} \log f_{\mathbf{V} \mid \Sigma^*}(\mathbf{v}_{mh}) \\
&= \sum_m \sum_h p_{mh} \log f_{\mathbf{V} \mid \Sigma^*}(\mathbf{v}_{mh})
\end{aligned} \tag{28}$$

where $f_{\mathbf{V} \mid \Sigma^*}$ is the Wishart PDF given optimal Σ , and $p_{mh} := \mathbb{P}(w_{mh} = 1 \mid \Omega_h; \Gamma)$ is the probability that network m dominates in a specific regime h . Inserting (27) into (25), we have

$$f_{\mathbf{V} \mid \Sigma^*}(\mathbf{v}_{mh}) = \frac{|\mathbf{v}_{mh}|^{(n_h - N - 1)/2}}{|\text{diag}(\mathbf{v}_{mh})|^{(n_h/2)}} \cdot \text{constant not depend on } \mathbf{v}_{mh}$$

hence

$$Q(\Gamma \mid \Gamma^{(current)}) = \sum_m \sum_h p_{mh}^{(current)} \left(\frac{n_h - N - 1}{2} \log |\mathbf{v}_{mh}| - \frac{n_h}{2} \log |\text{diag}(\mathbf{v}_{mh})| \right) + \text{constant} \tag{29}$$

where $\mathbf{v}_{mh} = n_h(1 - \Gamma_m)\Omega_h(1 - \Gamma_m)^\top$.

Another step that plays an important role in EM algorithm is the estimate of p_{mh} and p_m . Note that p_{mh} and p_m are two different quantities: p_{mh} is the posterior probability of $w_{mh} = 1$ for a specific h . On the other hand, p_m is the prior distribution of w_{mh} , which is the same for all regimes. According to Bayes rule

$$\begin{aligned}
p_{mh} &:= \mathbb{P}(w_{mh} = 1 \mid \Omega_h; \Gamma^{(current)}) \\
&= \mathbb{P}(w_{mh} = 1; \Gamma^{(current)}) \cdot \mathbb{P}(\Omega_h \mid w_{mh} = 1; \Gamma^{(current)}) \cdot \text{constant} \\
&= p_m^{(current)} \cdot f_{\mathbf{V}}^{(current)}(\mathbf{v}_{mh}) \cdot \text{constant}
\end{aligned} \tag{30}$$

where $p_m^{(current)}$ can be estimated by $p_m^{(current)} = \frac{1}{H} \sum_{h=1}^H p_{mh}^{(current)}$, $f_{\mathbf{v}}$ is defined in (25) and the normalizing constant is decided by $\sum_{m=1}^M p_{mh} = 1$ ¹³.

EM-Algorithm

Take an initial guess of Γ , then iterate between the following E-step and M-step.

E-step:

Update p_{mh} , the probability that network m dominates in regime h , given current estimates of Γ , according to (30). After that, we use the updated p_{mh} to construct

$$Q(\Gamma | \Gamma^{(current)}) = \sum_m \sum_h p_{mh}^{(current)} \left(\frac{n_h - N - 1}{2} \log |\mathbf{v}_{mh}| - \frac{n_h}{2} \log |diag(\mathbf{v}_{mh})| \right) + constant \quad (31)$$

where $\mathbf{v}_{mh} = n_h(1 - \Gamma_m)\Omega_h(1 - \Gamma_m)^\top$.

M-step:

Update estimates of Γ by maximizing $Q(\Gamma | \Gamma^{(current)})$. It is sufficient to update Γ_m separately by

$$\max_{\Gamma_m} \sum_h p_{mh}^{(current)} \left(\frac{n_h - N - 1}{2} \log |\mathbf{v}_{mh}| - \frac{n_h}{2} \log |diag(\mathbf{v}_{mh})| \right)$$

for each m .

4.4 The number of networks

To obtain the optimal number of networks, we apply the Bayesian information criterion. The Bayesian information criterion is defined as

$$\text{BIC} = \ln(n)k - 2 \ln(L).$$

where n is the number of samples, k is the number of parameters, and L is the maximum mixture log-likelihood function, see Fraley and Raftery (1998).

¹³This process is similar to the case of Gaussian mixture models, see Chapter 9 of Bishop (2006)

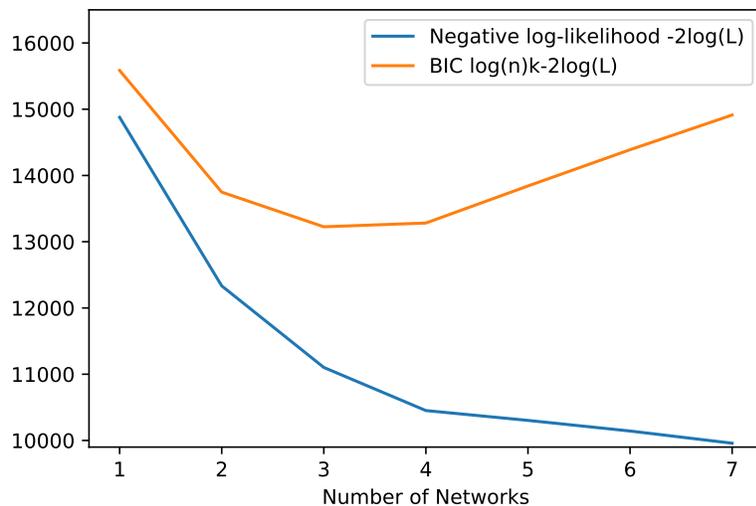


Figure 5: Negative log-likelihood and BIC vs. No. of networks

As shown in Figure 5, the optimal number of networks selected by Bayesian information criteria is 3.

4.5 Network Centrality

In our macro-identification process, we divide regimes according to macroeconomic shocks. According to the results of EM, it turns out some of those shocks play important roles in the network switching process, while others only contribute to the network switching marginally. In Table 4, regimes are grouped by networks in which they dominate. e.g., in regimes 1, 2, 3, 6, 14, 16, 17 and 18, network 1 dominates. On top of that, we list whether or not each macroeconomic factor is active in each regime.¹⁴ Active factors are denoted by 1 and others by 0.

¹⁴By active macroeconomic factor, we mean that the specific macroeconomic factor is in its top 25% quantile in that regime.

		Inflation	WTI	S&P	VIX	i_3M	USD	i_10Y-i_3M
Network 1	regime 1	0	0	0	0	0	0	0
	regime 2	1	0	0	0	0	0	0
	regime 3	0	1	0	0	0	0	0
	regime 6	1	0	0	1	0	0	0
	regime 14	0	0	0	1	1	1	0
	regime 16	1	0	0	0	0	0	1
	regime 17	0	1	0	0	0	0	1
	regime 18	1	1	0	0	0	0	1
Network 2	regime 4	0	0	1	0	0	0	0
	regime 8	1	0	1	0	1	0	0
	regime 9	0	0	1	0	0	1	0
	regime 10	1	0	1	0	0	1	0
	regime 11	0	0	0	1	0	1	0
	regime 12	0	0	1	0	1	1	0
	regime 13	1	0	1	0	1	1	0
Network 3	regime 5	0	0	0	1	0	0	0
	regime 7	0	1	0	1	0	0	0
	regime 15	0	0	0	0	0	0	1
	regime 19	0	0	0	1	0	0	1
	regime 20	1	0	0	1	0	0	1

Table 4: Network vs. Macro shocks

From Table 4, we can summarize what each network represents. Network 1, which includes the regime where no macroeconomic factor is active, can be interpreted as normal times. Network 2, on the other hand, is very likely to be an asset price shocks, because S&P and exchange rates are both moving. Network 2 also reflects some short-run monetary policy that is usually reflected in asset prices. Finally, Network 3 is clearly an uncertainty shock, as it corresponds to VIX and yield curve activities. We also find that those shocks have different transmission mechanisms within the networks as can be seen in Figure 7.

Now that we know there are three networks in the interbank financial system, a systemic risk measure, namely the Katz centrality, will be calculated for each of the networks. The Katz centrality (Katz, 1953; Junker and Schreiber, 2008) of a network with adjacency matrix Γ is defined as

$$\vec{C}_{\text{Katz}} = ((I - \alpha A^T)^{-1} - I) \vec{1}, \quad (32)$$

where α is a damping factor that satisfies $0 \leq \alpha < 1/|\lambda_{\max}|$. In our paper, we choose $\alpha = 0.5$.

The Katz centrality is a generalization of the degree centrality. Intuitively, a node in the graph is more important if it more often receives shocks from other nodes. Furthermore, the Katz centrality considers both the direct impact from other nodes as well as the cascade impact many steps ago. It assumes that both direct and indirect impacts affect the importance of a node, given that indirect impacts are discounted by a factor of α after each step. Apart from original applications in social networks and biological networks, Katz centrality has also been applied to evaluate systemic risk in financial networks, see (Thurner and Poledna, 2013; Temizsoy et al., 2017). The Katz centrality of the top 10 banks in the US in the three estimated networks is shown in Table 5.

To understand the economic intuition of the Katz centrality, we first look at an alternative definition of it

$$C_{\text{Katz}}(i) = \sum_{k=1}^{\infty} \sum_{n=1}^N \alpha^k (\Gamma^k)_{ni}$$

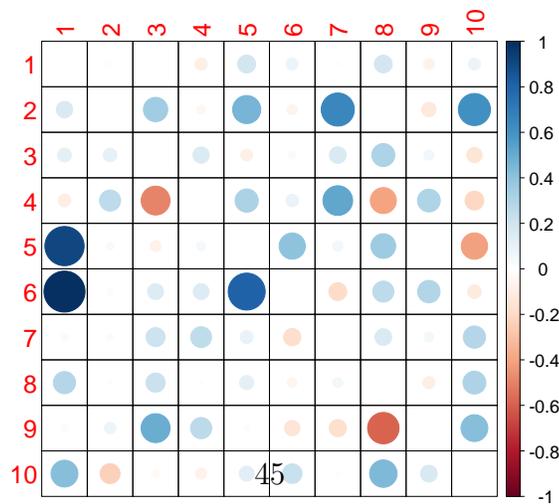
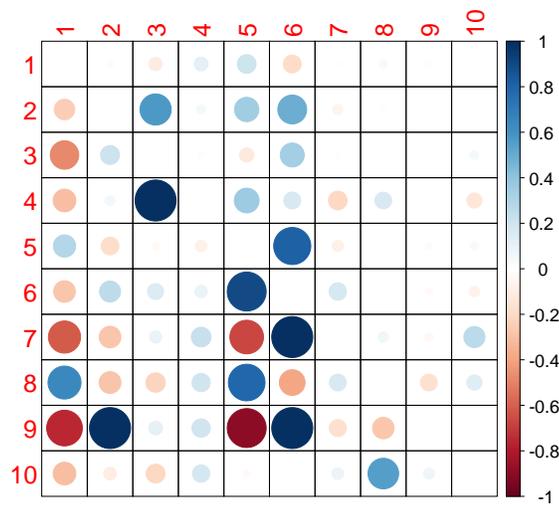
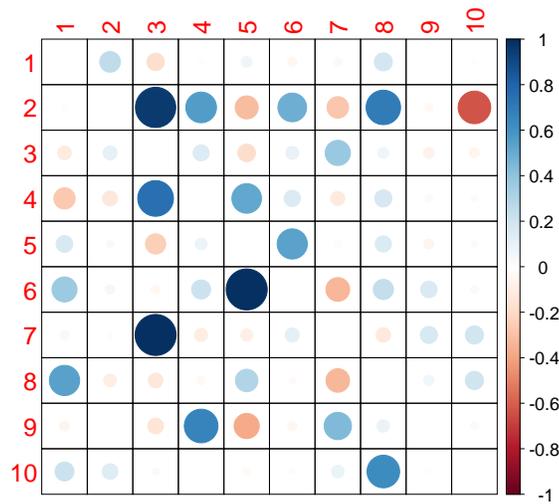


Figure 6: Estimated 3 networks

	Network 1	Network 2	Network 3
JPM	0.041 (0.005)	0.008 (0.013)	0.055 (0.015)
BAC	0.150 (0.018)	0.176 (0.017)	0.195 (0.045)
WFC	0.059 (0.006)	0.046 (0.013)	0.066 (0.011)
C	0.124 (0.008)	0.152 (0.022)	0.054 (0.039)
GS	0.100 (0.012)	0.122 (0.017)	0.139 (0.021)
MS	0.181 (0.016)	0.188 (0.021)	0.221 (0.025)
COF	0.117 (0.018)	0.054 (0.039)	0.075 (0.032)
HSBC	0.037 (0.016)	0.066 (0.036)	0.075 (0.015)
AXP	0.084 (0.020)	0.144 (0.033)	0.027 (0.016)
CSGN	0.107 (0.012)	0.044 (0.050)	0.095 (0.036)

Table 5: Estimates of network Katz centrality. Centrality values are normalized to sum up to 1. We use a discount factor $\alpha = 0.5$. Standard deviations (in brackets) are obtained by bootstrapping (2000 resamples) across regimes. In this case, regimes are decided by macroeconomic indicators. There are $H = 20$ regimes.

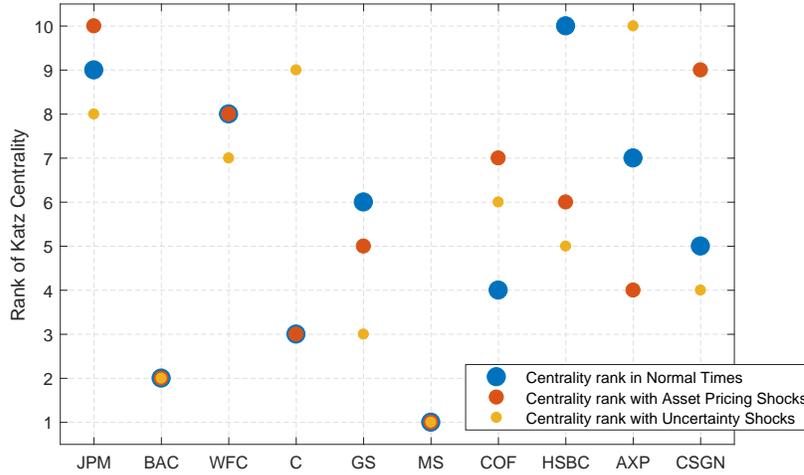


Figure 7: Rank of the estimated Katz centrality of in the 3 estimated networks.

From this definition, it is easy to see that the Katz centrality calculates the summation of an infinite series of impacts given a uniform shock to each bank. In our estimates, centrality values are normalized to sum up to 1, therefore each value \tilde{C}_i just means the portion of shock that is transmitted through bank i .

In summary, we find the data can be explained by 3 networks in the financial network among the top 10 banks in the US with our criteria. We reject the hypothesis of 1 network using an F-test and then use the Bayesian information criteria to conclude that 3 networks are optimal. With 4 or more networks, the model complexity penalty term in the BIC would stand out and reject the models. We are only applying our identification method on financial networks though, other applications of our identification method could give 4 or more networks as the optimal solution.

The three different networks we estimates imply different economic behaviors. The first network represents normal time, the second network represents an equity market shock, and the third network represents a VIX shock. Furthermore, the Katz centrality in different

networks is not the same. Interestingly, some banks are always systemically important (i.e., Bank of America and Morgan Stanley), but depending on the type of shocks other banks might change their rank of centrality.

Without the multi-network assumption in this paper, we would have incorrectly estimated a single network. In that case, the centrality rank assuming 1 network is given by JPM: 10, BAC: 2, WFC:9, C:7, GS:3, MS:1, COF:5, HSBC:8, AXP:6, CSGN:4. The resulting centrality measures, or any other systemic risk measures, will only be the weighted average of different scenarios in general. As a result, we lack the ability to identify banks that is systematic with respect to some specific shocks. Furthermore, with WFC for example, the centrality rank assuming 1 network is only 9th place, but it is actually more important in all three networks. Such an incorrect estimate of systemic risks could result in suboptimal decisions for financial practitioners, policymakers and regulators.

5 Conclusions

Understanding the interconnections within the financial system has been a first-order concern in developed economies since the 2008 global financial crises. In fact, macroeconomic prudential regulation needs to determine which banks and financial institutions are systemically important so they can be supervised closely. In the network language, it would mean that such financial institutions have a large centrality. Most of this analysis has been done either by concentrating on symmetric responses (computation based on correlations) or by observing a subset of financial contracts. This approach has been quite fruitful. In our view, however, both approaches might be incomplete.

On the one hand, the estimation of networks tends to obviate the directionality of the effects. In other words, Bank of America might have a large impact on American Express,

but the opposite might not be true. Network analysis in the literature tends to obviate this feature. The second approach, which concentrates on the detail description of the contingent contracts across banks, could represent a solution to this problem. However, it is virtually impossible to observe all possible contracts. Therefore, a market price of CDS is conceivably a more reliable measure of the actual exposure. This ambivalence implies that each procedure has a weakness that we have tried to address in the current paper. Furthermore, as has been shown in the literature, the nature of the transmission mechanisms changes with the shocks hitting the economy – meaning that the network is contingent on the state of the economy. We argue that the estimation of an asymmetric and contingent network requires a different identification method. We develop a methodology based on identification through heteroskedasticity. Applying this estimation method on CDS data of 10 large banks, we construct a financial network model and find that the data generating the model is consistent with three networks, each one contingent on one different macroeconomic shock.

Our results indicate that the systemically important bank depends on the type of shock that hits the economy – which is ultimately transmitted through a different network. Without our contingent financial network model, it is not possible to identify the importance of each bank in the financial system when a specific type of shock hits the economy. Indeed, we reject that the data is explained by a single network – suggesting that a policy designed based on that network would be inappropriate when a different transmission mechanism governs the dynamics of the system.

From the regulatory point of view, understanding the relative rankings on the financial institutions and how such ranking shifts in the sample is important. Our data is short and therefore we are limited in our ability to observe shocks that have not happened yet. For instance, we have not observed large positive productivity shocks, or relatively high inflation rates, or even high interest rates. Therefore, our conclusions are conditional to the sample we have seen. Within that sample, though, it is easy to identify tranquil times, periods

when there are shocks to asset prices, and times where uncertainty is high. In those settings, which bank is systemically important changes. Central banks and regulators need to pay attention to these changes if the proper macro-prudential regulation is to be designed. In fact, regulators with misleading information about the financial network may not be able to make the most appropriate policy decisions to minimize the impact of those shocks.

Furthermore, the application of our identification method for contingent networks is not limited to policymaking. For example, asset management practitioners could use our method to estimate the contingent network and allocate assets according to the dominate shock in a period of time, and macroeconomists could use our method to evaluate the impact of macroeconomic interventions. In general, how to model, estimate and intervene in shock contingent networks is still an open and important topic for future research.

Appendices

A A Single Network Endogenous Model

Assume two banks are related according to the following system of equations and shocks

$$x_t = \beta y_t + \epsilon_t \tag{33}$$

$$y_t = \alpha x_t + \eta_t \tag{34}$$

with reduced form,

$$x_t = \frac{1}{1 - \alpha\beta} (\beta\eta_t + \epsilon_t)$$
$$y_t = \frac{1}{1 - \alpha\beta} (\eta_t + \alpha\epsilon_t)$$

where η_t and ϵ_t are the structural shocks and α and β are describing the network.

As it is, this model cannot be estimated from the data. Equations (33) and (34) describe the behavior of the data entirely with 4 parameters/variables: two shocks ϵ and η and two parameters α and β . These four constitute the unknowns of the system. The problem of identification arises because there are three equations in four unknowns. The observable variables x and y have zero mean and in the data only three moments can be estimated; all from the variance-covariance matrix. What are the solutions to the problem? In economics, solutions tend to create circumstances in which an additional equation is added to the system of equations. For instance, the exclusion restriction in the instrumental variable approach boils down to assuming that one parameter is zero (the exclusion assumption).

Randomized controlled trials assume that all the variation is due to the treatment — again, this is implicitly assuming that there is no feedback effect ($\beta = 0$). This is a very reasonable assumption when the experiment is properly designed. All these solutions are making a parameter assumption (usually that a parameter is equal to zero). The identification through heteroskedasticity has a slightly different flavor. The easiest way to explain how identification through heteroskedasticity works is to show the system of equation. Assume that the parameters are stable and that the data has heteroskedasticity. For simplicity assume that there are two heteroskedastic regimes. In this case, it is possible to estimate one variance-covariance matrix in each regime.

$$\begin{aligned}\Omega_1 &= \begin{bmatrix} \text{var}(x_{t,1}) & \text{covar}(x_{t,1}, y_{t,1}) \\ & \text{var}(y_{t,1}) \end{bmatrix} = \frac{1}{(1 - \alpha\beta)^2} \begin{bmatrix} \sigma_{\epsilon,1}^2 + \beta^2\sigma_{\eta,1}^2 & \alpha\sigma_{\epsilon,1}^2 + \beta\sigma_{\eta,1}^2 \\ & \alpha^2\sigma_{\epsilon,1}^2 + \sigma_{\eta,1}^2 \end{bmatrix} \\ \Omega_2 &= \begin{bmatrix} \text{var}(x_{t,2}) & \text{covar}(x_{t,2}, y_{t,2}) \\ & \text{var}(y_{t,2}) \end{bmatrix} = \frac{1}{(1 - \alpha\beta)^2} \begin{bmatrix} \sigma_{\epsilon,2}^2 + \beta^2\sigma_{\eta,2}^2 & \alpha\sigma_{\epsilon,2}^2 + \beta\sigma_{\eta,2}^2 \\ & \alpha^2\sigma_{\epsilon,2}^2 + \sigma_{\eta,2}^2 \end{bmatrix}\end{aligned}$$

There are six unknowns in the system. The two parameters (α and β), and four variances ($\sigma_{\epsilon,1}^2$, $\sigma_{\epsilon,2}^2$, $\sigma_{\eta,1}^2$, and $\sigma_{\eta,2}^2$). As can be seen, there are six equations in six unknowns. This means that the system of equations is just identified.

Notice that even though in each regime the system is under-identified (fewer equations than unknowns) the system as a whole is identified. The key assumptions are two: that the structural shocks are indeed structural (they are uncorrelated) and that the parameters are stable. In the end, the parameter stability allows the heteroskedasticity to add additional equations — which helps solve the identification problem.

The intuition of the two endogenous variables case is as follows. First, the graphical representation of the joint residuals in this model always takes the form of a rotated ellipse. Second, the rotation is summarized by the variance-covariance matrix.

In equations (33) and (34), the only meaningful moment we can compute to estimate the

degree of contagion is the covariance matrix. An important question is then, what does the variance-covariance matrix represent? The errors in these models are distributed as a multinomial and their contours are ellipses. To fix concepts, let us start with a simple endogenous system of equations (33) and (34). The 95th percentile of the errors is distributed as a rotated ellipse. We can solve for two independent normal distributions from the structural equations as follows (with some abuse of notation)

$$\begin{aligned}\phi_1 &= \frac{x_t - \beta y_t}{\sigma_\epsilon} \sim N(0, 1) \\ \phi_2 &= \frac{y_t - \alpha x_t}{\sigma_\eta} \sim N(0, 1)\end{aligned}$$

Because ϕ_1 and ϕ_2 are independent with mean zero and variance one, it is possible to describe the ζ confidence interval as $\phi_1^2 + \phi_2^2 = \zeta$. This is exactly an ellipse. Substituting

$$\left(\frac{x_t - \beta y_t}{\sigma_\epsilon}\right)^2 + \left(\frac{y_t - \alpha x_t}{\sigma_\eta}\right)^2 = \zeta \tag{35}$$

The two axes of the ellipse cannot be computed in closed-form solution, but they depend on the slope of the curves (structural parameters) as well as the relative variances of the shocks. In Figure A.1 a graphical representation is shown. Suppose that the blue curve represents the supply and the red is the demand (when there are no shocks). Then x_t represent quantity and y_t represent price. Furthermore, the points reflect some random realization of structural shocks that leads to a point far from the depicted schedules. The ellipse represents the 90th percentile. In this particular case β is assumed to be negative (representing the “demand”), while α is positive. In Figure A.1, the variance of the demand shocks is larger than the variance of the shocks to the supply, hence, the ellipse is closely aligned with the supply curve. In the limit, if the variance of the demand is infinitely large, the ellipse would coincide exactly with the supply curve.

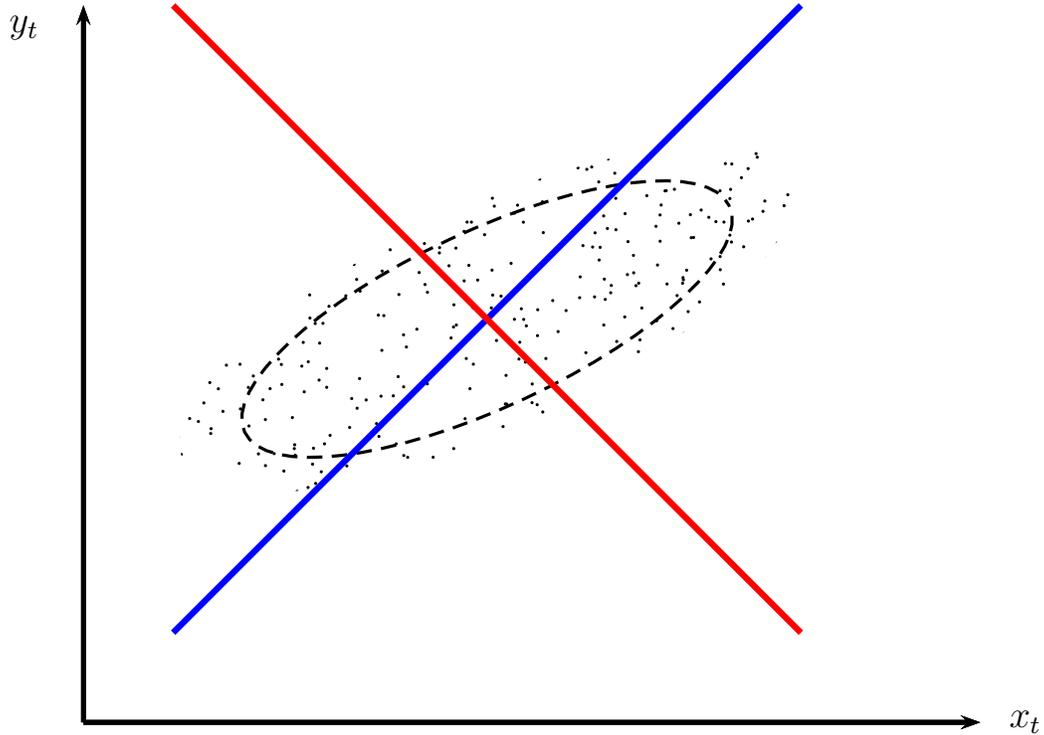


Figure A.1: Distribution of Errors

The form of the ellipse is also summarized by the variance-covariance matrix computed in the reduced form. Additionally, most of the methodologies we study are based on the variance-covariance matrix. Therefore, all the sources of bias can be tracked to it. Finally, as mentioned previously, in this model the only statistic that can be computed from the data — that allows us to recover the structural parameters — is the variance-covariance matrix.

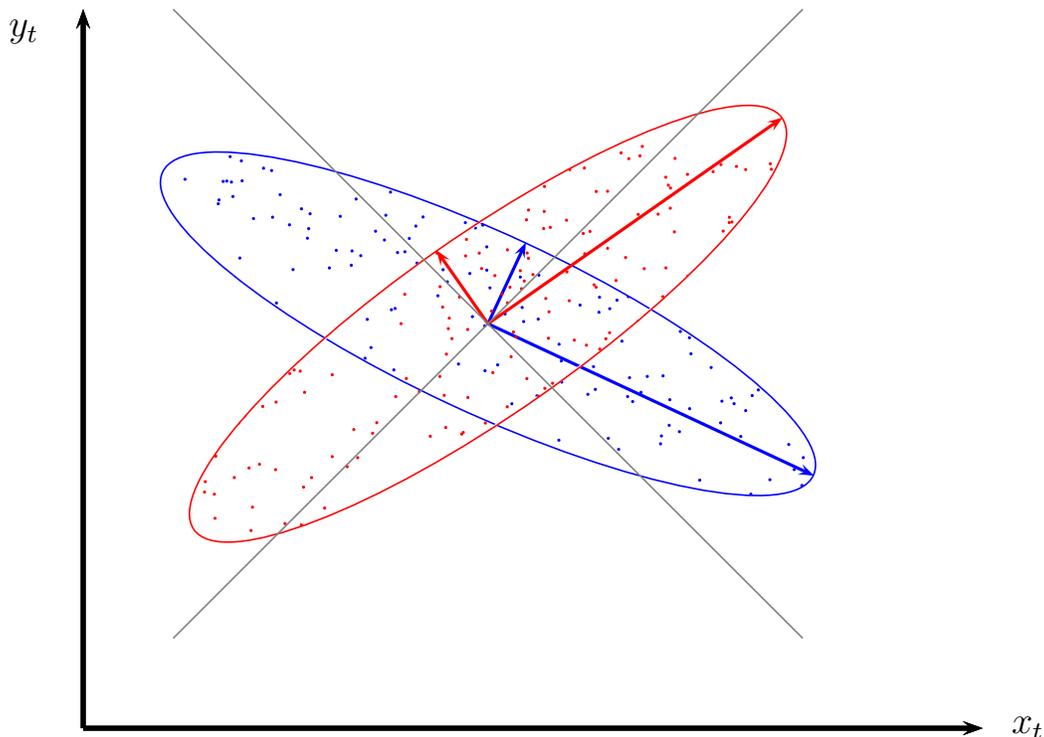


Figure A.2: Identification Through Heteroskedasticity

The intuition behind the identification through heteroskedasticity comes from the rotation of the residual ellipses. When the variances change, for the same parameters, the ellipses rotate. In Figure A.2, we show two cases: One when the shocks to the demand dominate (red), and one when the shocks to the supply dominate (blue). In particular, when the shocks to the demand dominate, then the ellipse approximates the supply curve. In fact, it is identical to the supply curve if the variance of the demand is infinite relative to the supply. Conversely, when the supply shocks are larger, then the long axis of the ellipse tilts toward the demand curve. It is this rotation of the ellipses when the relative variances shift that provides the identification.

It is instructive to re-state the underlying assumptions: structural shocks are uncorrelated

(quite uncontroversial) and parameters need to be stable across the regimes (so, this is a good technique to measure spillovers).

B Equivalent Formulation via Tensor Decomposition

We first show that the identification problem is equivalent to a tensor decomposition problem.

In previous sections, we identify multiple layers by matching the second moments

$$\Omega_h = (1 - \Gamma_1)^{-1} \Sigma_h (1 - \Gamma_1)^{-\top}$$

We define a new N -by- M matrix

$$A = (1 - \Gamma_1)^{-1}$$

Then we can write the moment matching equation as

$$\Omega_h = A \Sigma_h A^\top.$$

Because the matrix Σ_h is diagonal, we can further write

$$\Omega_h = \sum_{r=1}^N \vec{a}_r \sigma_{rh} \vec{a}_r^\top$$

where \vec{a}_r is the r th column of A and σ_{rh} is the r th diagonal entry of Σ_h . Because vector outer products can be written as tensor products, we can also write

$$\Omega_h = \sum_{r=1}^N (\vec{a}_r \otimes \vec{a}_r) \sigma_{rh}$$

where \otimes is the tensor product. Now if we stack all the second moments Ω_h along a third dimension, we obtain a N -by- N -by- H tensor $[\Omega_h]$ and it holds that

$$[\Omega_h] = \sum_{r=1}^N \vec{a}_r \otimes \vec{a}_r \otimes \vec{\sigma}_r \tag{36}$$

where $\vec{\sigma}_r = \left[\sigma_{r1} \quad \sigma_{r2} \quad \cdots \quad \sigma_{rH} \right]^\top$.

We can obtain estimates of Γ_1 by taking the rank- N tensor decomposition of $[\Omega_h]$.

According to the Kruskal's rank condition, if

$$\text{Krank}(a_r) + \text{Krank}(a_r) + \text{Krank}(\sigma_r) \geq 2R + 2$$

then the tensor decomposition problem has a unique solution.

C Proof of Lemma 2

Proof. First, constants are irrelevant, so we just have to prove

$$\Sigma_h^* = \arg \max_{\Sigma_h \in \mathbb{PD}} \frac{1}{|\Sigma_h|^{n_h/2}} |\mathbf{v}_{mh}|^{(n_h - N - 1)/2} e^{-(1/2) \text{tr}(\Sigma_h^{-1} \mathbf{v}_{mh})} = \frac{1}{n_h} \text{diag}(\mathbf{v}_{mh}) \quad (37)$$

Define D to be the new optimization variable with $D = \text{diag}(\mathbf{v}_{mh})/\Sigma_h$. Note by definition, D is a diagonal matrix. Now we insert $\Sigma_h = D^{-1} \times \text{diag}(\mathbf{v}_{mh})$ and change the optimization variable to D . Now the optimization problem becomes

$$D^* = \arg \max_{D \in \mathbb{PD}} \frac{1}{|D^{-1}|^{n_h/2} |\text{diag}(\mathbf{v}_{mh})|^{n_h/2}} |\mathbf{v}_{mh}|^{(n_h - N - 1)/2} e^{-(1/2) \text{tr}(D)} \quad (38)$$

and we just have to prove that the optimal D^* is a diagonal matrix with all entries equal to n_h . At this step, we have already separated \mathbf{v}_{mh} from this optimization problem. \mathbf{v}_{mh} is just a constant matrix that does not affect the optimal value of D^* . Leaving out all constants, and let d_i be the diagonal entries of D . Note that $|D| = \prod_i d_i$ and $\text{tr}(D) = \sum_i d_i$. Now the optimization problem becomes

$$d_i^* = \arg \max_{d_i > 0} \left(\prod_i d_i \right)^{n_h/2} e^{-\frac{1}{2} \sum_i d_i} \quad (39)$$

The solution of this optimization problem is given by $d_i^* = n_h$. □

D Credit Default Swap Data Details and Data Retrieval Process

To better assist understanding our results or reproducing our results, we list carefully the details of data we used in this paper.

We obtain the par mid spread of the credit default swap of the 10 target banks through the Thomson Reuters Eikon excel tool.

Step 1: CDS Ticker search

Due to the variety of CDS product exist in the market, we first need to decide which CDS we use. Thomson Reuters has decided a primary CDS for each bank through the ticker searching service. Input the command

```
=@TR("JPM;BAC;WFC;C;GS;MS;COF;HSBC.K;AXP;CSGN.S",  
"TR.CDSPrimaryCDSRic","CH=Fd RH=IN",B2)
```

into the Eikon excel tool, then we have the tickers of primary CDS products of the target companies. The obtained tickers are in Table A.1.

Most of those produces are 5 years CDS contracts traded in US. For HSBC and CSGN, they are 5 year CDS contracts traded in Europe.

Step 2: Retrieve par mid spread data for CDS

After obtaining those tickers, we request the actual par mid spread of those CDS products in the target date range. The command for retrieving spread data is

```
=@TR("JPM5YUSAX=R;BAC5YUSAX=R;WFC5YUSAX=R;C5YUSAX=R;  
GS5YUSAX=R;MS5YUSAX=R;COF5YUSAX=R;HSBA5YEUAM=R;AXP5YUSAX=R;
```

Bank RIC	Primary CDS RIC
JPM	JPM5YUSAX=R
BAC	BAC5YUSAX=R
WFC	WFC5YUSAX=R
C	C5YUSAX=R
GS	GS5YUSAX=R
MS	MS5YUSAX=R
COF	COF5YUSAX=R
HSBC.K	HSBA5YEUAM=R
AXP	AXP5YUSAX=R
CSGN.S	CSGN5YEUAM=R

Table A.1: RIC tickers for primary CDS products of target banks.

CSGN5YEUAM=R", "TR.PARMIDSPREAD", "Frq=D SDate=20090901 EDate=20170630
CH=Fd;IN RH=date",B2)

References

- Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015). Systemic risk and stability in financial networks. *The american economic review* 105(2), 564–608.
- Adrian, T. and M. K. Brunnermeier (2016). Covar. *American Economic Review* 106(7), 1705–1741. Predicting and measuring a financial institution’s contribution to systemic risk that internalizes externalities and avoids procyclicality.
- Allen, F., A. Babus, and E. Carletti (2012). Asset commonality, debt maturity and systemic risk. *Journal of Financial Economics* 104(3), 519–534.

- Allen, F. and D. Gale (2000). Financial contagion. *Journal of political economy* 108(1), 1–33.
- Altunbas, Y., L. Gambacorta, and D. Marques-Ibanez (2010). Bank risk and monetary policy. *Journal of Financial Stability* 6(3), 121–129.
- Anand, K., B. Craig, and G. Von Peter (2015). Filling in the blanks: Network structure and interbank contagion. *Quantitative Finance* 15(4), 625–636.
- Billio, M., M. Getmansky, A. W. Lo, and L. Pelizzon (2010). Econometric measures of systemic risk in the finance and insurance sectors. Technical report, National Bureau of Economic Research.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.
- Caballero, R. J. and A. Simsek (2013). Fire sales in a model of complexity. *The Journal of Finance* 68(6), 2549–2587.
- Cabrales, A., P. Gottardi, and F. Vega-Redondo (2014). Risk-sharing and contagion in networks.
- Caporin, M., L. Pelizzon, F. Ravazzolo, and R. Rigobon (2018). Measuring sovereign contagion in europe. *Journal of Financial Stability* 34, 150–181.
- Cavallo, A. and R. Rigobon (2016). The billion prices project: Using online prices for measurement and research. *Journal of Economic Perspectives* 30(2), 151–78.
- Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999). Monetary policy shocks: What have we learned and to what end? *Handbook of macroeconomics* 1, 65–148.
- Eichengreen, B., A. Mody, M. Nedeljkovic, and L. Sarno (2012). How the subprime crisis went global: evidence from bank credit default swap spreads. *Journal of International Money and Finance* 31(5), 1299–1318.

- Elliott, M., B. Golub, and M. O. Jackson (2014). Financial networks and contagion. *American Economic Review* 104(10), 3115–53.
- Elsinger, H., A. Lehar, and M. Summer (2013). Network models and systemic risk assessment. *Handbook on Systemic Risk* 1, 287–305.
- Forbes, K. J. and R. Rigobon (2002). No contagion, only interdependence: measuring stock market comovements. *The Journal of Finance* 57(5), 2223–2261.
- Fraley, C. and A. E. Raftery (1998). How many clusters? which clustering method? answers via model-based cluster analysis. *The computer journal* 41(8), 578–588.
- Freixas, X., B. M. Parigi, and J.-C. Rochet (2000). Systemic risk, interbank relations, and liquidity provision by the central bank. *Journal of money, credit and banking*, 611–638.
- Gai, P., A. Haldane, and S. Kapadia (2011). Complexity, concentration and contagion. *Journal of Monetary Economics* 58(5), 453–470.
- Gai, P. and S. Kapadia (2010). Contagion in financial networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 466(2120), 2401–2423.
- Girardi, G. and A. T. Ergün (2013). Systemic risk measurement: Multivariate garch estimation of covar. *Journal of Banking & Finance* 37(8), 3169–3180.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, 1029–1054.
- Junker, B. H. and F. Schreiber (2008). *Analysis of biological networks*, Volume 2. Wiley Online Library.
- Kalbaska, A. and M. Gałkowski (2012). Eurozone sovereign contagion: Evidence from the cds market (2005–2010). *Journal of Economic Behavior & Organization* 83(3), 657–673.

- Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika* 18(1), 39–43.
- Klein, R. and F. Vella (2009). Estimating the return to endogenous schooling decisions via conditional second moments. *Journal of Human Resources* 44(4), 1047–1065.
- Kritzman, M., Y. Li, S. Page, and R. Rigobon (2010). Principal components as a measure of systemic risk.
- Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business and Economic Statistics* 30, 67–80.
- Lewbel, A. (2018a). Identification and estimation using heteroscedasticity without instruments: The binary endogenous regressor case. *Economic Letters* 165, 10–12.
- Lewbel, A. (2018b). The identification zoo - meanings of identification in econometrics. *Journal of Economic Literature*.
- Little, R. J. and D. B. Rubin (2019). *Statistical analysis with missing data*, Volume 793. John Wiley & Sons.
- Merton, R. C., M. Billio, M. Getmansky, D. Gray, A. W. Lo, and L. Pelizzon (2013). On a new approach for analyzing and managing macrofinancial risks (corrected). *Financial Analysts Journal* 69(2), 22–33.
- Nakamura, E. and J. Steinsson (2018). Identification in macroeconomics. *NBER Working Paper* 23968.
- Onnela, J.-P., K. Kaski, and J. Kertész (2004). Clustering and information in correlation based financial networks. *The European Physical Journal B* 38(2), 353–362.
- Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics* 85(4), 777–792.

- Rigobon, R. and B. Sack (2003). Measuring the reaction of monetary policy to the stock market. *Quarterly Journal of Economics* 118(2), 639–669.
- Rigobon, R. and B. Sack (2004, February). The impact of monetary policy on asset prices. *Journal of Monetary Economics* 51, 1553–75.
- Rigobon, R. and B. Sack (2008). *Asset Prices and Monetary Policy*, Chapter Noisy Macroeconomic Announcements, Monetary Policy, and Asset Prices, pp. 335–370. National Bureau of Economic Research.
- Sentana, E. and G. Fiorentini (2001). Identification, estimation and testing of conditionally heteroskedastic factor models. *Journal of econometrics* 102(2), 143–164.
- Stegeman, A. and N. D. Sidiropoulos (2007). On kruskal’s uniqueness condition for the candecomp/parafac decomposition. *Linear Algebra and its applications* 420(2-3), 540–552.
- Temizsoy, A., G. Iori, and G. Montes-Rojas (2017). Network centrality and funding rates in the e-mid interbank market. *Journal of Financial Stability* 33, 346–365.
- Thurner, S. and S. Poledna (2013). Debtrank-transparency: Controlling systemic risk in financial networks. *Scientific reports* 3, 1888.
- Upper, C. (2011). Simulation methods to assess the danger of contagion in interbank markets. *Journal of Financial Stability* 7(3), 111–125.
- Wright, P. (1928). *The Tariff on Animal and Vegetable Oils*. New York: Macmillan Company.