

A System-wide Approach to Measure Connectivity in the Financial Sector*

Sumanta Basu¹, Sreyoshi Das², George Michailidis³ and Amiyatosh Purnanandam⁴

¹Department of Statistical Science, Cornell University

²Department of Economics, University of Michigan

³Department of Statistics, University of Florida

⁴Ross School of Business, University of Michigan

June 7, 2017

Abstract

We introduce and estimate a model that leverages a system-wide approach to identify systemically important financial institutions. Our Lasso penalized Vector Autoregressive (LVAR) model explicitly allows for the possibility of connectivity amongst all institutions under consideration: this is in sharp contrast with extant measures of systemic risk that, either explicitly or implicitly, estimate such connections using pair-wise relationships between institutions. Using simulations we show that our approach can provide considerable improvement over extant measures. We estimate our model for large financial institutions in the U.S. and show its usefulness in detecting systemically stressful periods and institutions.

*We are grateful to Viral Acharya, Rob Engle, Giampaolo Gabbi, Andreas Hagemann, Indrajit Mitra, Stefan Nagel, Uday Rajan, Toni Whited, and seminar participants at 2017 Conference on Banks, Systemic Risk, Measurement and Mitigation at Sapienza University, Cornell ORIE Colloquium, DSMM conference in San Francisco, FDIC, International Indian Statistical Association (IISA) Conference, International Society for Non-Parametric Statistics (ISNPS), Office of Financial Research (OFR), University of Michigan Finance brown bag, and Michigan Finance-Economics conference for helpful comments on the paper. All remaining errors are our own.

1 Introduction

There has been a growing interest in understanding and measuring *systemic risk*, largely driven by the events of the 2007-09 financial crisis. A number of such measures have been proposed, including conditional Value-at-Risk (CoVaR) ([Adrian and Brunnermeier \(2011\)](#)), CoRisk ([Chan-Lau et al. \(2009\)](#)), systemic expected shortfall (SES) ([Acharya et al. \(2012\)](#)) and, SRISK ([Brownlees and Engle \(2015\)](#)), to name a few. Another strand of literature proposes network connectivity of large financial institutions as a way to identify systemically important institutions based on the centrality of their role in an appropriately constructed network, e.g., network of the corresponding firms' stock returns (see [Billio et al. \(2012\)](#)).

There is broad agreement that systemic risk threatens the stability of the entire financial system and hence any associated risk measures should provide a systemwide perspective. However, there is relatively little theoretical guidance on how to measure systemic risk; therefore, understanding the econometric properties of the proposed measures of this risk becomes even more important.¹

Since systemic risk represents a property of the entire financial system, interconnectedness of the participants represents a key element. For example, highly interconnected institutions that are likely to fail pose a higher risk to the system due to the presence of multiple channels of transmission and contagion. Hence, at their core all proposed measures of systemic risk aim to reflect connectivity. For example, SES and CoVaR assess the association between a given financial institution's condition with that of the rest of the financial system and more broadly the economy. The larger the magnitude of these associations, the higher is the systemic risk of a given institution. Network based approaches directly aim to measure connectivity

¹Earlier theoretical work has mainly focused on banking and currency crises. These models provide extremely valuable insights into the microeconomic foundations of crisis, but they do not take us all the way to a measure of systemic risk that can be implemented in practice, e.g., see [Allen and Gale \(1998\)](#) and [Diamond and Dybvig \(1983\)](#) for some early contributions. Papers by [Battiston et al. \(2012\)](#), [Acemoglu et al. \(2015\)](#) and others have made considerable progress in the literature in recent years. These papers provide valuable insights into the shape of network structure, the mechanism of the shock propagation, and the resulting implications for the fragility of the system.

between financial institutions and subsequently derive summary network measures as proxies for systemic risk. However, extant measures of systemic risk often fall short of a true system-wide measure of connectivity. Our paper highlights this limitation of the current literature, and then proposes a solution.

We primarily focus on a network based approach akin to that adopted in [Billio et al. \(2012\)](#) to illustrate our key point. [Billio et al. \(2012\)](#) estimates a bivariate Granger causal association on the stock returns of large financial firms of the economy where firm A is said to be connected to firm B if A Granger-causes B , i.e., return of firm A at time t has additional predictive power in forecasting return of firm B at $t + 1$, over and above the lagged returns of firm B . While this is a useful starting point, such pairwise approach of learning network structures misses out on the system-wide connections. Specifically, a pairwise measure of statistical association between any two firms A and B gives the direct strength of connectivity between A and B , as well as indirect effects through all the other nodes in the network. As a result, a network based on such marginal effects of A on B does not pin down institutions that are key in propagating the risk in the system. We illustrate this issue in [Figure 1](#). The figure plots the true network structure for a three firm system. In this hypothetical system, there are 3 causal effects in the model: $B \rightarrow C$, $B \rightarrow A$, $C \rightarrow B$. However, due to indirect effect through B , there is additional (spurious) pairwise Granger causal effect $C \rightarrow A$. Measures such as SES and CoVaR partially mitigate this issue by considering statistical relationships between an institution and the system as a whole. However, even with these measures a similar concern arises since these models estimate the covariance of an institution with the rest of the system without conditioning it on all other participants. While our focus is on pairwise Granger causal network, we explore this issue for other measures further in [Section 4](#) of the paper.

Figure 1 around here

The key issue in the above example is that the pairwise metric does not take into consideration the effects of the third institution on the pair under consideration. Conceptually, the misspecification problem of the pairwise Granger causal effect is analogous to the well understood omitted variable bias in standard regression models. Statistically, the model parameters end up being inconsistently estimated, which in turn may lead to large economic costs; for example, a number of institutions that are not highly interconnected may end up being wrongly classified as interconnected under such an approach. Hence policy designs, such as linking a bank’s capital requirement based on their interconnectedness in the network, are likely to be problematic with such a structure. Similarly, such an approach may not be meaningful in identifying systemically important firms of the financial system.

One approach to correctly identify the interconnectedness structure of the system is to fit a VAR model that takes into consideration *all* interactions amongst the system’s components. This can be done, for example, by estimating the VAR model with all firms simultaneously, instead of a pair-wise approach. However, the number of parameters to be estimated even for the simplest lag-1 VAR model in this approach is quadratic in the number of institutions under consideration. For example, to estimate a full VAR(1) model for 100 financial institutions, we need over 10,000 time periods for estimation. In most practical applications, this seems infeasible. We suggest a statistical approach based on recent developments in higher dimensional statistics that overcomes this challenge.

We employ a regularized VAR model, using LASSO (Least Absolute Shrinkage and Selection Operator) techniques, that only focuses on estimating the *strongest* interconnections, while forcing weaker relationships to zero. The key statistical advantage of this approach is that we need significantly lower number of time points to estimate this model as compared to the classical estimation of the VAR model as long as the underlying network is approximately sparse. We provide an in-depth discussion of our statistical approach in Section 3 with additional technical details in the appendix. The method provides statistically consis-

tent estimates of the network’s interconnectedness, which constitutes the first step towards gaining insights about interconnectedness patterns during periods of financial calmness and juxtapose them to those during financial distress. As [Glasserman and Young \(2015\)](#) argue, the role of growing interconnectedness of the financial system is one of its least understood aspects.

On the methodological side, perhaps the closest to our system-wide approach is the one in [Demirer et al. \(2017\)](#), which builds upon the framework of [Diebold and Yilmaz \(2014\)](#). In a series of papers, Diebold and Yilmaz have advanced the literature by developing connectedness measures based on variance decomposition of banks. [Diebold and Yilmaz \(2014\)](#) used OLS estimation of VAR on a small number of firms (15 firms). To address the issue of high-dimensionality, [Demirer et al. \(2017\)](#) use penalized VAR estimation techniques, in particular an adaptive elastic net (AEnet) regularization term, to estimate the network of 150 global banks using daily data of their stock return volatility. The elastic net penalty has the attractive feature of producing less erratic estimates than lasso when the covariates in the regression model are highly correlated. However, we are not aware of any existing procedure to formally assess uncertainty of AEnet coefficients in high-dimensions. Our approach of debiased Lasso VAR comes with theoretically sound measures of statistical uncertainty, which provides greater confidence in inferences based on this approach. Our approaches complement each other to understand the underlying connectedness of financial institutions in high dimension.

After discussing the statistical underpinnings of our model, we conduct some simulation exercises to highlight the advantages of our measure over the existing ones. In our first simulation exercise, we simulate data on lead-lag relationship between financial institutions based on lag-1 VAR model. On the simulated data, we estimate connections based on both our model (which we refer to as Network Granger Causal model) and the bi-variate VAR model. Our model does considerably better in detecting the true network structure. We also

compute CoVaR and MES measures on this simulated data and show the improvement our model achieves.

The use of a first-order VAR model of stock returns may not be an innocuous assumption. In efficient markets, past stock returns of other financial institutions should not have any predictive power for explaining the return of any other institutions. Market inefficiency, slow diffusion of information and frictions such as short-selling restrictions can be a potential reasons for non-trivial dependence between the returns of different institutions over time. However, our paper does not rely on this specific form of interdependence across the institutions' returns. Using the idea of partial correlations, a system-wide approach can be taken to capture contemporaneous connectivity as well. Building on this idea, we next simulate a model that only has contemporaneous correlations across institutions' returns and contrast our approach with other models such as CoVaR and MES. Again, our model performs better in capturing the true connections. Given the very infrequent occurrence of actual systemic events that can be used to evaluate the relative performance of different models, our simulation exercise is especially important in establishing the usefulness of our approach.

In the final part of the paper, we estimate our model using the stock return data of three important sectors of the financial services industry, namely banks, broker-dealers and insurance companies. The financial institutions in these sectors are intricately related through both direct business relationships such as lending and borrowing, and through indirect relationship such as “spillover effects” through correlated trading or exposure to common assets.² Theoretical works such as [Allen and Gale \(2000\)](#), [Babus \(2013\)](#), [Acemoglu et al. \(2015\)](#) discuss direct linkage formation among firms through lending. On the other hand, some recent papers focus on connectedness via trading activities of firms. [Colla and Mele \(2010\)](#) discusses information network among investors while [Brunnermeier and Pedersen \(2009\)](#) shows how funding of traders with capital constraints and risk limits are affected by destabilizing

²[Billio et al. \(2012\)](#) discusses increased financial linkages across these types of institutions.

nature of margin-based trading . In this paper, we are agnostic about the reasons behind connections in the first place. Rather, our focus is on the measurement of the resulting interconnectedness.

Using our LASSO penalized lag-1 VAR model, we estimate the network structure over time, on a rolling basis, from year 1992 to 2012. We show that different measures of connectedness based on the number of firms connected to each other (degree) and the shortest path length from one firm to another in the network (closeness), exhibit sharp peaks just before important systemic events such as the dot-com related market crash in 2000 and the Lehman Brothers' failure in 2008.³ Thus our network is useful in providing information on the buildup of systemic risk in the financial system. Needless to say, with limited number of systemic events in the economy, we are unable to carry out any formal statistical test for the predictive power of our network. However, it is clear that our results line up well with identifiable periods of systemic risk in the economy.

Our network estimates allow us to detect institutions that are relatively more important in the network at any given point in time. Higher the degree of a firm, larger is the number of its immediate neighbors. Higher closeness, on the other hand, indicates how easily the firms can be accessed by other firms in the network. We find that AIG becomes one of the most important nodes in our network before and during the recent financial crisis. This finding is consistent with anecdotal evidence that highlights the central role of AIG in the economy during the 2007-2009 period. We provide the ranking of institutions at different points in time during our sample period, and these rankings can be useful inputs to policy decisions on the detection of systemically important institutions. Based on our estimates we find that banks that were closely linked to AIG experienced larger negative returns in the immediate aftermath of the failure of Lehman, providing confidence in our estimation method.

³There are several possible measures of centrality in networks such as degree, closeness, betweenness and eigenvector. Without a clear theoretical guidance, it is unclear which measure is most suited for systemic risk applications. Hence we present our results for two most widely used measures used often in studies of network model.

Our network estimate picks up strong relationships, which are likely to be more meaningful for policy decisions. We contrast our estimated network with that in [Billio et al. \(2012\)](#) which is significantly more dense. Said differently, in their pairwise network, institutions on average are connected to several others since the estimation does not parse out indirect relationships between institutions. Thus the *pairwise* Granger causal approach ends up with too many connections between institutions as opposed to our *network* Granger causal approach.

In summary, our paper contributes to the literature by estimating the network structure in a statistically principled way, specifically a measure of network that is consistent and mitigates, to a large extent, the omitted variable bias inherent in pairwise methods. Since any error in the misclassification of systemically important institution can be very costly for the economy, our paper provides a considerable improvement in designing and implementing efficient macro-prudential regulations. Our approach can be useful in a number of different settings where researchers are likely to be interested in both direct and indirect linkages between several firms in a network. For example, our methodology can be useful in detecting supplier-customer stock return relationships for a large number of firms. Similarly, our method can be helpful in estimating the effect of common owners or board members on firm policies. Our paper provides self-contained guidance on estimating a true Network Granger Causal model for applied researchers in different areas of finance and economics.

Section 2 expands on the biases created by pairwise approach and highlight the limitations of extant measures of systemic risk. Section 3 proposes our Lasso penalized VAR measure. In Section 4, we show the usefulness of our measure, compared to existing measures, on simulated data sets. Section 5 presents the estimation result with actual data for 75 largest financial institutions of the U.S. Section 6 concludes.

2 Pair-wise versus system-wide approaches

We elaborate on the problem statement and potential biases created by extant measures in this section. Throughout this paper, we use $A_{i\cdot}$ and $A_{\cdot j}$ to denote the i^{th} row and j^{th} column of a matrix A , respectively. We also use the standard notations for norms of a p -dimensional vector $\|v\|_\infty = \max_{j=1,\dots,p} |v_j|$, $\|v\|_1 = \sum_{j=1}^p |v_j|$. For a $m \times n$ matrix A , we denote its Frobenius norm as $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2}$.

Consider a network of 15 institutions with 5 hubs, each with one central firm. In each hub, the middle firm is central and propagates shocks to other firms. Firms on the periphery, on the other hand, do not propagate any shocks to other firms (see Figure 2). Such a dynamics can be modeled by assuming a data generating process in which the middle firm's return in period t affects the returns of both firms on the periphery in period $t + 1$. We capture this idea by simulating data as per the dynamics below:

$$\begin{aligned} R_2^{t+1} &= 0.6 * R_2^t + \epsilon_2^{t+1} \\ R_j^{t+1} &= 0.6 * R_j^t + 0.4 * R_2^t + \epsilon_j^{t+1}, \quad j = 1, 3 \\ \epsilon_j^t &\stackrel{i.i.d.}{\sim} N(0, 1) \end{aligned}$$

We simulate this model with 500 independent draws. Based on the simulated data, we fit a pair-wise VAR model in line with [Billio et al. \(2012\)](#). For each pair of firms, we estimate the lead-lag relationship between their returns using an OLS model. It is worth emphasizing that in this approach each estimation exercise ignores the effect of all other firm's returns on the returns of the pair under consideration.

The true network as well as the estimated pairwise Granger causal network are depicted in Figure 2. The estimated network structure detects significant relationships between the adjacent peripheral firms in each hub in addition to the relationship between the central firm and the rest. Thus, the estimated network structure provides an incorrect picture of the true

network. The reason is simple. The pairwise model ignores the fact that returns of both peripheral-firms are driven solely by the central firm. Ignoring the effect of the central firm’s returns while estimating correlations between the returns of firms on the periphery, leads to false positives connections. In other words, the pairwise model ignores the conditional independence in returns of firms on the edges, conditional on the central firm’s returns.

Figure 2 around here

After estimating the network structure, researchers often use statistics such as the degree of a node (i.e., number of important connections a particular node has) as a measure of the importance of the node in the network. In the above example, the pair-wise model estimates a degree of 2 for both the adjacent nodes, as compared to its true degree of 1. Thus, the use of this network structure can lead to misleading inferences. An immediate solution to this problem is to estimate the VAR model simultaneously with all firms in the system. However, such an approach is not feasible with standard techniques due to data limitations. For example, if we have 100 large institutions in the system, then a VAR(1) model needs to estimate 10,000 (100×100) parameters! This is often impossible due to relatively fewer samples and regime changes in the underlying system. Our proposed method overcomes this problem of dimensionality and allows us to estimate the model structure in a very wide range of situations.

While it is relatively straightforward to see the difference between a pair-wise and a system-wide approach in the case of VAR model discussed above, even other models such as MES and CoVaR face this challenge. For example, consider CoVaR. It measures the value-at-risk of the entire financial system conditional on the value-at-risk of a given institution. For firm i value-at-risk at a confidence level q represents the extent of losses that will not be exceeded with a probability greater than q . CoVaR measures the probability that the entire system is in distress (i.e., the return of the entire system is below some threshold)

conditional of bank i hitting its VaR limit. A related measure, $\Delta CoVaR$ measures the difference in CoVaR when the firm i is in its median state compared to the same firm being in a distressful state. More formally:

$$\begin{aligned}\mathbb{P}(R^i \leq VaR_q^i) &= q \\ \mathbb{P}(R^{system} \leq CoVaR_q^{system|R^i=VaR_q^i} \mid R^i = VaR_q^i) &= q \\ \Delta CoVaR_q^{system|i} &= CoVaR_q^{system|R^i=VaR_q^i} - CoVaR_q^{system|R^i=VaR_{50\%}^i}\end{aligned}$$

As can be seen from the above discussion, CoVaR only conditions on the distress of one financial institution at a time. Thus it misses out the effect of all other firm's returns on the system, and just like pairwise VaR it attributes all the indirect linkages as a direct linkage between firm i and the system. For example, assume that JP Morgan Chase is the most vital bank in the system in the sense that its distress leads to distress of the entire system as well as a specific bank, Citi Bank. Even if Citi Bank, in our example, is systemically unimportant, CoVaR is likely to pick it up as an important systemic bank. The underlying issue is the same: CoVaR of Citi Bank does not consider the indirect effect of JP Morgan Chase.

MES, defined as the expected return of firm i when the system is at its lower tail, provides some improvement by conditioning on the system as a whole. However, it still computes a pairwise measure. Formally, MES is defined as follows (we take negative of the expected return so that the measure increases in systemic risk:

$$MES = -E(R^i \mid R^{system} \leq R_q^{system})$$

If firm j is the central node that affects both firm i and the system as a whole, then we will find a significant relationship between firm i 's returns and the system as a whole in a model that excludes firm j from it.

Ideally, we want to compute the CoVaR and MES measures of an institution after conditioning on the effect of *all* other firms in the system. For example, the notions of CoVaR and MES can be generalized in a system-wide fashion by including the omitted firms in the conditioning set as follows:

$$\mathbb{P}\left(R^{system} \leq CoVaR^{system|R_q^i} \mid R^i = VaR_q^i, R^j = VaR_{50\%}^j, R^k = VaR_{50\%}^k, \dots\right) = q$$

However, estimation of such measures will face similar statistical challenges due to over-parameterization, which will require additional econometric considerations. For expositional simplicity, we first discuss our modified VAR model and later return to a discussion of these other measures of systemic risk.

3 Model and Method Description

To overcome the limitations presented above, we adopt an approach that has both sound statistical and economic properties. At a very broad level, our statistical approach forces weak relationships among institutions in the network to zero, allowing us to take a true system-wide approach in estimating the model with limited data. In economic terms, this approach is both sensible and useful for policy designs. As we discuss in detail later in the paper, numerous studies have shown that financial institutions form trading or counter-party relationships with only a handful of other institutions. Hence, the assumption of sparsity that underlies our estimation is reasonable in our context. Second, when regulator have limited resources, it is advantageous to focus on stronger connections in the network. Our model allows us to do this.

We estimate the network connectivity among p institutions based on a p -dimensional VAR(1) model of stock returns (after suitable transformation to reduce nonstationarity). The transition matrix of this model reflects strengths of lead-lag relationships between returns

of two institutions, *conditional* on the returns of all the other ones in the sample. To ensure consistent estimation of our model with limited sample size ($n \ll p$), we assume *sparsity* of the true underlying financial network, and motivate this assumption by pointing to empirical evidence in section 3.1. The posited sparsity assumption implies that a large number of elements in the transition matrix are zero, and hence fewer parameters need to be estimated from the available data.

As we describe in section 3.2, such a sparse VAR model can be consistently estimated using a penalized (Lasso) regression framework with small sample size. However, using the sparse Lasso VAR estimates directly to assess network connectivity faces two issues - (i) this estimate does not come with associated uncertainty measures (e.g. confidence intervals), and (ii) sparsity of the network relies on a non-obvious choice of a tuning parameter. Our proposed debiased Lasso VAR estimates mitigate both issues by allowing us to formally test for Granger causality, and form a network with statistically significant relationships as edges. The problem of selecting the critical tuning parameter then reduces to the familiar specification of significance level in hypothesis testing. By varying the level of significance (e.g., 1%, 5%, 10%), we can change the levels of sparsity in our estimated networks. Given that we carry out *simultaneously* p^2 tests (one for each debiased edge in the network), we need to correct for the well known multiple comparisons problem. After doing so, the resulting significant edges are used to construct the Granger causal network of interest, which is summarized by using various standard network measures such as degree and closeness to detect highly connected and thus systemically important institutions.

3.1 VAR models and network Granger causality

We model the process of stock returns of p firms $X^t = (X_1^t, \dots, X_p^t)'$ using a p -dimensional Gaussian VAR(1) model.⁴⁵

$$X^t = AX^{t-1} + \varepsilon^t, \quad \varepsilon^t \sim N(0, \Sigma_\varepsilon), \quad \Sigma_\varepsilon = \text{diag}(\sigma_1^2, \dots, \sigma_p^2) \quad (1)$$

In this model, the $p \times p$ transition matrix A can be viewed as a weighted, directed network $G = (V, E)$ amongst financial institutions, with the set of nodes $V = \{1, 2, \dots, p\}$ and the set of edges $E = \{(i, j) : A_{ij} \neq 0\}$. The weight of an edge (i, j) , denoted by $|A_{ij}|$ measures the strength of connections. For ease of presentation, we work with the undirected, unweighted *skeleton* of the network G , denoted by $\mathcal{S}(G)$, where there is an edge $i - j$ between institutions i and j if $\max\{|A_{ij}|, |A_{ji}|\} \neq 0$.

The VAR model allows one to generalize pairwise Granger causality towards Granger's original definition of causality ([Granger, 1969, 1980](#)). A series X_1 Granger-causes another series X_2 if

$$\sigma^2(X_2^{t+1}|\mathcal{I}(t)) < \sigma^2(X_2^t|\mathcal{I}(t) - \mathcal{I}_{X_1}(t)),$$

where $\sigma^2(A|B)$ denotes the variance of the prediction error, when predicting A using the best linear predictor constructed from information set B , and $\mathcal{I}(t)$ captures *all available information in the universe* up to time t . For pairwise Granger causality analysis, the information set $\mathcal{I}(t)$ is restricted to the information in the two series X_1 and X_2 up to time t . A joint VAR model allows one to expand the set $\mathcal{I}(t)$ to include information contained in *all* p series X_1, X_2, \dots, X_p . Conditioning on this larger information set is the *central theme* of our system-wide approach, as we also emphasize in section [4](#) in the context of

⁴We chose the VAR order to be 1 for ease of exposition. Networks can be estimated by combining information of transition matrices from different lags in a VAR(d) models ([Basu et al., 2015](#)).

⁵In the data analysis, we use residuals of a GARCH model fitted to the univariate series of returns. Other suitable transformations can be applied to adjust for non-Gaussian heteroskedasticity in data. Our statistical methodology is general and can be applied on other characteristics of the institutions; e.g., volatilities (after log transformation), leverage ratios etc.

contemporaneous dependence. To emphasize its importance in constructing the network representation of the system, we refer to this notion as *network Granger causality* (Basu et al., 2015). The entries of the VAR transition matrix A capture the network of Granger causal relationships with respect to this larger information set.

The choice of the information set $\mathcal{I}(t)$ is an important consideration in multivariate Granger causality analysis, well-known in the time series and econometrics literature. Failure to include relevant information outside the two series under investigation often results in a spurious Granger causal relationship among the observed series, which essentially captures indirect effects coming via the unobserved omitted variables. Another view of using VAR models to estimate network connectivity among stock returns of p financial institutions is also related to the general theory of graphical models popular in statistics and machine learning (Wainwright and Jordan, 2008), since the transition matrix A of a Gaussian VAR(1) model with diagonal Σ_ε determines the adjacency matrix of the Directed Acyclic Graph (DAG) which characterizes the conditional independence relationships among firm characteristics in their joint distribution (Eichler, 2012).

Sparsity of Financial Networks. We assume the network is *sparse*, i.e., the number of edges present in the network ($s := \|A\|_0 = \sum_{i,j=1}^p \mathbb{1}[A_{ij} \neq 0]$) is very small compared to the total number of possible edges p^2 . For example, in a network with 100 institutions, we have 10,000 parameters in a first-order VAR model. We require the true number of interconnections in a 100 institution network to be much smaller than 10,000. This is a reasonable assumption for our application. First, each financial institution is unlikely to form strong relationships with all others in the sample simply due to the costs involved in starting and maintaining such relationships. This is especially true in information-sensitive markets involving non-trivial search costs (e.g., see discussion in Gofman (2016)), where institutions often rely on repeat transactions with a relatively smaller set of institutions. Empirical evidence from inter-bank relationships provide strong support for this assertion.

For example, [Soramäki et al. \(2007\)](#) analyze daily networks in the first quarter of 2004 using interbank payments transferred between commercial banks over the Fedwire. Based on actual data they find few highly connected banks and the great majority of banks having few counterparties. That the degree distribution (number of counterparties for each bank) roughly follows the power law distribution with few core banks and several small banks is reported for several interbank market across the world (e.g [Bech and Atalay \(2010\)](#), [Boss et al. \(2004\)](#), [Iori et al. \(2008\)](#), [Craig and Von Peter \(2014\)](#), [Blasques et al. \(2015\)](#)). If the underlying network structure is not very sparse but has a few strong and many weak relationships, our model will be able to detect strong relationships forcing the weaker ones to zero ([Bühlmann and van de Geer, 2015](#); [van de Geer and Stucky, 2016](#)). Again, from an economic viewpoint this is a reasonable property of our model since we are mainly interested in strong connectivity relationships to begin with.

In the next section, we provide a short overview of the existing machinery for estimating large VAR models and describe our method, which builds upon a bias-corrected Lasso procedure originally proposed in [Javanmard and Montanari \(2014\)](#) and extended in this paper for time dependent data.

3.2 Estimating large VAR models

Historically, the most common method for estimating the transition matrix A is on an equation-by-equation basis, by ordinary least squares (OLS) regression of X_i^t on $X_1^{t-1}, X_2^{t-1}, \dots, X_p^{t-1}$, for $i = 1, \dots, p$. However, the OLS estimate is ill-defined when the number of predictors is larger than the number of observations i.e, $p > n$. A VAR(1) model with p variables requires estimation of p^2 free parameters, which in turn requires at least $O(p^2)$ samples for meaningful estimation. Therefore, without imposing any additional restrictions on the parameters, it is not possible to estimate such a VAR model.

Penalized VAR estimation with Lasso. Recent advances in high-dimensional statis-

tics have established that it is possible to estimate a VAR model with relatively few samples, if the underlying transition matrix is appropriately sparse. In the context of regression problems, several sparsity-inducing methods have been introduced, arguably the most popular among them being the Least Absolute Shrinkage and Selection Operator (Lasso) (Tibshirani, 1996). Recently Basu and Michailidis (2015) have established that the Lasso VAR estimates are consistent in high-dimensional settings, i.e., assuming p grows with n , possibly at a faster rate. More precisely, if the number of non-zero elements of the transition matrix $s \ll p^2$, then much fewer sample (than what is required for OLS estimation) is sufficient for consistent estimation of A . An element of the estimated sparse transition matrix, $\hat{A}_{i,j}$, can then be used to denote the edge strength between nodes i and j . Barigozzi and Brownlees (2013) proposed a similar Lasso based VAR estimation procedure for network estimation.

The equation-by-equation estimate of Lasso VAR is defined as

$$\hat{A}_{i\cdot} = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \frac{1}{n} \|Y_{\cdot i} - \mathbf{X}\beta\|^2 + \lambda_i \|\beta\|_1, \quad i = 1, \dots, p$$

Here $\|\beta\|_1 := \sum_{j=1}^p |\beta_j|$ is the ℓ_1 -norm penalty, which encourages sparsity in the solution by shrinking smaller coordinates to zero. Using \hat{A} to construct a sparse estimate of network faces the following two issues. The first is the choice of tuning parameters. Lasso VAR minimizes, for every $i = 1, \dots, p$, a residual sum of squared errors (RSS) plus λ_i times the sum of $\sum_{j=1}^p |A_{ij}|$, where λ_i is a tuning parameter controlling the degree of sparsity in $\hat{A}_{i\cdot}$. Essentially, Lasso augments an OLS minimization with a penalty term that penalizes non-zero coefficients, and higher values of λ_i encourage sparser estimates. Similar to OLS estimates, Lasso penalized least squares estimates of VAR can be obtained by p separate Lasso regressions and it entails selection of p tuning parameters λ_i 's. With limited sample sizes, cross-validation and other data-driven strategies of selecting λ_i 's fail to provide robust guidelines for such choices. The second issue is that Lasso estimates, unlike OLS, do not come with an associated measure of uncertainty. The main reason is that the use of penalty

introduces bias in the estimate, which is not easy to quantify in closed form. As a result, developing central limit theorems and associated inference machinery (p-values, confidence intervals) has been a challenge in the practice of Lasso.

Statistical Inference with debiased Lasso VAR. To address these two technical issues, we build upon a recently proposed method of debiasing Lasso estimates (Javanmard and Montanari, 2014). It provides a substantial correction to the bias of Lasso and in turn allows assessing uncertainty of the estimated network edges. Also, in order to reduce the degree of subjectivity in tuning parameter selection, this method uses a theory-driven choice of λ_i 's obtained using the strategy of scaled Lasso (Sun and Zhang, 2012). In this work, we extend this method to time dependent data settings.

We start by elaborating on the second point. The theoretical literature of Lasso suggests that Lasso estimates are consistent for a choice of λ_i which scales with the noise standard deviation σ_i , which is unknown in reality (Bühlmann and Van De Geer, 2011). The scaled Lasso procedure (Sun and Zhang, 2012) suggests a work-around by minimizing a squared error loss function penalized for both large $|A_{ij}|$ and σ_i , and provides an estimate $\hat{\sigma}_i$. Debiased Lasso VAR starts by obtaining equation-by-equation Lasso estimates $\hat{A}_{i\cdot}$, obtained by plugging-in $\hat{\sigma}_i$ in the theory-driven choice of tuning parameters λ_i . In the next step, we conduct a bias correction of \hat{A} using $\tilde{A}_{i\cdot} = \hat{A}_{i\cdot} + \frac{1}{n}M\mathbf{X}'(Y_{i\cdot} - \mathbf{X}\hat{A}_{i\cdot})$, where the matrix M (see Appendix A for complete description) is a pseudo-inverse of the sample covariance matrix ⁶. We show that the bias corrected estimates $\tilde{A}_{i\cdot}$ have asymptotically zero mean, finite variance and use the formula $P_{ij} = 2[1 - \Phi(\frac{\sqrt{n}|\tilde{A}_{ij}|}{\hat{\sigma}_i[M\hat{\Sigma}_X M']_{jj}})]$ (Javanmard and Montanari, 2014) to calculate p-values for the hypothesis tests of interest $H_0 : A_{ij} = 0$ vs. $H_A : A_{ij} \neq 0$.

An estimate \hat{A} of the VAR transition matrix can be used to construct a weighted, directed network. An edge is present from node j to node i if A_{ij} is significant at a pre-specified threshold $\alpha > 0$.

⁶Such a bias correction is in the spirit of a single step of Newton-Raphson or Fisher scoring algorithms in classical statistics, with suitable modifications to allow for lack of regularity in high-dimension

The choice of the significance threshold α is important, since constructing the directed network amounts to performing $p(p-1)$ hypothesis tests. For large p , this requires a correction for multiple testing to avoid the problem of high false positives. The standard Bonferroni criterion for controlling the family-wise error rate (FWER) is the most conservative one, but it suffers from low power. We use a less stringent criterion of multiple testing, proposed in [Benjamini and Hochberg \(1995\)](#), to control the False Discovery Rate (FDR). FDR is the expected proportion of falsely rejected hypotheses over the total number of rejected hypotheses. Thus, a 20% false discovery rate would imply that, on average, 1 out of 5 selected edges is falsely detected. The procedure was originally proposed for independent test statistics, and its validity for test statistics with positive regression dependency was established in [Benjamini and Yekutieli \(2001\)](#).

Network construction and Centrality with VAR estimates. The topology of a weighted, directed network with edges significant at a level α (after correcting for multiple testing), or its undirected, unweighted skeleton $\mathcal{S}(G)$, can be explored by standard visualization software or by calculating network centrality measures. In [Section 5](#), we have used two centrality measures, degree and closeness, of $\mathcal{S}(G)$ to identify central institutions and monitor the degree of connectedness in different components of the US financial sector (e.g. banks, insurance companies and broker dealers). We provide more details on the centrality measures in [Section 5](#). However, before applying our model to the data, in the next section we estimate our model on simulated data and contrast it with other measures of systemic risk. This is an important exercise to gain insights on the performance of these measures in stylized settings. Due to the limited number of systemic events, it is almost impossible to empirically validate these measures. The next best alternative is to study the efficacy of these measures on simulated data, which is presented next.

4 Simulation Results

In this section, we conduct some simulation experiments to highlight the benefit of our approach over existing measures. In Section 4.1, we focus on differences between pairwise VAR and Lasso-VAR in estimating the network structure. In Section 4.2, we undertake more extensive simulations to show that the limitations of pairwise approach apply more broadly to other extant measures of systemic risk as well, including MES and CoVaR. The key intuition is similar: these measures estimate the association between the system and a firm, one firm at a time, which stops short of a true system-wide approach. Finally, in Section 4.3, we change our data generating process from VAR to allow for contemporaneous correlation structure, and argue that an appropriate statistical method for measuring partial correlation is more suitable than extant pairwise approaches like MES and CoVaR. Thus, our results are not specific to a given set of economic assumptions that result in a lead-lag relationship in the returns of financial institutions. Rather, our model can be used to refine a whole range of statistical estimation in this area.

4.1 Granger causality and Network Granger causality

Figure 3 around here

Figure 4 around here

In this section, we conduct a small numerical experiment to demonstrate the advantage of network Granger causal estimates using debiased Lasso VAR (referred to as “LVAR”) over pairwise Granger causal estimates with standard pairwise VAR models. We simulate 100 datasets, each of size $n = 500$, from a 15-dimensional Gaussian VAR(1) model (1) (i.e., $p = 15$). The transition matrix A has the following structure: $A_{j-1,j} = A_{j+1,j} = 0.6$, for $j = 2, 5, 8, 11, 14$; $A_{ii} = 0.8$ for $i = 1, \dots, 5$; and $A_{ij} = 0$ otherwise. The noise variance is

set to $\sigma^2 = 1$. This model captures a directed network with five hubs of size 3 each, with 1 central node affecting 2 neighbors. Thus, in this hypothetical network, only 5 of the 15 firms are systemically important.

The average performances of LVAR and pairwise VAR estimates in recovering the true network skeleton are displayed in Figures 3 and 4. The left Panel of Figure 3, shows the skeleton of the true network, with 5 non-overlapping hubs. In the right panel, we plot the “average” network estimated by Lasso VAR (BH correction used with a threshold 20%), where the grayscale of each edge represents the proportion of times (out of 100 datasets) that edge was significant. Similarly, the middle panel shows that “average” network estimated by pairwise VAR (significance threshold set at 5%). The results in the middle panel clearly show that the pairwise VAR model detects too-many connections compared to the true network. In this model, the edges are significant either through direct connectivity or through indirect effects of connectivity emanating from a common neighbor. For instance, the estimated pairwise VAR networks select edges between firms 1 and 3, which share a common neighbor 2. Networks estimated using LVAR do not show any such patterns, and thus they are closer to the true network.

Figure 4 illustrates that this pattern of selecting high false positive is stable across datasets, and is not an artifact of a few simulated runs. The number of edges selected by pairwise VAR (blue) and debiased Lasso VAR (red) on each of the 100 estimated networks are plotted. The figure clearly shows that pairwise VAR method selects at least 15 edges in all the datasets, while LVAR selects only 10 – 15 edges. This is expected since LVAR takes into consideration the partial dependence between firms while pairwise VAR captures the marginal dependence.

The above results demonstrate the potential limitation of pairwise approach in identifying systemically important institutions. As shown in Figure 3, the pairwise VAR approach identifies all three firms 1, 2 and 3 as central, while in truth only firm 2 is central to the economy.

Such misclassification of systemically important institutions can have crucial implications for the detection of risk and a range of policies that depend on systemic risk.

4.2 Comparison with MES and CoVaR

In the next simulation experiment, we simulate firm returns from a Gaussian VAR(1) model, where the transition matrix corresponds to the adjacency matrix of a network described in Figure 5. To enrich our experiment we now add five firms in the network that are isolated: i.e., not at all connected to the system. The network has $p = 20$ firms, of which $\{1, 2, 3, 4, 5\}$ are isolated, i.e., they are not affected by shocks on the other firms. There are 3 central/risky firms in this universe $\{8, 13, 18\}$, each of which transmits shock to four other firms. Based on $n = 500$ returns simulated from this model, we calculate MES, CoVaR and degrees of different firms in pairwise and Lasso VAR networks. The results are reported in Figure 6. The top panel shows that except the five isolated firms, all the firms are deemed as risky in MES, CoVaR and pairwise VAR. Moreover, with a slight exception to MES, the true central firms $\{8, 13, 18\}$ are hard to detect among the 15 connected firms in this universe. In contrast, Lasso VAR captures the true network structure and ranks the three central firms as highly risky compared to the other 15 firms.

Figure 5 around here

Figure 6 around here

4.3 Contemporaneous Correlation Structure

In this section, we show that the importance of delineating direct vs. indirect associations amongst firms is prominent even when the connection among firm returns is contemporaneous instead of intertemporal (i.e., the lead-lag relationship). This exercise is also useful

in stressing the point that our approach does not depend on whether one takes a strong view on the informational efficiency of the markets or not. To demonstrate this, we generate firm returns from a multivariate Gaussian distribution, where the partial correlation among firms encode the conditional relationship described in the network 5. We simulate $T = 500$ returns from this distribution, and report the estimated MES, CoVaR and firm degrees in pairwise VAR in Figure 7. We simulate data from a $p = 20$ -dimensional Gaussian distribution with correlated components, where the conditional independence among the nodes follows the network structure in Figure 5. In particular, we construct a matrix Θ as follows: for each $j \in \{8, 13, 18\}$, we set $\Theta_{ij} = \Theta_{ji} = 0.5$, where $j \in \{i - 2, i - 1, i + 1, i + 2\}$. For every other pairs $\{i, j\}$, $\Theta_{ij} = 0$. To ensure the positive definiteness preserving the network structure, the inverse covariance matrix is generated as $\Theta + (|\lambda_{\min}| + 0.2)I$, where λ_{\min} is the minimum eigenvalue of Θ . The inverse covariance matrix is contain information on the partial correlations and is routinely used in Gaussian graphical modeling (see Appendix B for more details).

Figure 7 around here

Since MES and ΔCoVaR measure contemporaneous association between each firm and the system, these measures are highest for the central firms $\{8, 13, 18\}$, however the firms affected by these three central firms are also close. Since there is no intertemporal dependence, pairwise VAR does not detect any Granger causal relationship as expected. The same holds for LVAR. However, we show that a bias corrected version of Graphical Lasso (Friedman et al. (2008); Jankova et al. (2015)), a method for calculating partial correlation in high-dimension, correctly detects the central firms as more risky than the other 15 firms. Similar to the network Granger causality, partial correlation measures the correlation between each pair of firm returns, conditioning on the returns of all the other firms under consideration. The firm pairs (i, j) with strong partial correlation relationships can be recovered using node-

wise regression, i.e., regressing R_i on the returns of all the other firms, and looking at the coefficient of R_j (Meinshausen and Buhlmann, 2006). An alternative approach utilizes the fact that the partial correlation structures among the components of a multivariate Gaussian random variable $\mathbf{X} \sim N(0, \Sigma)$ can be obtained from the inverse covariance matrix $\Theta = \Sigma^{-1}$. Based on these connections, the graphical Lasso (Glasso) estimates Θ use a Lasso penalized maximum likelihood method to estimate Θ :

$$\hat{\Theta} := \underset{\Theta \succeq 0}{\operatorname{argmax}} \quad \log \det \Theta - \operatorname{tr}(S\Theta) - \lambda \sum_{i \neq j} |\Theta_{ij}|,$$

where S is the sample covariance matrix, λ is a tuning parameter controlling the degree of sparsity and $\succeq 0$ denotes that the function is maximized over non-negative definite matrices. Both of these approaches are commonly used in the statistics literature to build partial correlation networks from high-dimensional data sets. In recent work, Brownlees et al. (2015) used Glasso based estimates to construct a network amongst firms based on their realized volatilities. We use a bias corrected version of Graphical Lasso, recently proposed in Jankova et al. (2015), which provides a measure of uncertainty of the edge weights. We provide further details on the estimation exercise in Appendix B.

Overall, these simulation results establish the usefulness of our approach in estimating the true network structure. We now proceed with the estimation exercise with actual data on stock returns of large financial firms in the U.S.

5 Empirical Application

We estimate the LVAR model to detect the Network Ganger Causality structure on a subset of the data set used by Billio et al. (2012).

5.1 Data Description and Summary Statistics

We use monthly returns data from January, 1990 to December, 2012 for three financial sectors, namely banks (BA), primary broker/dealers (PB) and insurance companies (INS) available at the University of Chicago’s Center for Research in Security Prices (CRSP) and retrieved from Wharton Research Data Service (WRDS). We denote firms with Standard Industrial Classification (SIC) from 6000 to 6199 as banks, from 6200 to 6299 as broker/dealers and from 6300 to 6499 as insurance companies. We divide the data into 3-year rolling windows, retaining only the institutions that have complete data in that window. To create our final data set, we keep the top 25 institutions in terms of market capitalization in each sector in every time window.

Our final sample covers 225 different institutions spanned over 23 years period. Figures 8 and 9 show the mean and standard deviation (in %) of monthly stock returns across different sectors in each 3-year rolling window. As expected, the average returns are significantly lower and the standard deviations significantly higher during the 2007-2009 period, compared to any other period in our sample. Another period of significant volatility in the sample is the Russian financial crisis in 1998. Also, looking across sectors, all three experienced stress during the 2007-2009 crisis, whereas around 1998 it was predominantly the broker-dealers (PB) who exhibit high volatility.

Figure 8 around here

Figure 9 around here

5.2 Network estimation and Measures of connectedness

In order to estimate our network, we consider the Generalized AutoRegressive Conditional Heteroscedasticity (GARCH(1,1)) as our baseline model for returns of individual firms. This

allows us to remove any effect of heteroskedasticity from contaminating our LVAR measure. Since accurate estimation of Granger causal relationships relies crucially on the stationarity of the underlying data generating process (Lütkepohl, 2005), raw returns with high heteroskedasticity are not appropriate for constructing Granger causal networks. The approach of using GARCH fitted residuals was also adopted in Billio et al. (2012). Multivariate GARCH models like Dynamic conditional correlation (DCC) (Engle, 2002) were not applicable due to high-dimensionality in our data set with ($n = 36$ time points, $p = 75$ firms), but are potential alternatives to univariate GARCH ones, if the sample size is sufficiently large. We note that by denoting an institution's return at time t as $R_{i,t}$, a GARCH(1,1) specification implies the following.

$$\begin{aligned} R_{i,t} &= \mu_i + \sigma_{i,t}\epsilon_{i,t}, & \epsilon_{i,t} &\sim N(0,1) \\ \sigma_{i,t}^2 &= \alpha_0 + \alpha_1\epsilon_{i,t-1}^2 + \beta_1\sigma_{i,t-1}^2 \end{aligned} \tag{2}$$

We estimate the GARCH(1,1) parameters μ_i , $\sigma_{i,t}$, α_0 , α_1 and β_1 for each of the 75 institutions in every time window. Then we fit our debiased Lasso VAR (LVAR) model to the estimated Garch fitted returns, namely $\hat{\epsilon}_{i,t} = \frac{R_{i,t} - \hat{\mu}_i}{\hat{\sigma}_{i,t}}$ in every window. Our LVAR network thus defined has 75 nodes, each corresponding to a financial institution and unweighted non-directional edges such that an edge between institution i and j denotes either that i Granger-causes j , or j Granger-causes i or both, after a BH p-value correction at a 20% threshold level of FDR in the estimated LVAR model.

5.3 Comparison of Pairwise and Lasso Penalized VARs

We estimate the model on a rolling basis every month with data from the previous 36-months. Thus we obtain a network structure for every month in the sample. Similarly, following Billio et al. (2012), we estimate pairwise VARs for the 75 largest firms in every time window as

before and define unweighted non-directional edges such that an edge between institution i and j denotes either that i Granger-causes j , or j Granger-causes i or both at the 5% level of significance.

In Figure 10 we plot the graphs of networks estimated using the pairwise VAR and the LVAR models for two periods overlapping financial crises. The upper and lower panel depict the networks estimated for windows October 1995 - September 1998 and August 2006 - July 2009, respectively. Both types of network plots show high connectivity during crises. However, as expected, the pairwise VAR model estimates a far denser network. In comparison, the LVAR network is sparse and identifies AIG and Goldman Sachs as key central nodes during the 2007-09 period. The benefit of the LVAR model over traditional techniques can be easily seen from these figures. First, it allows us to pin down highly interconnected periods in a cleaner manner and second, it provides a stronger separation between important institutions such as AIG and Goldman Sachs and the rest, compared to the pair-wise model.

Figure 10 around here

Consistent with our simulation results, the pairwise network captures both direct and indirect linkages between the two firms in the real data as well. This in turn results in several false positives. Our refined measure, on the other hand, is able to separate out weaker connections from stronger ones, and hence it provides a more meaningful measure. Since there are limited systemic events during the sample period, it is hard to empirically assess the validity of these models with any reasonable degree of precision. It is, therefore, even more important to rely on statistically principled techniques for future applications of network models. In the remainder of this section we discuss our results and findings in more details to establish the usefulness of our measure in understanding system-wide connectivity.

5.4 Time Series of Summary Statistics

In our first test, we study the evolution of system connectivity based on our measure. In order to do so, we summarize the estimated networks using two primary measures of centrality well known in the network literature, namely degree and closeness.

$$\begin{aligned}\text{Degree of node } i &= \text{deg}(i) = \text{number of edges adjacent to node } i \\ \text{Closeness of node } i &= \frac{1}{\sum_{i \neq j} d(i, j)}\end{aligned}$$

where $d(i, j)$ = shortest path length between node i and j , i.e., number of edges constituting the shortest path between i and j . If there is no path between nodes i and j , then the total number of nodes is used as the shortest path length. While average degree measures the average number of direct neighbors, i.e., connectivity in the network, average closeness measures the shortest number of steps in which a node can be accessed from another node.

Figure 11 and Figure 12 plot average degree and closeness, respectively of our estimated network over 3-year rolling windows. These time series plots show that connectivity, measured either by count of neighbors or distance between nodes, increases before and during systemically important events. In both figures, we mark a few key events of the last decade at the time window when it is first included in the sample. In both figures we see two bigger cycles, one starting around 1998 and another around 2008. The former coincides with the Russian default and LTCM bankruptcy in late 1998 and the latter marks the financial crisis of 2007-2009. In between the two, there is another prominent cycle of increased connectivity starting around 2002 that coincides with the growth of mortgage-backed securities (e.g., see the pattern in MBS growth over this time period in [Ashcraft et al. \(2010\)](#), Figure 3) and the increased connectivity of different sectors of the market through holdings of these securities as well as increased interlinkages through insurance contracts.

Figure 11 around here

Figure 12 around here

The time-series results show that our network measure is sensible in detecting large systemic events. To contrast our measure with pairwise network model, in Fig 13 we plot the evolution of connectedness based on the two models. Note that it is not useful to directly compare the number of connections over time based on the two models, since the pairwise VAR has always significantly higher number of connections. A meaningful measure should be based on deviation from historical levels of connections – disproportionate increase or decrease in connectivity measures compared to historical numbers provides more meaningful information on the buildup of systemic risk in the economy. Thus, we scale the degree centrality of both network models in different rolling windows by the historical average of degree centrality over all rolling windows spanning 1990-2012. Figure 13 provides the results. Both models are able to detect the 2008-09 financial crisis, however, LVAR model does a much better job around the Russian/LTCM default. It is comforting to see the sharp spike in LVAR model-based connectivity in periods leading up to both the important events during our sample period.

Figure 13 around here

It is clear that the key feature of our model is to separate out relatively stronger connections from the weaker ones. Hence, a key benefit of our approach is cross-sectional in nature, namely our model better identifies firms that are systemically more important than the others in a stressful situation. We had shown this advantage with simulated data in Section 4. Now, we identify the important institutions in real data using our model. In Figure 14 we show the list of important firms based on our connectivity measures, and Table 1 contains firm names with ticker symbols. Since the estimated networks exhibit different levels of overall sparsity in different time periods, raw degree centrality of a firm is not ideal to

capture its relative importance in the system. So in each time period, we take the normalized degree of firms, i.e., $(\text{degree} - \text{average node degree})/(\text{standard deviation of node degrees})$, as a measure of systemic importance of the firm in that time period. We list firms with highest degree in networks estimated using 3 year historical data starting May, 2007 and then re-estimating the network every two months. We see that AIG emerges as one of the highest degree nodes as early as July, 2008. We also see the increasing connectivity of Goldman Sachs from March, 2009 onwards. These estimates line up well with the anecdotal evidence on the importance of these institutions, especially AIG, during the financial crisis period. More importantly from a regulatory perspective, the separation between AIG and the second most important institution in our network is stark. Figure 15 reproduces the figure based on pair-wise VAR. In this model too, AIG and GS come up as important institutions, but the separation between AIG and the next firm is much smaller than our model. Thus, when we separate out all the indirect connections in the network, AIG emerges as a significantly more important institution than what one would conclude based on a model that captures the effect of both direct and indirect connections. Second, our model continues to identify AIG as a relatively more important institution even in 2009-2010 period, compared to the corresponding estimation based on pairwise VAR model. Again, the result shows that there are non-trivial practical implications emanating from the estimation method employed.

Figure 14 around here

Figure 15 around here

Table 1 around here

5.5 Results around the Lehman Brothers Failure Event

We exploit the failure of Lehman Brothers in September 2008 as a shock to the system, and use this event to shed light on the usefulness of our network in detecting interconnected firms. On September 10, 2008 Lehman Brothers puts itself up for sale, but does not find a buyer. The U.S. government refuses to step in and ultimately the firm announced its bankruptcy filing on the eve of September 15, 2008. There was considerable government intervention immediately following its collapse. However, in the short window of time from September 10 – September 16, there was significant ambiguity about the bailout possibility. We expect firms connected to Lehman to experience large negative returns during this period. That is indeed the case based on our network estimation. Lehman has two direct connections in the network – AIG and Cigna. As shown in Table 2, AIG experienced large negative returns of -60.8% on September 15. CIGNA had a negative return of -2.9% on the day. Both these firms continue to experience large negative returns till September 18, 2008, when the U.S. government announced a rescue package for AIG. Extending the analysis to the neighbors of Lehman’s neighbors, the Table also produces returns for this event window for firms connected to AIG and CIGNA. They all experience large negative returns on September 15, 2008, with AIG’s neighbors experiencing generally more negative returns than CIGNA’s neighbors. As this analysis illustrate, a useful feature of our model is that we can trace the effect of a negative shock on a firm on the entire network by tracing its effects through the direct linkages. Pairwise analysis doesn’t lend itself to such an experiment due to the confounding indirect effects.

Table 2 around here

5.6 Inter-sectoral Connectivity

Our model allows us to study both within and across sector connectivity. Even since great depression, there has been a number of policy interventions in banking industry that are primarily motivated by concerns about connections across banking, broker-dealer, and insurance sector. A prominent example is the imposition of the Glass-Steagall Act in 1933 that prohibited commercial banks from engaging in investment banking activities, such as underwriting of securities or investment in certain class of securities with their own money or their client's money. Some of the key provisions of the Act were repealed during our estimation period through the enactment of Gramm-Leach-Bliley (GLB) Act of 1999. The GLB Act removed barriers between the commercial banks, broker-dealers and insurance sector. Thus we expect the inter-sectoral connectivity to increase around this period. While the Act itself was finally passed in 1999, the real effect of this act was felt in the market starting from 1998 itself. In 1998, Citicorp, a commercial bank, merged with the insurance company Travelers Group to form a conglomerate combining banking, securities and insurance services under one large group. This merger was in violation of the original Glass-Steagall Act at the time, but after the enactment of GLB Act a year later, it was given a legal status on a retrospective basis. For our network, this is an important event: by law banking, insurance, and broker-dealer sectors are expected to show increased connectivity during this period.

We plot the evolution of inter-sector linkage between the insurance sector and the other two sectors in Figure 16. The figure demonstrates that insurance sector became more connected with both the broker-dealer and banking sector in 1998-1999. These results show that our network topology is consistent with the intended consequence of the repeal of Glass-Steagall Act that increased the connectivity across sectors. Overall our results are consistent with broad changes in the markets and regulations.

6 Conclusions

We propose a measure of network connectivity based on a system-wide approach. Unlike extant measures that rely on pairwise approach, we estimate the connections across all firms in a system-wide sense. Such an improvement is important for measures of risk that are designed to detect system-wide effects. While we use measure based on stock returns to illustrate the usefulness of our approach, our model can also be applied to other sensible measures of firm characteristics such as volatility and value-at-risk.

Our simulation exercises highlight the usefulness of taking a systemic approach suggested by our model – it separates out direct linkages from the indirect ones, which in turn allows us to pin down the source of shock propagation in a system. Several policy proposals, such as linking capital requirements to measures of systemic risk, crucially depend on an accurate measure of this risk. Any misclassification, therefore, is likely to be costly to the economy. Our measure minimizes the possibility of such misclassifications. Finally, we apply our method to large financial institutions of the U.S. and show that our model is able to capture both systemic events and systemically important institutions in a meaningful manner.

A Estimation of large VAR models

We discuss statistical issues for estimating VAR models using ordinary least squares (OLS) when the sample size (n) is small compared to the number of time series (p), and describe how Lasso based penalized estimation methods can be used to overcome them. We conclude with a description of our multiple testing correction methods to construct networks based on fitted VAR models.

In low-dimensional problems ($n > p$), the most common method for estimating VAR models is ordinary least squares (OLS) regression of X^t on X^{t-1} (Lütkepohl, 2005). Formally, given $n + 1$ observations $\{X^0, X^1, \dots, X^n\}$ from the stationary VAR process (1), one forms

autoregressive design

$$\underbrace{\begin{bmatrix} (X^{n+1})' \\ \vdots \\ (X^1)' \end{bmatrix}}_{\mathbf{Y}} = \underbrace{\begin{bmatrix} (X^n)' \\ \vdots \\ (X^0)' \end{bmatrix}}_{\mathbf{X}} A' + \underbrace{\begin{bmatrix} (\varepsilon^{n+1})' \\ \vdots \\ (\varepsilon^1)' \end{bmatrix}}_E \quad (3)$$

The OLS estimate of the VAR transition matrix A is then obtained by conducting p separate, equation-by-equation OLS regressions to estimate the rows of A . Formally,

$$\hat{A}_{i:}^{OLS} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \frac{1}{n} \|\mathbf{Y}_{:i} - \mathbf{X}\beta\|^2, \quad \text{for } i = 1, \dots, p. \quad (4)$$

In classical, low-dimensional asymptotics (p fixed, $n \rightarrow \infty$), \hat{A}^{OLS} is a consistent estimate of A and $\sqrt{n}(\hat{A}^{OLS} - A)$ is asymptotically normal with finite variance-covariance matrix. This allows conducting formal hypothesis tests of Granger causality $H_0 : A_{ij} = 0$ vs. $H_1 : A_{ij} \neq 0$, for all $1 \leq i, j \leq p$, and construct a network of significant Granger causal estimates in a system-wide fashion ⁷.

In a high-dimensional setting with $n < p$, equation-by-equation estimation (4) with OLS is no longer possible. Even for $p < n$, the overall estimation error $\|\hat{A}^{OLS} - A\|_F^2$ is of the order of $O(p^2/n)$, which means one needs at least $O(p^2)$ samples for meaningful estimation. Unfortunately, without further assumptions on the network structure, this is the minimal requirement since we are indeed estimating p^2 free parameters.

Under assumption of sparsity of the true network A ($\|A\|_0 := \sum_{i,j=1}^p \mathbb{1}[A_{ij} \neq 0] = s$, $s \ll n$), classical *subset selection* procedures like best subset, forward, backward and stepwise regression can potentially be used to replace OLS in (4). However, their statistical properties in the $n < p$ setting are unknown, and they are found to be unstable in empirical

⁷Note that this is different from the approach of [Billio et al. \(2012\)](#), who fit separate bivariate VAR models for different pairs of firms (i, j) , $1 \leq i, j \leq p$.

applications (Breiman, 1995). Another alternative to OLS in such situations is *shrinkage* methods like ridge regression which also appears in the literature of Bayesian VAR. Ridge regression shrinks weak coefficients towards zero to reduce the variance of \hat{A} and produce meaningful estimates, but introduces bias in them. More importantly, interpretation of ridge estimates is not obvious since it does not perform explicit variable selection. Also, due to the added bias of ridge regression, inference machinery in high-dimension has not been developed.

Our choice of Lasso (Least Absolute Selection and Shrinkage Operator) is motivated by its ability to provide an attractive middle ground - it shrinks regression coefficients to reduce variance and make consistent estimation possible in high-dimension, and at the same time performs automatic variable selection by setting weaker coefficients exactly to zero. The resulting estimates are sparse and easier to interpret. Similar to ridge, lasso estimates are biased and statistical inference with them remained a challenging problem for a long time. However, recent developments in high-dimensional statistics have provided means to correct the bias and carry out formal tests of significance of Lasso coefficients in a regression problem. We use Lasso, followed by a bias correction, to estimate our VAR model.

Penalized VAR estimation with Lasso: The equation-by-equation estimate of Lasso VAR is defined as

$$\hat{A}_{i\cdot} = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \frac{1}{n} \|Y_{\cdot i} - \mathbf{X}\beta\|^2 + \lambda_i \|\beta\|_1, \quad i = 1, \dots, p \quad (5)$$

Here $\|\beta\|_1 := \sum_{j=1}^p |\beta_j|$ is the ℓ_1 -norm penalty, which encourages sparsity in the solution, by shrinking smaller coordinates to zero. The tuning parameter λ_i controls the degree of sparsity in the estimate, larger values of λ_i result in sparser $\hat{A}_{i\cdot}$.

Note that under the model (1) with diagonal Σ_ϵ , equation-by-equation estimate of Lasso

VAR indeed coincides with the penalized maximum likelihood estimate (MLE)

$$\hat{A} := \operatorname{argmin}_{A \in \mathbb{R}^{p \times p}} \frac{1}{2n} \|\hat{A} - A\|_F^2 + \sum_{i=1}^p \lambda_i \|A_{i:}\|_1 \quad (6)$$

Since the equation-by-equation estimate is equivalent to p separate Lasso estimates, in our discussion we focus on the generic Lasso estimation of a linear model $Y_{n \times 1} = \mathbf{X} \beta_{p \times 1}^0 + \epsilon_{n \times 1}$, given by

$$\hat{\beta} := \operatorname{argmin}_{\beta \in \mathbb{R}^p} \frac{1}{2n} \|Y - \mathbf{X}\beta\|^2 + \lambda \|\beta\|_1. \quad (7)$$

For estimating the i^{th} row of A using Lasso, the errors in the above regression take the form $\epsilon = [\epsilon_i^n : \dots : \epsilon_i^1]'$, the true coefficients $\beta^0 = A'_{i:}$, and the penalty λ_i is chosen based on $\sigma_i = sd(\epsilon_i^1)$ (see next section). The design matrix $\mathbf{X} = [X^n : \dots : X^0]'$ is the same across all p regressions. The facts that the rows of the design matrix are not i.i.d. and the error vector ϵ is correlated with the design matrix X are specific to VAR estimation problems, and violate the assumptions under which statistical properties of Lasso and debiased Lasso have been studied in literature. We provide some new theoretical analysis to justify their validity in the context of VAR estimation.

Choice of tuning parameters: In practice, choosing the “best” tuning parameter λ is cumbersome and depends on the context of the problem. AIC, BIC or Cross-validation (CV) guided choice of λ are commonly used, although they are known to perform poorly in high-dimensional problems, where $n \ll p$. Since our sample size is small, we use a theory-driven, plug-in estimate rather than cross-validation or data-driven strategies. The theoretical choice of $\lambda \propto \sigma \sqrt{\log p/n}$ (Bühlmann and Van De Geer, 2011) requires knowledge of the error standard deviation $\sigma = \sqrt{\text{Var}(\epsilon_1)}$, which is seldom available in practice. To mitigate these problems, we use the scaled lasso algorithm in (Sun and Zhang, 2012) to obtain an estimate of $\hat{\sigma}$, and choose $\lambda = C\hat{\sigma}\sqrt{\log p/n}$. Scaled Lasso solves the following

convex optimization problem

$$(\hat{\beta}, \hat{\sigma}) \equiv \underset{\beta \in \mathbb{R}^p, \sigma > 0}{\operatorname{argmin}} \frac{1}{2\sigma n} \|Y - \mathbf{X}\beta\|^2 + \frac{\sigma}{2} + \tilde{\lambda} \sum_{j=1}^p |\beta_j|. \quad (8)$$

with $\tilde{\lambda} = C\sqrt{\log p/n}$, for some constant C that does not depend on model parameters. Scaled Lasso enjoys similar theoretical properties as Lasso in high-dimensional problems, but does not require knowledge of error standard deviation σ in the choice of tuning parameter. Rather, it provides as a by-product an estimate of σ which can be used for follow-up analyses as in hypothesis testing and confidence interval construction for the regression coefficients $\hat{\beta}_j$.

Consistency of Lasso VAR in high-dimension. [Basu and Michailidis \(2015\)](#) have established that the Lasso VAR estimates are consistent in high-dimensional settings, i.e., assuming p grows with n , possibly at a faster rate. In particular, under a double asymptotic framework where both $p, n \rightarrow \infty$, $p = O(n^\alpha)$ for any $\alpha > 0$, and the true sparsity $s = o(n)$, it follows from the results of Section 4 in [Basu and Michailidis \(2015\)](#) that $\|\hat{\beta} - \beta^0\|^2 = O_P(s \log p/n)$ with a choice of $\lambda \propto \sqrt{\log p/n}$ and the underlying Gaussian VAR process is stable ([Lütkepohl, 2005](#)). This rate of convergence demonstrates the remarkable advantage of Lasso (also reported in several other works involving i.i.d. data): modulo a cost of $\log(p^2) = 2 \log(p)$ for searching the locations of non-zero coordinates in A , one needs merely $O(s)$ samples to estimate the VAR coefficients, which is the same as if one *a priori* knew the positions of the s non-zero edges and were only estimating the s free parameters of edge strengths. So, for problems where $s \log p \ll p^2$, Lasso VAR achieves comparable estimation accuracy as OLS with much smaller sample size.

Bias Correction of Lasso VAR estimates:

Despite the nice estimation accuracy of Lasso VARs as above, there are two limitations of using it directly for network construction. First, the shrinkage effect of Lasso introduces a

bias in estimating the edge strength, which can be potentially large in a finite-sample setting. Second, the Lasso VAR estimates \hat{A}_{ij} do not come with any measure of uncertainty.

(Javanmard and Montanari, 2014) proposes a methodology to bias-correct the VAR estimates so as to draw statistical inference. Bias correction of nonlinear estimates is a common technique in classical statistics (Cordeiro and McCullagh, 1991; Cordeiro and Vasconcellos, 1997). For high-dimensional regression problems, Zhang and Zhang (2014) first proposed a bias correction method for constructing confidence intervals of the individual regression coefficients. In parallel lines of work, van de Geer et al. (2014); Javanmard and Montanari (2014) also proposed bias corrected versions of Lasso for linear regression. For a more detailed discussion of the intuition behind bias correction, we refer the readers to the excellent review article Dezeure et al. (2015).

In order to correct the bias of Lasso VAR estimates, we first construct a matrix M , which can be viewed as an approximate inverse of the sample covariance matrix $\hat{\Sigma}_X = \mathbf{X}'\mathbf{X}/n$. Given a tuning parameter $\mu > 0$ (chosen in the order of $\sqrt{\log p/n}$), the j^{th} row of the matrix M , $1 \leq j \leq p$, is obtained by solving the following convex program

$$\begin{aligned} & \text{minimize} && m' \hat{\Sigma} m \\ & \text{subject to} && \left\| \hat{\Sigma} m - e_j \right\|_{\infty} \leq \mu, \end{aligned} \tag{9}$$

where $e_j \in \mathbb{R}^p$ is the vector with 1 at the j^{th} position and zero at all the other coordinates. If any one of the p convex programs is not feasible, the matrix M is set to identity⁸.

With the new matrix M , for any given $j \in \{1, \dots, p\}$, the debiased Lasso estimate is given by

$$\tilde{\beta} = \hat{\beta} + \frac{1}{n} M \mathbf{X}' (Y - \mathbf{X} \hat{\beta}), \tag{10}$$

⁸Using arguments of Lemma 23 in Javanmard and Montanari (2014) together with Proposition 2.4 in Basu and Michailidis (2015), we can show that $\Sigma^{-1} = (\text{Var}(X^1))^{-1}$ is a solution to 9 (hence the constrained optimization problem is feasible) with high probability.

where $\hat{\beta}$ is a solution of (7).

The intuition of debiasing is simple and can be explained in a low-dimensional context assuming M is exactly $\hat{\Sigma}^{-1}$. Let $\delta = \hat{\beta} - \beta$ denote the bias of lasso estimate $\hat{\beta}$. Then (10) can be expressed as

$$\begin{aligned}\tilde{\beta} &= \beta^0 + \delta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(-\mathbf{X}\delta + \epsilon) \\ &= \beta^0 + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon,\end{aligned}$$

which is identical to the ordinary least squares estimate. Since ϵ is zero mean Gaussian, under suitable regularity assumption on \mathbf{X} which ensures that the variance does not blow up, the second term is asymptotically negligible, making the bias of $\tilde{\beta}$ orders of magnitude smaller than the bias of $\hat{\beta}$. Another way to motivate the debiasing procedure is to view the bias correction step as an approximate Newton-Raphson iterate, since $(1/n)\mathbf{X}'(Y - \mathbf{X}\hat{\beta})$ and $\hat{\Sigma}$ can be viewed as approximate gradient and Hessian of the least-squares loss function evaluated at the current iterate $\hat{\beta}$. The classic method of Fisher's scoring uses a similar one-step update to a consistent estimate to reduce its variance and make it efficient (Le Cam, 1956).

The estimation error of debiased Lasso, after proper rescaling, allows the decomposition

$$\sqrt{n}(\tilde{\beta} - \beta^0) = \frac{1}{\sqrt{n}}M\mathbf{X}'\epsilon + \Delta, \text{ where } \Delta = -\sqrt{n}(M\hat{\Sigma} - I)(\hat{\beta} - \beta^0). \quad (11)$$

Suppose the tuning parameters λ and μ are chosen to be of the order $\sqrt{\log p/n}$. In a double asymptotic regime $p, n \rightarrow \infty$ mentioned above, Proposition 4.1 in Basu and Michailidis (2015) established that $\|\hat{\beta} - \beta^0\|_1$ is $O_P(s\sqrt{\log p/n})$. This, together with (9), implies that $\|\Delta\|_\infty$ is of the order $O_P(s \log p/\sqrt{n})$. Hence, the bias term Δ is asymptotically negligible when $s \log p = o(\sqrt{n})$, and it is possible to conduct inference using only the asymptotic distribution of the first term.

Inference with debiased Lasso. We discuss construction of p -values for the hypothesis $H_0 : \beta_j = 0$ vs. $H_A : \beta_j \neq 0$, for a fixed $j \in \{1, \dots, p\}$, under the double asymptotic regime $n, p \rightarrow \infty$ where $s\sqrt{\log p}/n \rightarrow 0$. Leveraging the asymptotic negligibility of the bias term in (11), we use the method proposed in Javanmard and Montanari (2014) to construct p -values for testing significance of the individual edges A_{ij} . Formally, for every $i, j \in \{1, \dots, p\}$, the p -value for testing

$$H_0 : \beta_j^0 = 0 \text{ vs. } H_A : \beta_j^0 \neq 0$$

is given by

$$P_j = 2 \left[1 - \Phi \left(\frac{\sqrt{n}|\tilde{\beta}_j|}{\hat{\sigma}[M\hat{\Sigma}_X M']_{jj}} \right) \right] \quad (12)$$

where $\Phi(\cdot)$ is the standard normal cdf, and $\hat{\sigma}$ is a consistent estimate of error standard deviation σ obtained using scaled Lasso (8) with $\tilde{\lambda} := 10\sqrt{2\log p/n}$, as suggested in the theoretical analysis of Javanmard and Montanari (2014).

Network construction with VAR estimates: Using the estimates of p lasso problems as row vectors, we construct our debiased Lasso VAR estimate \tilde{A} . This matrix can be used to estimate the weighted, directed network described in section 3. An edge is present from node j to node i if A_{ij} is significant at a pre-specified threshold $\alpha > 0$.

Choice of the significance threshold α is important, since constructing the directed network amounts to performing $p(p-1)$ hypothesis tests. For large p , this requires a correction for multiple testing to avoid the problem of high false positives. The standard Bonferroni criterion for controlling the family-wise error rate (FWER) is the most conservative one, but it suffers from low power. We use a less stringent criterion of multiple testing, proposed in Benjamini and Hochberg (1995), to control the False Discovery Rate (FDR). False Discovery Rate is the expected proportion of falsely rejected hypotheses over the total number of rejected hypotheses. Thus, a 20% false discovery rate would imply that, on average, 1 out of 5 selected edges is falsely detected. The procedure was originally proposed for independent

Algorithm 1: Network GC using Debiased Lasso VAR

Input: Data: $\{X^0, \dots, X^n\}$, $X^t \in \mathbb{R}^p$, upper bound on false discovery rates (FDR): α
 $\mathbf{X} \leftarrow [X^{n-1} : \dots : X^0]'$
for $i \leftarrow 1$ **to** p **do**
 $Y \leftarrow [X_i^n : \dots : X_i^1]$
 Estimate noise standard deviation $\hat{\sigma}_i$ using scaled Lasso (8)
 Set tuning parameter of Lasso $\lambda \leftarrow \hat{\sigma}_i \sqrt{\log p/n}$
 Calculate Lasso VAR estimate $\hat{A}_{i\cdot}$ using (7)
 Calculate debiased Lasso VAR estimate $\tilde{A}_{i\cdot}$ using (10)
 for $j \leftarrow 1$ **to** p **do**
 Calculate p -value P_{ij} for testing $H_0 : A_{ij} = 0$ vs. $H_A : A_{ij} \neq 0$ using (12)
 end
end
Adjust p -values P_{ij} , for $1 \leq i, j \leq p$, for multiple testing using Benjamini-Hochberg procedure, controlling for FDR at α
Set $\tilde{A}_{ij} \leftarrow 0$, for all i, j with $P_{ij} \geq \alpha$
Output: Estimated weighted adjacency matrix \tilde{A}

test statistics, and its validity for test statistics with positive regression dependency was established in [Benjamini and Yekutieli \(2001\)](#).

The Benjamini-Hochberg (BH) procedure works as follows. Suppose we are testing m null hypotheses H_1, H_2, \dots, H_m with p -values P_1, \dots, P_m . Given the sequence of ordered p -values $P_{(1)} \leq \dots \leq P_{(m)}$, the following twp-step procedure controls the FDR at a level $\alpha > 0$:

1. find the largest integer $k \geq 1$ such that $P_{(k)} \leq \frac{k}{m}\alpha$;
2. reject all null hypotheses $H_{(i)}$, for $i = 1, \dots, k$.

The topology of a weighted, directed network with edges significant at a level α (after correcting for multiple testing), or its undirected, unweighted skeleton $\mathcal{S}(G)$, can be explored by standard visualization software or by calculating network centrality measures described in section 5.

The complete algorithm for calculating weighted adjacency matrix \tilde{A} based on debiased Lasso VAR is described in Algorithm 1.

B Estimating Contemporaneous Dependence with de-biased graphical Lasso

We assume the stock returns of p firms across n time points, after appropriate transformations (log, GARCH filter etc.) to reduce nonstationarity, are stored in a $n \times p$ matrix X , where X_{ij} denotes the return of firm j at time point i . We center the columns of X to have zero mean and unit variance.

Assuming the returns are independent across time, and the p -dimensional vector of returns follow a $N(0, \Sigma)$ distribution, returns of firms i and j are conditionally independent given the rest of the firms if and only if their partial correlation is zero. Since $\Theta = \Sigma^{-1}$ contains partial covariances between firm returns, returns of firms i and j are partially uncorrelated if and only if $\Theta_{ij} = 0$. Hence, the set $E = \{(i, j) | \Theta_{ij} \neq 0\}$ can be used as the set of edges of a graph among the p firms. This procedure is also known as Gaussian graphical modeling.

In classical low-dimensional setting (p fixed, $n \rightarrow \infty$), the maximum likelihood estimate of Θ is given by S^{-1} , where $S := X'X/n$ is the sample covariance matrix. In high-dimensional setting ($n \ll p$), S is not invertible and the MLE is not uniquely defined. In such cases, under sparsity assumption on Θ , the graphical Lasso (Glasso) estimate (Friedman et al., 2008) $\hat{\Theta}$ is defined as the minimizer of penalized negative log likelihood

$$\hat{\Theta} = \underset{\Theta \succeq 0}{\operatorname{argmin}} tr(S\Theta) - \log \det(\Theta) + \lambda \sum_{i \neq j} |\Theta_{ij}|,$$

where λ is a tuning parameter controlling the level of sparsity in $\hat{\Theta}$, and $\Theta \succeq 0$ indicates that we minimize over the set of nonnegative definite matrices.

The Glasso estimate is known to be consistent for Θ in high-dimensional settings, as long as Θ is sufficiently sparse (Ravikumar et al., 2011). However, just like Lasso, the Glasso

estimate is also known to be biased, and formal statistical inference (confidence intervals, hypothesis testing) with $\hat{\Theta}_{ij}$ is not possible without correcting the bias.

The debiased graphical Lasso (DGlasso), proposed recently in [Jankova et al. \(2015\)](#), corrects the bias of Glasso and provides a method for formally testing presence or absence of an edge between firms i and j ($H_0: \Theta_{ij} = 0$ vs. $H_A: \Theta_{ij} \neq 0$). Below we provide a brief description of DGlasso and refer the readers to [Jankova et al. \(2015\)](#) for more details.

Since columns of X are scaled to have zero mean and unit variance, we set $\lambda = \sqrt{\log p/n}$, as suggested in [Jankova et al. \(2015\)](#). Starting with the Glasso estimate $\hat{\Theta}$ and the sample covariance S , the debiased Glasso estimate is defined as

$$\hat{T} = \text{vec}(\hat{\Theta}) - \hat{\Theta} \otimes \hat{\Theta} \text{vec}(S - \hat{\Theta}),$$

where $\text{vec}(A) = [A_{11}, \dots, A_{p1}, \dots, A_{1p}, \dots, A_{pp}]'$ is a vector formed by stacking the columns of a $p \times p$ matrix A , and $A \otimes B = ((A_{ij}B))_{1 \leq i, j \leq p}$ is the $p^2 \times p^2$ Kronecker product of matrices A and B . Similar to the bias correction of Lasso described in [Appendix A](#), the above bias correction can also be viewed as an approximate Newton-Raphson step.

Under some regularity conditions on Θ , [Jankova et al. \(2015\)](#) established that the entries of the debiased estimate \hat{T} have the following asymptotic distribution

$$\sqrt{n} \left(\hat{T}_{ij} - \Theta_{ij} \right) / \hat{\sigma}_n \xrightarrow{d} N(0, 1), \quad \hat{\sigma}_n^2 := \hat{\Theta}_{ij}^2 + \hat{\Theta}_{ii} \hat{\Theta}_{jj}.$$

Leveraging this asymptotic distribution, we can formally test presence or absence of individual edges in the Gaussian graphical model

$$H_0 : \Theta_{ij} = 0 \quad \text{vs.} \quad H_A : \Theta_{ij} \neq 0$$

and calculate p -values using the formula $1 - \Phi \left(\sqrt{n} |\hat{T}_{ij}| / \hat{\sigma}_n \right)$, where $\Phi(\cdot)$ is the cumulative

distribution function of standard normal density.

The network of partial correlations among the returns of p firms can be constructed using only statistically significant edges after correcting for multiple comparisons using Benjamini-Hochberg procedure described in appendix A. Unlike the directed network constructed using the transition matrix of a VAR model, the network of contemporaneous relationships is undirected since Θ is already a symmetric matrix. So we can work with the estimated network directly and calculate centrality measures.

C Computing debiased Lasso and Glasso

The simulation and real data analyses in this paper using the statistical software R. We calculated debiased Lasso using the R codes available on the webpage of the authors of Javanmard and Montanari (2014) at <http://web.stanford.edu/~montanar/sslasso/> with the default choices of tuning parameters, and implemented the algorithm of Jankova et al. (2015) for debiased graphical Lasso ourselves. We used the R function `p.adjust()` to implement the Benjamini-Hochberg procedure for multiple testing corrections. For the empirical analysis, univariate GARCH models were fitted using R package `fGarch`.

References

- D. Acemoglu, A. E. Ozdaglar, and A. Tahbaz-Salehi. Systemic risk in endogenous financial networks. *Available at SSRN 2553900*, 2015.
- V. Acharya, R. Engle, and M. Richardson. Capital shortfall: A new approach to ranking and regulating systemic risks. *The American Economic Review*, pages 59–64, 2012.
- T. Adrian and M. K. Brunnermeier. Covar. Technical report, National Bureau of Economic Research, 2011.
- F. Allen and D. Gale. Optimal financial crises. *The journal of finance*, 53(4):1245–1284, 1998.
- F. Allen and D. Gale. Financial contagion. *Journal of political economy*, 108(1):1–33, 2000.
- A. B. Ashcraft, P. Goldsmith-Pinkham, and J. I. Vickery. Mbs ratings and the mortgage credit boom. 2010.
- A. Babus. The formation of financial networks. 2013.
- M. Barigozzi and C. Brownlees. Nets: Network estimation for time series. *Available at SSRN 2249909*, 2013.
- S. Basu and G. Michailidis. Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 43(4):1535–1567, 2015.
- S. Basu, A. Shojaie, and G. Michailidis. Network granger causality with inherent grouping structure. *Journal of Machine Learning Research*, 16:417–453, 2015.
- S. Battiston, D. D. Gatti, M. Gallegati, B. Greenwald, and J. E. Stiglitz. Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk. *Journal of Economic Dynamics and Control*, 36(8):1121–1141, 2012.

- M. L. Bech and E. Atalay. The topology of the federal funds market. *Physica A: Statistical Mechanics and its Applications*, 389(22):5223–5246, 2010.
- Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300, 1995.
- Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188, 2001.
- M. Billio, M. Getmansky, A. W. Lo, and L. Pelizzon. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559, 2012.
- F. Blasques, F. Bräuning, and I. Van Lelyveld. A dynamic network model of the unsecured interbank lending market. 2015.
- M. Boss, H. Elsinger, M. Summer, and S. Thurner 4. Network topology of the interbank market. *Quantitative Finance*, 4(6):677–684, 2004.
- L. Breiman. Better subset regression using the nonnegative garrote. *Technometrics*, 37(4):373–384, 1995.
- C. T. Brownlees and R. F. Engle. Srisk: A conditional capital shortfall measure of systemic risk. *Available at SSRN 1611229*, 2015.
- C. T. Brownlees, E. Nualart, and Y. Sun. Realized networks. *Available at SSRN 2537986*, 2015.
- M. K. Brunnermeier and L. H. Pedersen. Market liquidity and funding liquidity. *Review of Financial studies*, 22(6):2201–2238, 2009.

- P. Bühlmann and S. Van De Geer. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.
- P. Bühlmann and S. van de Geer. High-dimensional inference in misspecified linear models. *Electronic Journal of Statistics*, 9(1):1449–1473, 2015.
- J. A. Chan-Lau, M. Espinosa, K. Giesecke, and J. A. Solé. Assessing the systemic implications of financial linkages. *IMF global financial stability report*, 2, 2009.
- P. Colla and A. Mele. Information linkages and correlated trading. *Review of Financial Studies*, 23(1):203–246, 2010.
- G. M. Cordeiro and P. McCullagh. Bias correction in generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 629–643, 1991.
- G. M. Cordeiro and K. L. Vasconcellos. Bias correction for a class of multivariate nonlinear regression models. *Statistics & probability letters*, 35(2):155–164, 1997.
- B. Craig and G. Von Peter. Interbank tiering and money center banks. *Journal of Financial Intermediation*, 23(3):322–347, 2014.
- M. Demirer, F. X. Diebold, L. Liu, and K. Yilmaz. Estimating global bank network connectedness. 2017.
- R. Dezeure, P. Bühlmann, L. Meier, N. Meinshausen, et al. High-dimensional inference: Confidence intervals, p -values and r-software hdi. *Statistical Science*, 30(4):533–558, 2015.
- D. W. Diamond and P. H. Dybvig. Bank runs, deposit insurance, and liquidity. *Journal of political economy*, 91(3):401–419, 1983.
- F. X. Diebold and K. Yilmaz. On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134, 2014.

- M. Eichler. Graphical modelling of multivariate time series. *Probability Theory and Related Fields*, 153(1-2):233–268, 2012.
- R. Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002.
- J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.
- P. Glasserman and H. P. Young. How likely is contagion in financial networks? *Journal of Banking & Finance*, 50:383–399, 2015.
- M. Gofman. Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions. *Available at SSRN 2194357, Forthcoming in Journal of Financial Economics*, 2016.
- C. W. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.
- C. W. Granger. Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and control*, 2:329–352, 1980.
- G. Iori, G. De Masi, O. V. Precup, G. Gabbi, and G. Caldarelli. A network analysis of the italian overnight money market. *Journal of Economic Dynamics and Control*, 32(1):259–278, 2008.
- J. Jankova, S. van de Geer, et al. Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics*, 9(1):1205–1229, 2015.
- A. Javanmard and A. Montanari. Confidence intervals and hypothesis testing for high-

- dimensional regression. *The Journal of Machine Learning Research*, 15(1):2869–2909, 2014.
- L. Le Cam. On the asymptotic theory of estimation and testing hypotheses. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California, 1956.
- H. Lütkepohl. *New introduction to multiple time series analysis*. Springer, 2005.
- N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the lasso. *Ann. Statist.*, 34(3):1436–1462, 06 2006. doi: 10.1214/009053606000000281. URL <http://dx.doi.org/10.1214/009053606000000281>.
- P. Ravikumar, M. J. Wainwright, G. Raskutti, and B. Yu. High-dimensional covariance estimation by minimizing 1-penalized log-determinant divergence. *Electronic Journal of Statistics*, 5:935–980, 2011.
- K. Soramäki, M. L. Bech, J. Arnold, R. J. Glass, and W. E. Beyeler. The topology of interbank payment flows. *Physica A: Statistical Mechanics and its Applications*, 379(1): 317–333, 2007.
- T. Sun and C.-H. Zhang. Scaled sparse linear regression. *Biometrika*, page ass043, 2012.
- R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.
- S. van de Geer and B. Stucky. χ^2 -confidence sets in high-dimensional regression. In *Statistical Analysis for High-Dimensional Data*, pages 279–306. Springer, 2016.
- S. van de Geer, P. Bühlmann, Y. Ritov, and R. Dezeure. On asymptotically optimal con-

- fidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3): 1166–1202, 2014.
- M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1-2):1–305, 2008.
- C.-H. Zhang and S. S. Zhang. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):217–242, 2014.

Figure 1: A schematic representation of VAR(1) model with $p = 3$ firms A, B, C. There are 3 network Granger causal effects in the model: $B \rightarrow C$, $B \rightarrow A$, $C \rightarrow B$. However, due to indirect effect through B, there is additional pairwise Granger causal effect $C \rightarrow A$.

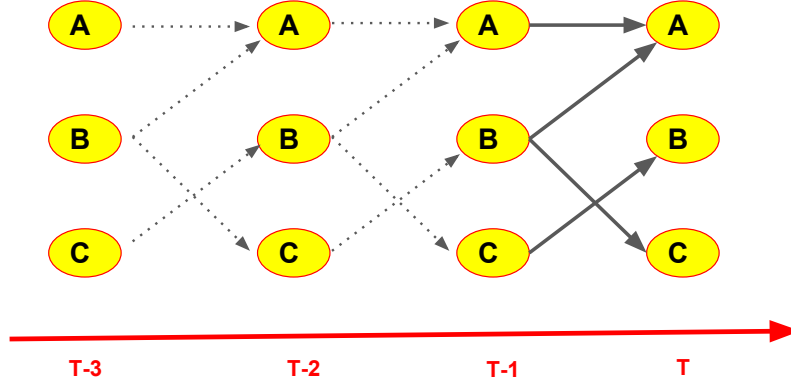
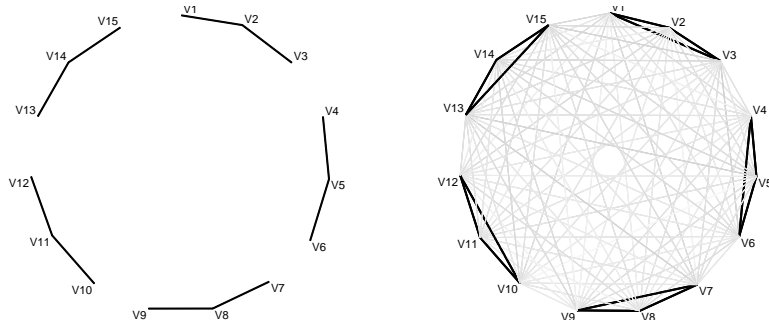


Figure 2: Simulation



True network and the network estimated by pairwise GC in a simulated universe with 15 firms. In this universe, there are 5 central firms $\{2, 5, 8, 11, 14\}$ - each affecting two different firms and forming a network with 5 hubs. In addition to the true network edges, the pairwise GC method picks up additional edges between each pair of non-central firms in each hub. The shade of the edges are darker proportional to the number of times they are estimated.

Figure 3: A simulated network estimation ($n = 500$, $p = 15$) with pairwise VAR and de-biased Lasso VAR (LVAR). The true network (left) has 5 hubs, each of size 3. Pairwise VAR (middle) estimates marginal association and captures indirect effects, and hence the estimated network (middle panel) has 5 complete cycles. Lasso VAR (right), on the other hand, estimates conditional dependence and accurately identifies the structure of the 5 hubs, including the central node and the neighbors. The grayscales of edges represents the proportion of times an edge was detected by Lasso VAR and pairwise VAR in 100 simulated datasets from the true network.

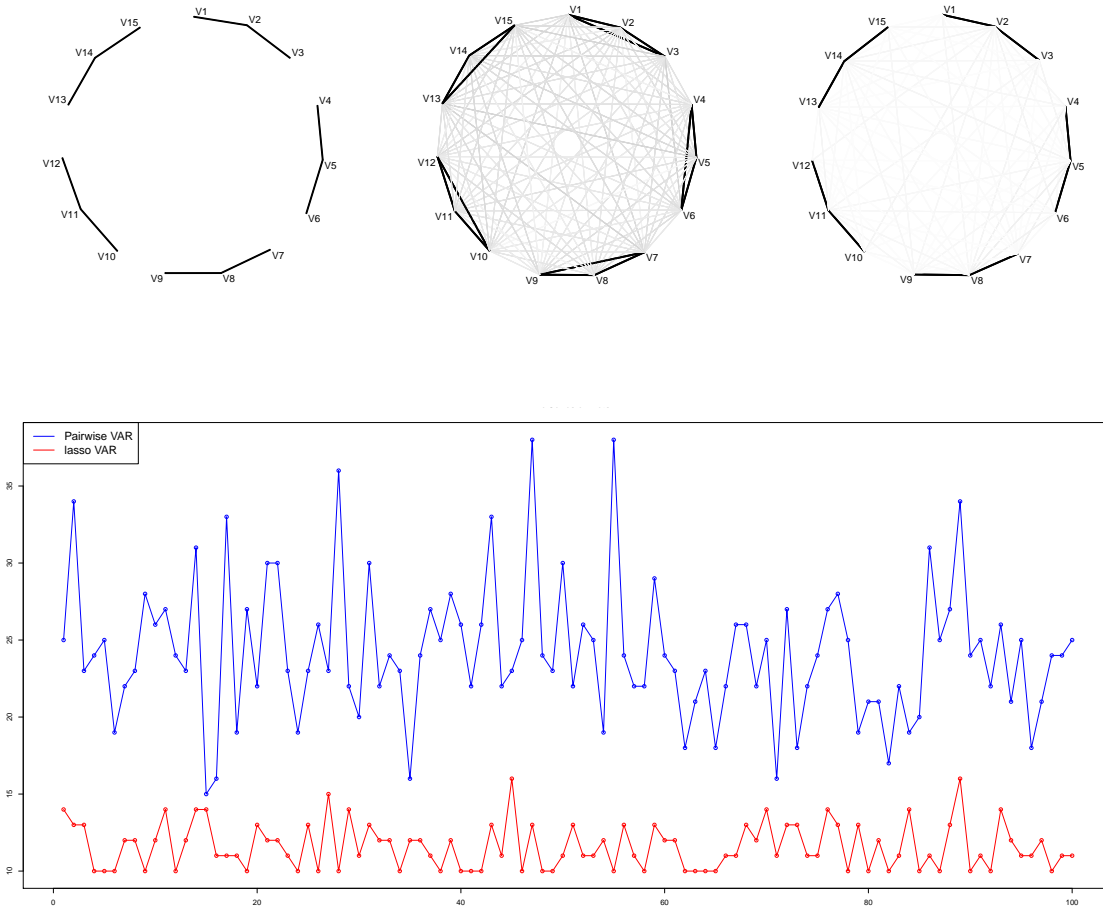


Figure 4: Edge discovery in the simulated network estimation problem of Figure 3. The total number of significant edges discovered by Lasso VAR and pairwise VAR in 100 simulated datasets from the true network are plotted. Pairwise VAR (blue) selects at least 15 edges in all instances, while de-biased Lasso VAR (red) selects much fewer edges, between 10 and 15, consistently across different datasets. The number of edges in the true network is 10, as shown in the left panel of Figure 3.

Figure 5: True Network with 5 isolated firms $\{1, 2, 3, 4, 5\}$, three central firms $\{8, 13, 18\}$ each with 4 neighbors. The returns were simulated based on a Gaussian VAR(1) model with a transition matrix A with the above network structure. In particular, we set $A_{ii} = 0.7$ for $i = 1, \dots, 20$. Also, for every $j \in \{8, 13, 18\}$ and $i \in \{j - 2, j - 1, j + 1, j + 2\}$, we set $A_{ij} = 0.6 + \eta_{ij}$, where $\eta_{ij} \stackrel{i.i.d.}{\sim} \text{uniform}(0, 0.05)$. For all other pairs $\{i, j\}$, we set $A_{ij} = 0$.

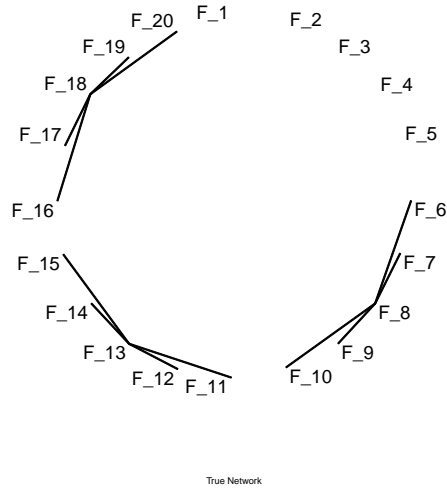


Figure 6: Boxplots of systemic risk measures based on 100 simulated datasets of size $n = 500$ generated from a VAR(1) model described in Figure 5. For the pairwise measures MES, ΔCoVaR and pairwise VAR, the first 5 isolated firms have the lowest systemic risk measure. However, the systemic risk measures of the central nodes $\{8, 13, 18\}$ are not significantly different from the peripheral nodes. In LVAR, the degrees of the central nodes are significantly different from the rest, and hence identification of risky nodes is easier.

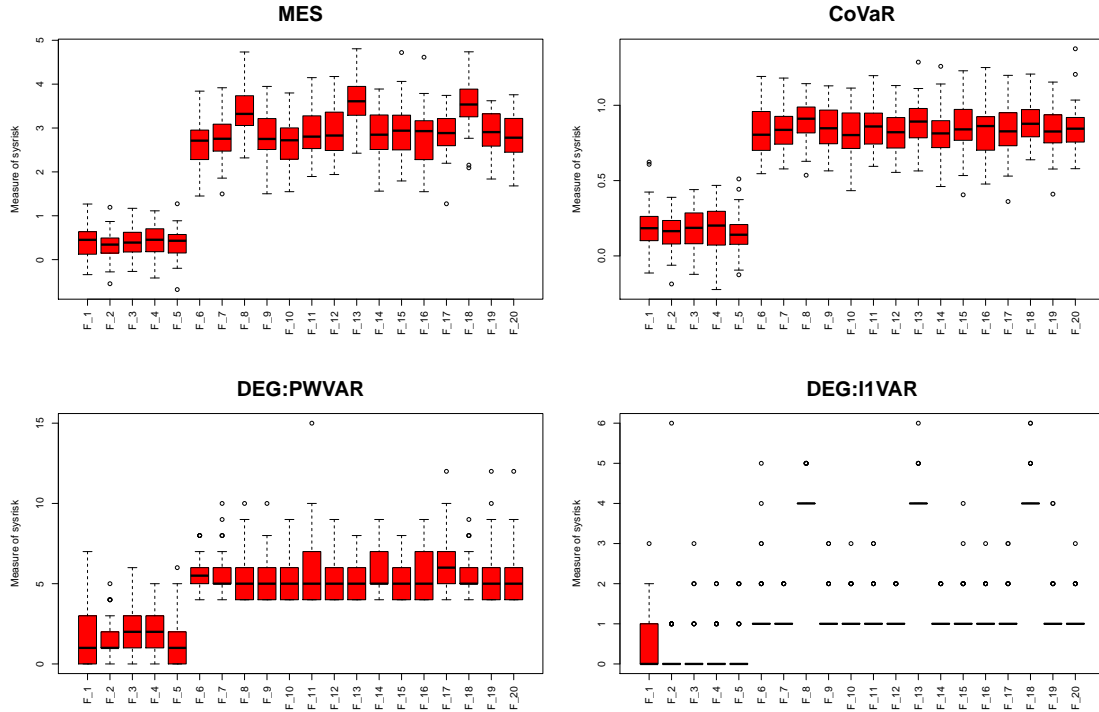


Figure 7: Boxplots of systemic risk measures based on 100 simulated datasets of size $T = 500$ with only contemporaneous dependence among nodes. The data are generated from a Gaussian graphical model with a true network structure of Figure 5, see Section 4.3 for more details. We report the performance of three pairwise measures: MES, ΔCoVaR , pairwise VAR, and a system-wide measure, viz., debiased graphical lasso. For the pairwise measures MES, ΔCoVaR and pairwise VAR, the first 5 isolated firms have the lowest systemic risk measure. However, the systemic risk measures of the central nodes $\{8, 13, 18\}$ are not significantly different from the peripheral nodes. In networks estimated by debiased graphical lasso, the degrees of the central nodes are significantly different from the rest, and hence identification of risky nodes is easier.

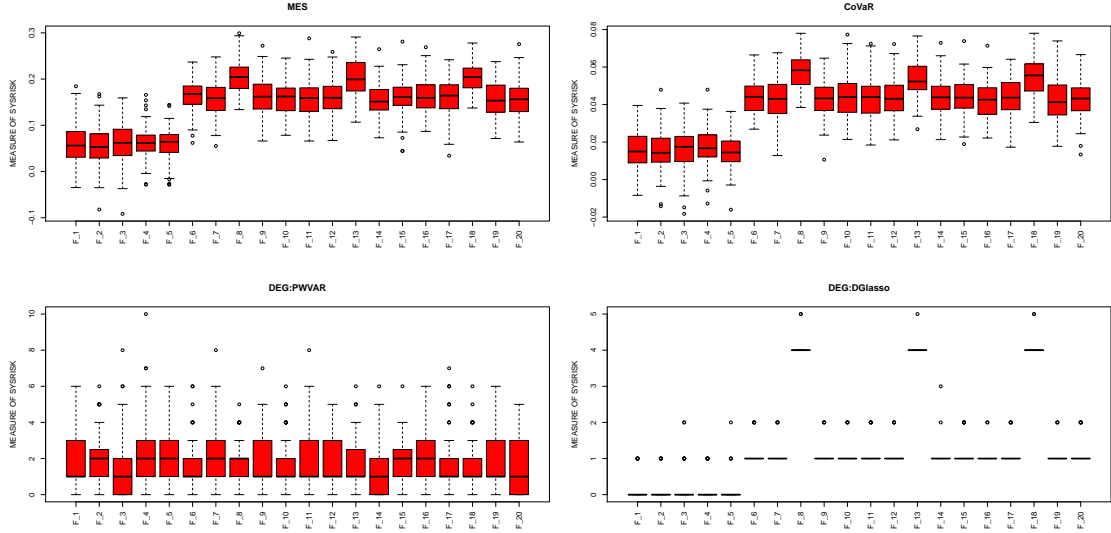


Figure 8: Average monthly return of firms used in the empirical analysis of Section 5 over 3-year rolling windows spanning 1990 – 2012. In each window, 25 largest firms (in terms of market capitalization) from three sectors - Banks (BA), primary broker-dealers (PB), and insurance firms (INS), are included.

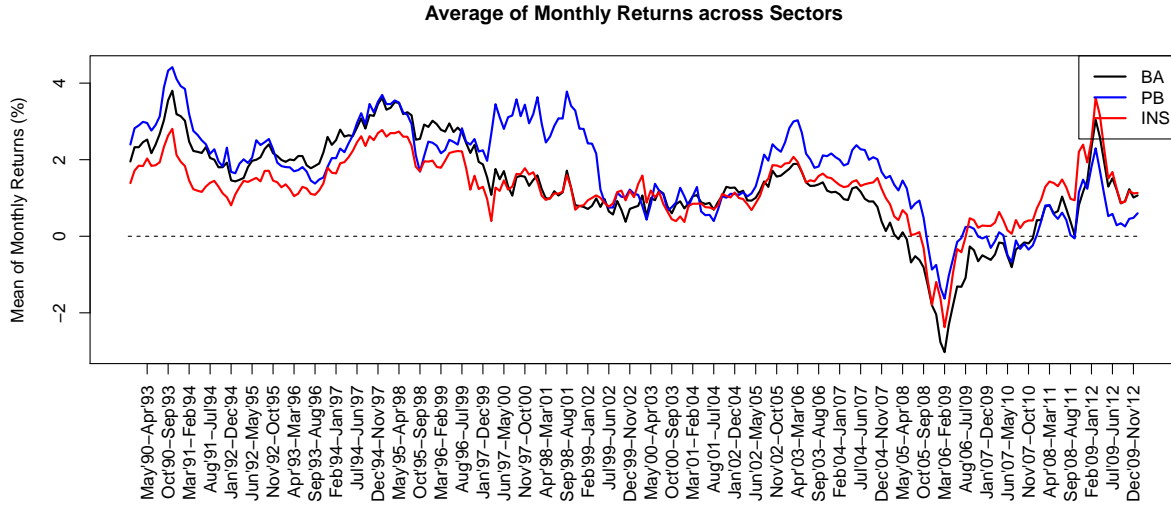


Figure 9: Standard deviation of monthly returns of firms used in the empirical analysis of Section 5 over 3-year rolling windows spanning 1990 – 2012. In each window, 25 largest firms (in terms of market capitalization) from three sectors - Banks (BA), primary broker-dealers (PB), and insurance firms (INS), are included.

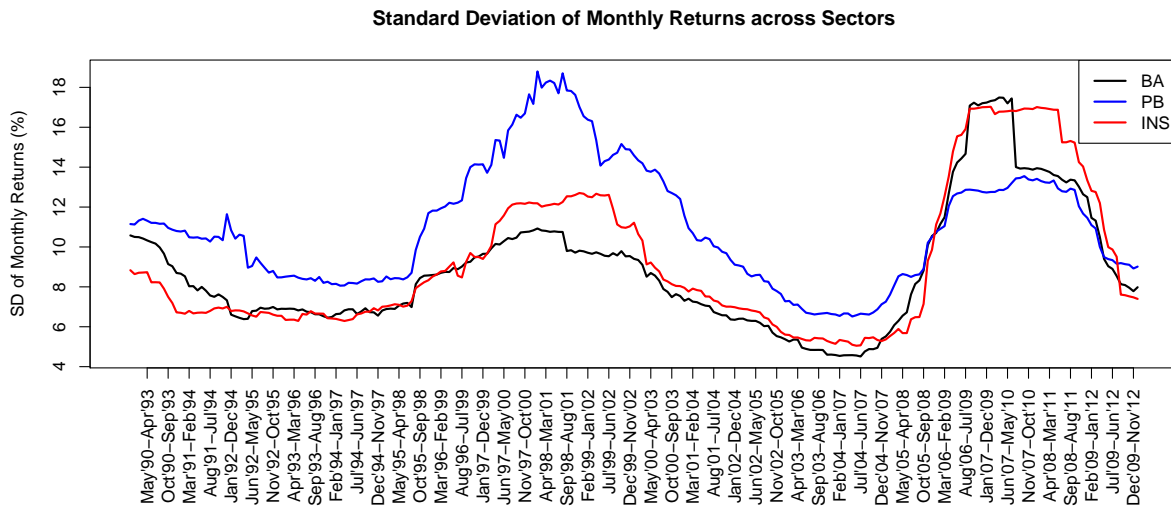


Figure 10: Networks estimated by pairwise VAR and lasso VAR on the time horizons (a) Oct 1995 - September 1998, and (b) August 2006 - July 2009. During both crisis periods, networks estimated by Lasso VAR have substantially fewer connections than the networks estimated by pairwise VAR. During the 2007-2009 crisis, AIG, Bank of America and Goldman Sachs emerge as the three highly connected firms in the three sectors - Insurance (INS), Banks (BA) and primary Broker/Dealer(PB).

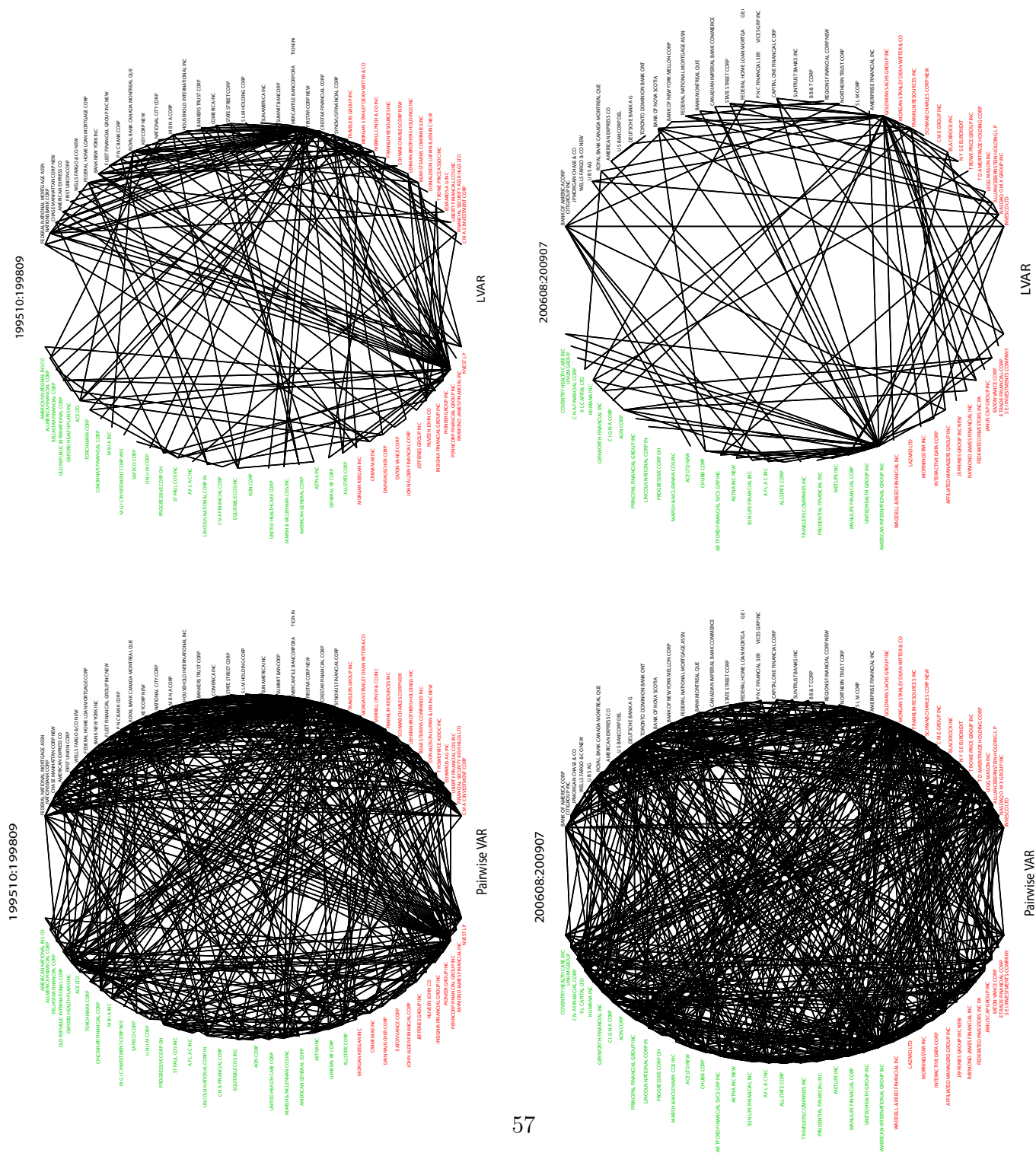


Figure 11: Average degree of LVAR networks based on monthly returns of 75 largest firms, estimated separately for 3-year rolling windows spanning 1990 – 2012. Vertical dotted lines indicate important systemic events. Average degree increases around systemic events, showing higher connectivity among financial institutions.

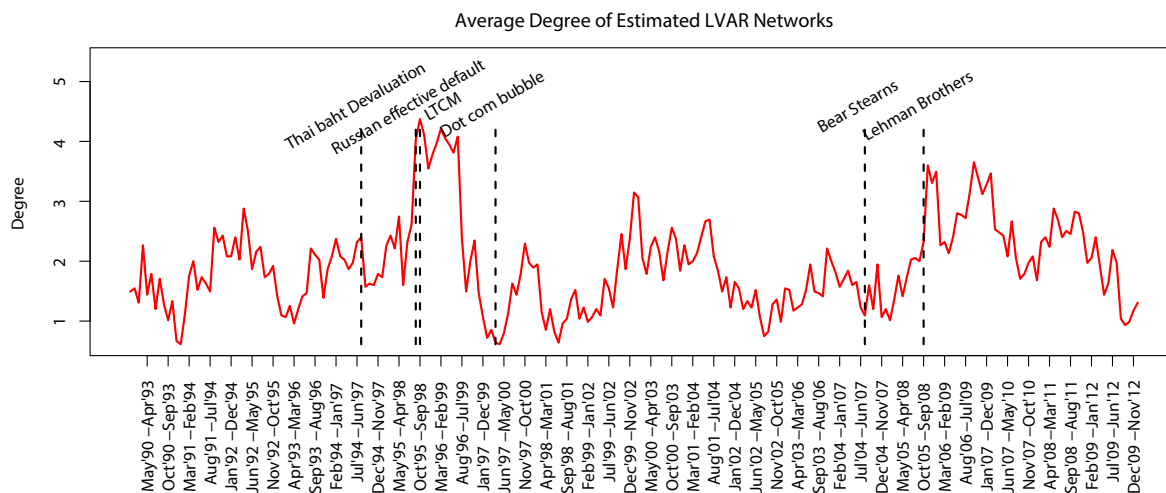


Figure 12: Average closeness of LVAR networks based on monthly returns of 75 largest firms, estimated separately for 3-year rolling windows spanning 1990 – 2012. Vertical dotted lines indicate important systemic events. Average closeness increases around systemic events, showing higher connectivity among financial institutions.

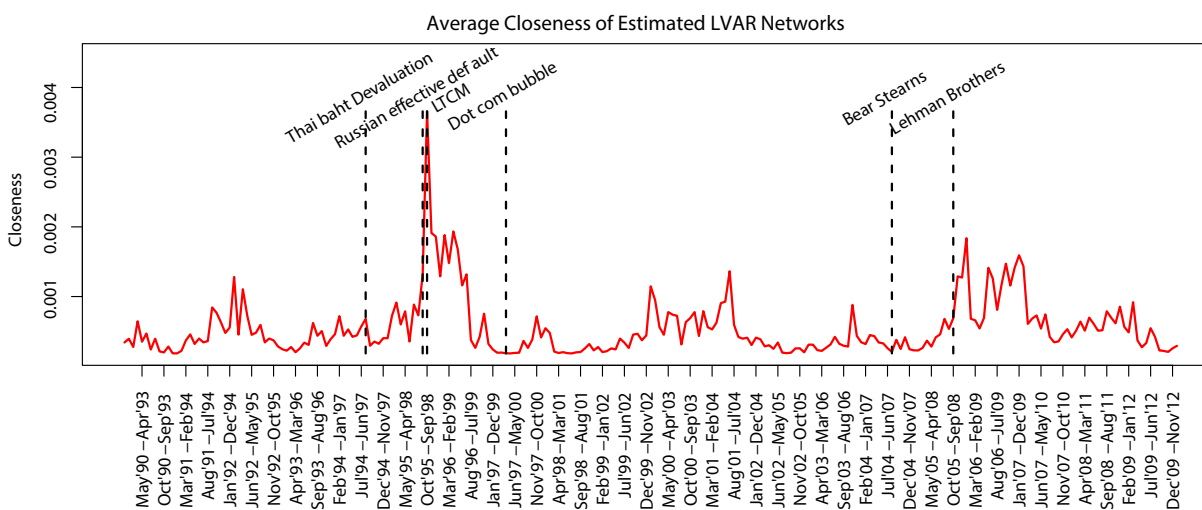


Figure 13: Evolution of average degree of return networks, scaled by their historical average (over 1990 – 2012), for LVAR and pairwise VAR. Around LTCM crisis and Russian effective default, connectivity of LVAR networks increased sharply compared to a pairwise VAR network.

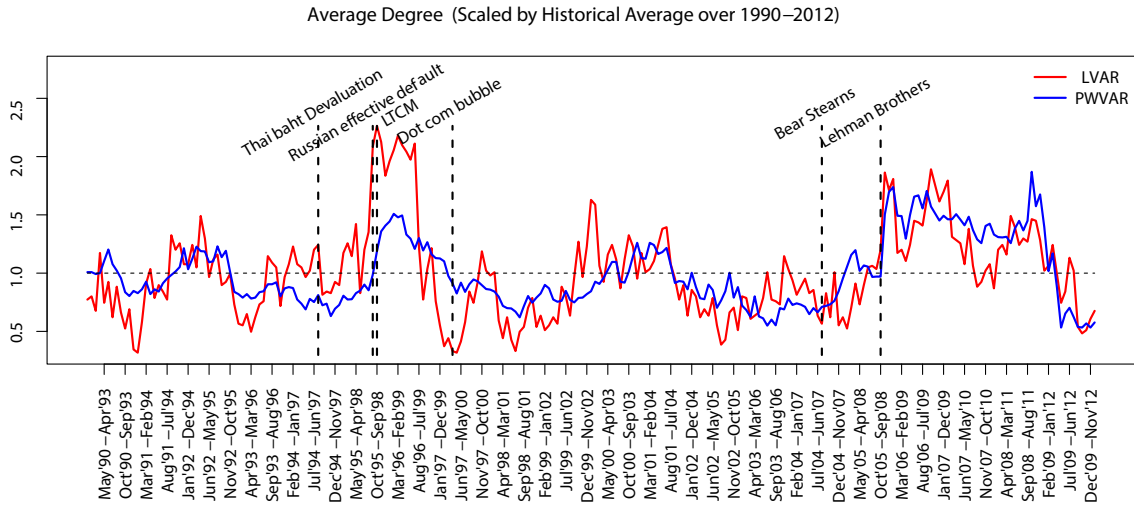


Figure 14: Firms with highest number of connections in LVAR networks, estimated using 3 years of monthly returns. The horizontal axis plots the last month of each window, and the vertical axis displays the degree of a firm, standardized by the mean and standard deviation of degrees of all the firms in the network.

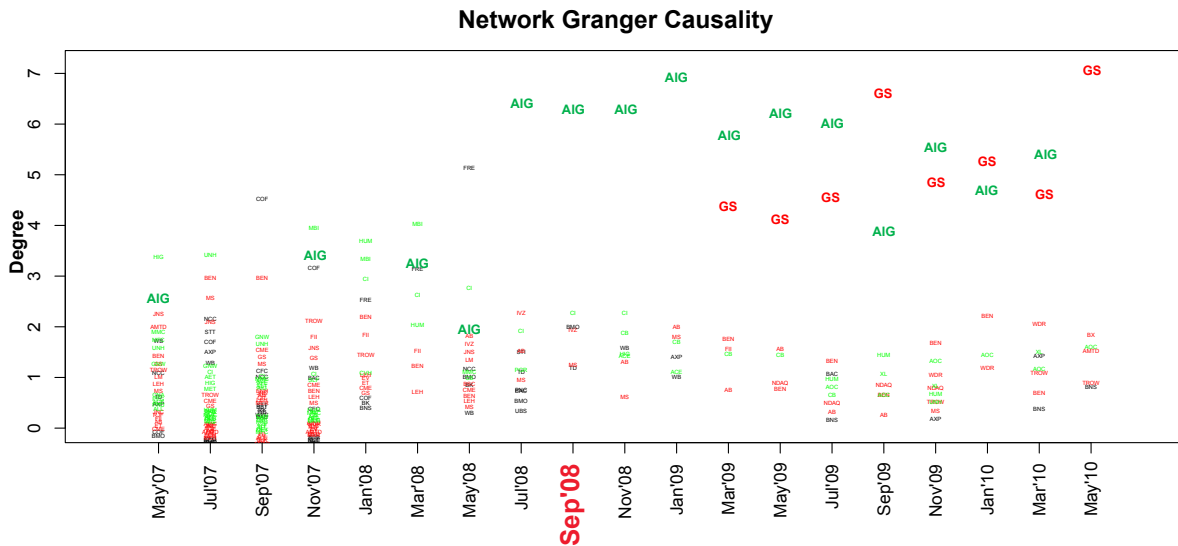


Figure 15: Firms with highest number of connections in pairwise VAR networks, estimated using 3 years of monthly returns. The horizontal axis plots the last month of each window, and the vertical axis displays the degree of a firm, standardized by the mean and standard deviation of degrees of all the firms in the network.

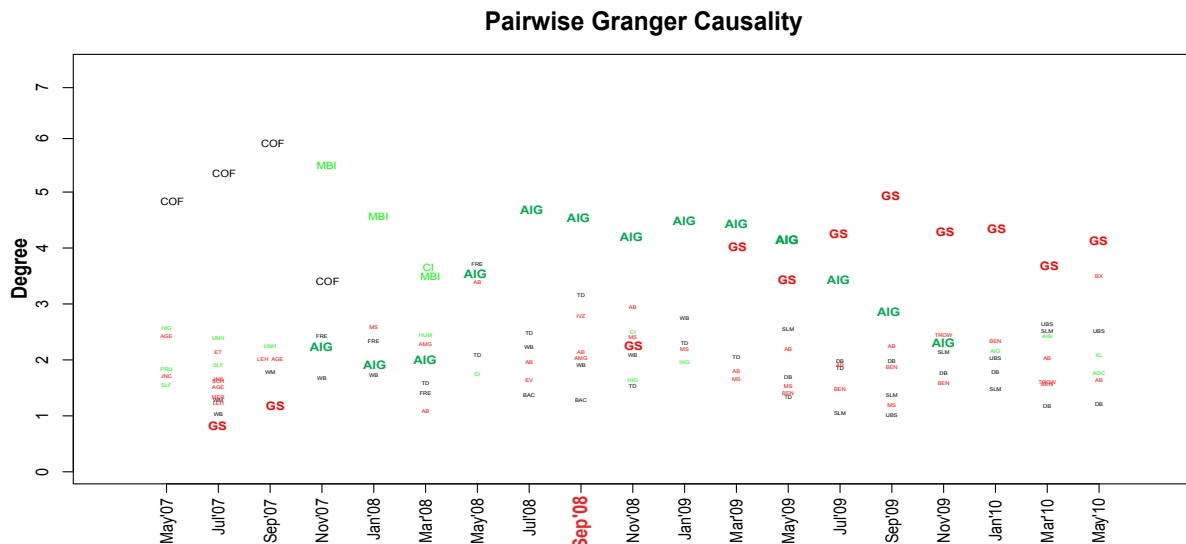


Figure 16: Comparison of within- and between- sectoral connectivities for the Insurance sector in estimated LVAR networks. The lines plot, for each of the three sectors, the total number of connections (edges) with firms in other sectors, as a ratio of the number of edges among firms within the sector.

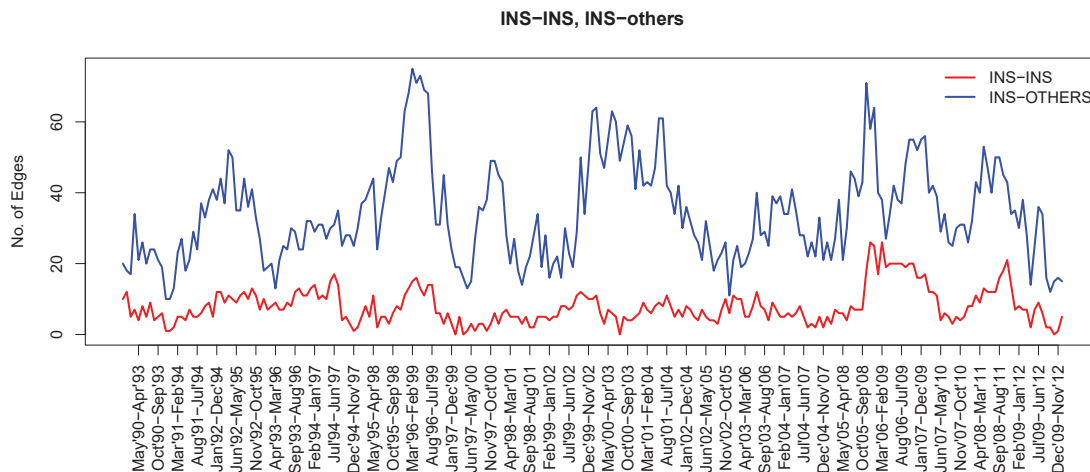


Table 1: Firm Names, Sectors and Ticker Symbols

Ticker	Sector	Firm Name
AB	PB	ALLIANCEBERNSTEIN HOLDING L P
AIG	INS	AMERICAN INTERNATIONAL GROUP INC
AMTD	PB	T D AMERITRADE HOLDING CORP
AOC	INS	AON CORP
AXP	BA	AMERICAN EXPRESS CO
BAC	BA	BANK OF AMERICA CORP
BEN	PB	FRANKLIN RESOURCES INC
BK	BA	BANK OF NEW YORK MELLON CORP
BMO	BA	BANK MONTREAL QUE
BNS	BA	BANK OF NOVA SCOTIA
BX	PB	BLACKSTONE GROUP L P
CB	INS	CHUBB CORP
CFC	BA	COUNTRYWIDE FINANCIAL CORP
CI	INS	C I G N A CORP
CME	PB	C M E GROUP INC
COF	BA	CAPITAL ONE FINANCIAL CORP
FII	PB	FEDERATED INVESTORS INC PA
FRE	BA	FEDERAL HOME LOAN MORTGAGE CORP
GNW	INS	GENWORTH FINANCIAL INC
GS	PB	GOLDMAN SACHS GROUP INC
HIG	INS	HARTFORD FINANCIAL SVCS GRP INC
HUM	INS	HUMANA INC
IVZ	PB	INVESCO LTD
JNS	PB	JANUS CAP GROUP INC
LEH	PB	LEHMAN BROTHERS HOLDINGS INC
LM	PB	LEGG MASON INC
MBI	INS	M B I A INC
MS	PB	MORGAN STANLEY DEAN WITTER & CO
NCC	BA	NATIONAL CITY CORP
NDAQ	PB	NASDAQ O M X GROUP INC
STI	BA	SUNTRUST BANKS INC
TD	BA	TORONTO DOMINION BANK ONT
TROW	PB	T ROWE PRICE GROUP INC
UNH	INS	UNITEDHEALTH GROUP INC
WB	BA	WACHOVIA CORP 2ND NEW
WDR	PB	WADDELL & REED FINANCIAL INC

Table 2: Lehman Brothers Failure Event

Lehman Brothers Neighbors			
	AIG	CIGNA	S&P500 Return
9/10/2008	-4.7%	1.4%	0.6%
9/11/2008	0.3%	1.9%	1.4%
9/12/2008	-30.8%	0.0%	0.2%
9/15/2008	-60.8%	-2.9%	-4.6%
9/16/2008	-21.2%	-3.9%	1.7%
9/17/2008	-45.3%	-6.2%	-4.7%

Panel A: Returns of Most Connected AIG neighbors

	Invesco Ltd	Suntrust Banks	Morgan Stanley	Bank of New York	Citi
9/10/2008	1.7%	-4.9%	-3.7%	1.5%	-1.1%
9/11/2008	1.9%	1.5%	-0.5%	3.4%	-0.4%
9/12/2008	1.0%	3.0%	-3.8%	0.4%	-3.5%
9/15/2008	-6.6%	-3.0%	-13.5%	-8.5%	-15.1%
9/16/2008	4.4%	6.7%	-10.8%	3.2%	3.3%
9/17/2008	-5.9%	-2.7%	-24.2%	-12.3%	-10.9%

Panel B: Returns of Most Connected CIGNA neighbors

	Bank of New York	Regions Financial Corp	Chubb Corp	Ace Ltd New
9/10/2008	1.5%	-2.7%	2.0%	-0.5%
9/11/2008	3.4%	-1.1%	0.3%	0.3%
9/12/2008	0.4%	5.1%	0.4%	0.8%
9/15/2008	-8.5%	-4.0%	-2.1%	1.2%
9/16/2008	3.2%	5.6%	13.5%	8.5%
9/17/2008	-12.3%	-7.4%	-5.0%	-3.0%