Response to Reviewers

1 Response to AE

Two referees reviewed the paper. Both found the paper has potential for AOAS. They raised serious issues. Please address all the questions/issues that were raised.

We would like to thank the AE and the two reviewers for their helpful comments and suggestions. We believe the revised manuscript addresses all the concerns raised by the reviewers. In particular, we have made the following major changes in response to the reviewers’ comments.

1. Contribution over other debiased VAR models. Reviewer 2 raised the point of limited novelty in theory and methodology over other debiased VAR models [Zheng and Raskutti, 2018, Krampe et al., 2018]. In the revised introduction, we clarified the motivation and innovation of our proposed debiased Lasso VAR.

Our objective is to build interpretable financial networks from small sample data sets arising in the analysis of monthly returns (as in Billio et al., 2012). We found that the estimator of the inverse covariance matrix, proposed for de-biasing Lasso in [van de Geer et al., 2014] and also adopted in the above papers, was not powerful enough and led to networks with no edges in most time periods. On the other hand, the debiasing method of Javanmard and Montanari [2014], which explicitly maximizes power in finite samples by solving a convex optimization problem, led to robust and compelling empirical results. This empirical observation motivated us to investigate the asymptotic theory of DLVAR.

The additional theoretical challenge associated with our proposed DLVAR stems from the convex optimization procedure used in [Javanmard and Montanari, 2014], which was not encountered in the above papers since they use analytically more tractable approaches akin to node-wise regression.

In the revision, we also expanded on another methodological aspect of DLVAR: using false discovery rate (FDR) to correct for multiple testing of edge significance in the network. Since finite sample properties of FDR have not been investigated in the VAR context, we added new simulation studies to show that this approach is able to recover the true network structure, as measured by precision, recall and F1 score, in finite sample settings.

2. Expanding the data analysis. Reviewer 2 suggested extending the data period from 2012 to present. In section 4.8 of the revised manuscript, we included a complete analysis describing the evolution of DLVAR networks of large U.S. financial firms in the time period 2012-2021 May.

While there was no major crisis in 2012-2019, our method was able to detect increased connectivity (see Figure 2 in our response to Reviewer 2) starting March 2020, shortly after the COVID-19 shock hit the US financial market. This period coincides with a number of important policies taken by the Federal Reserve Bank. For instance, On March 15, 2020, the Fed announced that it would buy at least $500 billion in Treasury securities and $200 billion in government-guaranteed mortgage-backed securities over the coming months (Source: https://www.brookings.edu/research/fed-response-to-covid19/). These measures were intended to increase liquidity and ensure credit flow in the financial market. So it is not surprising that DLVAR networks became denser around this period.

3. Simulation studies. Reviewer 1 suggested checking the finite sample performance of FDR thresholding in a simulation study. In the revised manuscript, we have conducted a number of simulation studies in Section 3. Since our primary interest is in identifying too-central-to-fail firms, we have investigated the performance of DLVAR in recovering two types of networks (see Figure 1 in our response to Reviewer 1):
(i) networks with isolated components of size 5, each component has a hub node affecting the other 4 nodes;

(ii) a network whose edge set matches a real financial network estimated by DLVAR a month before Lehman Brothers declared bankruptcy (August 2008).

We show that DLVAR achieves promising precision, recall and F1 score, when recovering the true network structure in finite sample settings.

4. Extending asymptotic theory beyond Gaussian VAR(1). Reviewer 2 suggested extending DLVAR beyond VAR(1), and relax the Gaussianity assumption. In the revised manuscript, we noted that constructing a single network by combining different transition matrices in a VAR(d) model is a research question of independent interest, and beyond the scope of this work. However, our asymptotic theory can be extended to handle non-Gaussian VAR models of higher order by leveraging existing results in the literature.

In the revised manuscript (see Appendix B), we have outlined how to extend the asymptotic theory from VAR(1) to VAR(d) models, and also how to relax the Gaussian error assumption to more general subGaussian and subexponential errors. The key observations are the following:

(i) It is well-known in the literature that a p-dimensional VAR(d) process $X_t = \sum_{\ell=1}^{d} A_{\ell} X_{t-\ell} + \varepsilon_t$ can be equivalently expressed as a dp-dimensional VAR(1) process $\tilde{X}_t = A_{\tilde{\ell}} X_{t-1} + \tilde{\varepsilon}_t$. Consistency of DLVAR can be established by assuming regularity conditions on the spectrum the characteristic polynomial $A(z) := I - \sum_{\ell=1}^{d} A_{\ell} z^\ell$, and the inverse covariance matrix of $\tilde{X}_t$.

(ii) As expected, the convergence rates of DLVAR changes from $O\left(\frac{s\log p}{\sqrt{n}}\right)$ when errors have heavier tails than Gaussian. In particular, it changes to $O\left(\frac{s(\log p)^{2\zeta}}{\sqrt{n}}\right)$, where $\zeta = 1/2$ for sub-Gaussian, and $\zeta = 4$ for sub-exponential. Since these rates only depend on the dimension $p$ through its logarithm, the asymptotic validity in high-dimension still remains.

Since the extension to VAR(d), $d > 1$, is not the focus of our network construction methodology, and the main technical ingredients needed for the extension to non-Gaussian VAR already exist in the literature, we chose to keep this discussion in the appendix instead of the main text.

All substantial changes are marked in red in the revised manuscript. We hope the revised manuscript will be suitable for publication in AOAS. We will be happy to provide more information or clarification if needed.

Next, we provide point-by-point responses to the issues raised by the two reviewers.

2 Response to Reviewer 1

1. To employ the debias procedure, we may need good estimator of the precision matrix. To obtain the precision matrix, the authors suggested the procedure (2.5). The condition $\|m\|_1 \leq \mu_1$ may imply the sparsity of the precision matrix. Then we can also use the CLIME (Cai et al., 2011). Is it possible to use the CLIME in the proposed procedure? If so, it would be interesting compare these two procedures. On the other hand, the authors need to investigate and discuss asymptotic behaviors of the proposed method such as consistency and convergence rate.

Thanks for raising this important point. A reasonable estimator $\hat{\Theta}$ of the true precision matrix $\Theta$ is indeed required for the procedure to work. Interestingly, a close inspection of the proofs for high-dimensional inference procedures shows that for bias correction and valid inference, one merely needs the rows of $(\hat{\Theta}\Sigma - I)$ to be sufficiently small. The node-wise regression based estimator $\hat{\Theta}$ proposed in van de Geer et al. [2014] satisfies this property, and we expect CLIME to do so as well.
However, as [Javanmard and Montanari 2014] points out, an optimization criterion can be devised to explicitly find, among all estimators of $\Theta$ satisfying the above properties, one that minimizes the standard errors of the debiased regression coefficients. Therefore, in our small sample setting, we used their de-biasing procedure to boost power.

In our empirical analysis, we experimented with the existing method of [van de Geer et al. 2014], implemented in the R package hdi, but did not recover any significant network edges among US financial firms in most time windows. The approach of [Javanmard and Montanari 2014], however, recovered many interesting network edges. This was expected given the higher power offered by this approach.

Since the CLIME method of constructing precision matrix estimator $\hat{\Theta}$ does not explicitly attempt to improve power in finite samples, we did not pursue this direction. Developing a formal method which will use CLIME along the line of [Javanmard and Montanari 2014] is definitely interesting, but we view this as beyond the scope of this work.

We used the side constraint $\|m\|_1 \leq \mu_1$ only for analytical convenience (see [Loh and Wainwright 2012] for similar assumption), and did not use it in practice. As we clarify in the revised manuscript, for large enough $\mu_1$ the constraint is not binding, so it does not enforce sparsity and the algorithm reduces to the one proposed in [Javanmard and Montanari 2014]. This choice of large $\mu_1$ is different in spirit from CLIME, where $\|m\|_1$ is minimized to find a sparse inverse covariance estimator. We have explained this difference in section 2 of the revised manuscript where we describe the algorithm. See also our response to Reviewer 2 (item 3) for additional clarifications to address the gap between theory and practice.

2. For the convergence rate, there are too many notations. It would be better to discuss each term and provide specific convergence rate under some reasonable cases.

Thank you for this suggestion. Since we study the asymptotics of DLVAR in a regime where $p \to \infty$, it is only fair to assume that the model parameters $A$ and $\Sigma_\varepsilon$ will also change as the sample size $n \to \infty$. Following the standard exposition in high-dimensional statistics, we have chosen to explicitly show how the parameters also affect the convergence rates.

In the revised manuscript, we made the following changes:

(a) We used $O_p(1)$ instead of the full formula in the theoretical analysis whenever possible,
(b) We have added more explanation and interpretations the notations for transition matrix, error variance and sparsity-inducing parameters,
(c) We have added a short discussion under Assumption 2.1. Here, we consider a reasonable case where the eigenvalues of the error variance-covariance $\Sigma_\varepsilon$ and the characteristic polynomial $A(z)$ remain bounded away from 0 and $\infty$. In this scenario, the terms involving $\kappa(\Sigma_\varepsilon)$ and $A$ will not appear in the convergence rates, which will then largely match analogous results for high-dimensional regression with i.i.d. data.

3. They suggested the statistics to conduct the FDR procedure. To verify the proposed procedure, they need to check the finite sample performance via a simulation study.

Thank you for your suggestion. In the revised manuscript (see Section 3), we conducted two simulation studies to assess the finite sample performance of the FDR procedure for VAR network estimation.

In the first study (left panel of Figure 1), we simulated time series from $p = 25, 50$ and $75$ dimensional Gaussian VAR(1) models, corresponding to potential 625, 2500 and 5625 VAR parameters. The network topology consisted of disjoint hubs, each of size 5. We applied DLVAR with 1% and 5% FDR correction, and reported the precision, recall and F1 scores (harmonic mean of precision and recall) in the task of recovering the true networks.
Figure 1: Network topology used in simulation studies. [Left] Network consists of isolated components. Each component has 1 hub node with outgoing edges to 4 other nodes. [Right] Topology of network obtained from real data by DLVAR in the time period September 2005 - August 2008. The large blue node represents AIG, which was detected as a central node in our analysis.

In the second study (right panel of Figure 1), we simulated a $p = 75$ dimensional VAR(1) model whose network skeleton matches the DLVAR network obtained from our real data analysis in the time period 2005 September - 2008 August, which was two weeks before the Lehman Brothers declared bankruptcy. Our results (Section 3 of the revised manuscript) show that as long as the true network is sparse, but the signal strength is large, sample sizes as small as $n = 150$ can provide impressive network recovery in terms of F1 score (close to 90%).

4. For the empirical analysis, they mentioned that they used the institutes which have complete data in the window. What is the meaning of this? You remove the company, when it has some missing data point for any reason? Since they analyze the crisis period, it is important to clarify this.

Thanks for your comment. We have removed a firm from our analysis when it did not have any data after a specific time point in our 36-month time window. Since we analyzed the largest 25 firms in each sector in each time window, there were no incomplete records in the database of their returns. The only cases where missing data appeared were when a firm ceased to exist in the database. For instance, Lehman Brothers declared bankruptcy in September 2008, so we only had this firm in our analysis for the 36-month time windows that ended in August 2008 or earlier.

In the revised manuscript, we have clarified this in section 4.1.

5. (Minor Comment) In Assumption 2., “wher” should be “where”.

Thanks for the correction. We have fixed these errors and completed a thorough revision to avoid typos and grammatical errors.

3 Response to Reviewer 2

The paper is well-written and addresses an important topic of systemic risks. There are however limited contributions in methodology and theory. In particular, DLVAR is an extension of the de-biased Lasso in high-dimensional regression from independent and identically distributed case (van de Geer et al., 2014;
Figure 2: Evolution of degree connectivity of DLVAR networks after the financial crisis. The most prominent peak happens in 2020 March, shortly after the COVID-19 shock hit the stock market.

Javanmard and Montanari, 2014) to the VAR case. The debiased VAR models have been proposed by Zheng and Raskutti (2018); Krampe, Kreiss and Paparoditis (2018), who also provided high-dimensional asymptotic theory. Meanwhile, it provided a good empirical study on the stock markets with a number of interesting empirical findings.

Thanks for your comment on the importance and exposition of our empirical findings. As we explain in our response to the AE (see item 1), the key motivation of this work is to develop a statistical method that is well-suited for learning financial networks from high-dimensional stock returns data. In our empirical analysis of monthly stock returns (originally investigated in Billio et al. [2012]), we have \( p = 75, T = 36 \). In this data set, we find that a DLVAR based on van de Geer et al. [2014] led to mostly empty networks, while the DLVAR based on Javanmard and Montanari [2014] that explicitly maximizes power in finite sample, provided interpretable networks. This empirical finding led us to investigate high-dimensional asymptotics of our proposed DLVAR. The theory needed additional work due to a separate convex optimization step used in Javanmard and Montanari [2014]. This does not arise in Zheng and Raskutti [2018] and Krampe et al. [2018], who adopted a node-wise regression approach proposed in van de Geer et al. [2014].

1. **Strongly suggest extending data period from 2012 to recent.**

Thank you for this suggestion. As we mentioned in our response to the AE (see item 2), in the revised manuscript, we have added a separate subsection (Section 4.8 in the revised manuscript) that monitors the evolution of DLVAR networks from the end of crisis period to this year (Jan 2010 - May 2021).

The majority of this period is not marked by any big systemic event. The only period of substantial distress is in the first quarter of 2020, when the COVID-19 shock hit the financial market. Unlike the financial crisis of 2007-09, which was preceded by a long period of increased linkage during the growth of subprime mortgages, the COVID-19 was truly a short-term exogenous shock and not a result of growing systemic risk. In order to mitigate its impact on the market, the Federal Reserve took a number of regulatory policy that injected capital in the market and increased liquidity to avoid a crash. It is anticipated that the market participants adjusted their positions in response to Fed policy, which could have led to increased linkages in the financial sector.

In our analysis, we see that the DLVAR network started to become dense in March 2020, and its edge density remained at an all time high since the financial crisis of 2007-09 (see Figure 2). Two other periods that exhibit moderate spikes in connectivity span the first two quarters of 2015 and 2019. As we point out in our revised manuscript, these appear a few months after the Federal Reserve’s major policy changes to stop the expansion and explicit tapering of assets in the financial market. To what extent these peaks reflect a delayed reaction to the Fed policy, however, is beyond the scope of this work and remains a topic of future investigation.
2. The current setup is VAR(1) under Gaussian assumption, see Equation (2.1). It will be more interesting to add flexibility of considering VAR(p), and also relax Gaussianity.

Thank you for your comment. As described in our response to the AE (see item 4), we have added Appendix B in the revised manuscript outlining how to extend the asymptotic theory for VAR(d) setting, $d > 1$, and for VAR models with subGaussian and subexponential errors.

The focus on VAR(1) in this work is motivated by our intent to contrast the financial networks with that of [Billio et al., 2012], who also use pairwise VAR(1). For VAR model of higher order ($d > 1$), there is no universal way to aggregate the transition matrices of different orders and build a single network. Two potential directions are to use cumulative impulse response functions or forecast error variance decomposition [Demirer et al., 2018]. Their properties in high-dimension are not well-understood, so we do not pursue this direction in this work. However, as we discuss in the revised manuscript, the asymptotics of DLVAR can be extended to VAR($d$) setting.

The Gaussian assumption can be relaxed to sub-Gaussian and sub-exponential errors following the argument in [Sun et al., 2018]. The main difference from the Gaussian case is that the convergence rate changes from $O\left(\frac{s \log p}{\sqrt{n}}\right)$ to $O\left(\frac{s(\log p)^2 \zeta}{\sqrt{n}}\right)$, where $\zeta = 1/2$ for sub-Gaussian, and $\zeta = 4$ for sub-exponential. Since these rates only depend on the dimension $p$ through its logarithm, validity of our procedure can still be established in a high-dimensional regime.

3. There is an additional side constraint $\|m_1\| \leq \mu_1$ for theoretical results. It is also stated that in empirical analysis, the constraint is not binding for any $\mu_1 \geq 6$. It needs further clarification the possible gap between theoretical and empirical setup.

Thank you for this comment. Reviewer 1 also asked for further clarification of this additional side constraint. In our response to Reviewer 1 (item 1), we have explained the theoretical rationale behind this constraint, and added a discussion in the revised manuscript below equation (2.5).

To address the gap between theory and practice, we provide the following clarification. As Assumption 2.1 and Proposition 2.3 show, consistent estimation is possible in an asymptotic regime where $K_\Theta = \sum_{r,s} |\Theta_{rs}|$ does not grow very fast as $n \to \infty$, and we choose $\mu_1 = O(K_\Theta)$. Since $K_\Theta$ is unknown in practice and we are not tuning over $\mu_1$, we performed a robustness check. We investigated the range of values of $\mu_1$ for which our results remain the same (i.e. the constraint is not binding). Since we see that for $\mu_1$ as small as 6 the results remain the same, it increases our confidence that our findings are reasonably robust to possible misspecification of our assumption on $K_\Theta$.

In the revised manuscript, we have included this explanation in Appendix C: Details of Computation.

4. The choice of $\alpha$. How much difference between Bonferroni and the multiple testing in your case?

As expected, Bonferroni correction for multiple testing led to very few edges in our network compared to the FDR correction proposed by Benjamini and Hochberg (BH). It did not identify many of the systemic events as accurately as the BH procedure. In Figure 3, we compare the evolution of degree connectivity in the post-crisis period using the two methods ($\alpha = 0.2$ in both cases). As we can see, the Bonferroni correction did not detect the increased interlinkages in the market shortly after the COVID-19 shock as accurately as the BH method.

5. Figure 1. Resolution should be increased. Hard to see the name labels Moreover, what is the role of location of firms? Obviously the location of firms changes from top case to bottom case.

In the revised manuscript, we decided to split the figures in two, and moved the networks for the 1998 period to the supplement (see Figure 16). We have also increased the font size. The locations of firms do not have any particular meaning. In each time window, we selected the top 25 firms from each of the three sectors based on their market capitalization. This sometimes led to
Figure 3: Evolution of degree centrality of DLVAR networks in the post-crisis period. The red line corresponds to the Benjamini-Hochberg (BH) procedure controlling the false discovery rate (FDR) at $\alpha = 20\%$. The black line corresponds to Bonferroni correction controlling the family wise error rate (FWER) at $\alpha = 20\%$.

the same firms appearing in different orders in the node set when we constructed networks in different time windows.

6. The visual illustration is not consistent, or at least incomplete, with text description. For example, Figure 2 and Figure 3. In addition to the highlighted extreme events, there are some unlabelled peaks around e.g. May’00-Apr’03, Jan’07-Dec’09 (average degree) and Mar’06-Feb’09 (closeness). What happened there? Meaningful high degree or uninterpretable situation? Moreover there are drops around Jan’97-Dec’99 (labelled as Dot-com)? Why is it so low? Does it mean the model is unable to catch it?

Thank you for this observation. There are indeed peaks and drops in the evolution of network connectivity which do not coincide with known systemic events. Since a detailed investigation into the economic activities during these period is beyond the scope of this paper, we did not highlight these in the original manuscript. However, our inter-sectoral connectivity analysis (section 4.6 and Figure 8 in the revised manuscript) shows that many of the unexplained peaks correspond to increased connectivity between the insurance sector and other sectors, sometimes after big systemic events. In particular,

(a) May’00-Apr’03: this is inside the prominent cycle of increased connectivity starting around 2002 mentioned in the original manuscript. It coincides with the growth of mortgage-backed securities (MBS), and the increase in inter-sectoral connectivity through holdings of these securities as well as interlinkages through insurance contracts.

(b) Jan’07-Dec’09, Mar’06-Feb’09: A visual comparison of the blue and red series in Figure 8 of the revised manuscript suggests that these peaks, appearing a few months after dotcom bubble and the financial crisis, also coincide with periods of unusually high connectivity between the insurance sector and other sectors.

(c) Jan’97-Dec’99: yes, you are right. Our model does not detect increased connectivity among financial firms during the dot com bubble. Since we do not include tech firms in our analysis, it is not clear to us whether this is a limitation of our data set or the proposed method.

Since we have not done an in-depth economic analysis into the insurance sector activities during these periods, we were reluctant to interpret these peaks. However, for the sake of completeness, we have adjusted the text around Figures 3 and 4 to report these descriptive findings.

7. It is stated that “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 …” However the time labels of Figures 9 and 10 do not exactly match the description. Please correct.
Thanks for pointing this out. We have adjusted the text in the revised manuscript as follows:

“During the financial crisis of 2007-09, the average returns started to drop in 2007 and began a recovery in early 2009. Their standard deviations (volatility) also increased consistently during this period. During the Russian financial crisis of 1998, the average returns did not experience significant drop but the standard deviations increased, and remained high throughout the Dotcom bubble (1998-2000) period.”

References


A HIGH-DIMENSIONAL APPROACH TO MEASURE CONNECTIVITY IN THE FINANCIAL SECTOR

By Sumanta Basu1, Sreyoshi Das2, George Michailidis3 and Amiyatosh Purnanandam4

1Department of Statistics and Data Science, Cornell University sumbose@cornell.edu
2Bates White LLC, sdas@bateswhite.com
3Department of Statistics and Informatics Institute, University of Florida, gmichail@ufl.edu
4Ross School of Business, University of Michigan, amiyatos@umich.edu

Data-driven network models to measure systemic risk in the financial sector and identify “too-connected-to-fail” institutions are becoming increasingly common in financial applications. Existing statistical methods for building such networks either take a pairwise approach of fitting many bivariate models, or a system-wide approach of fitting penalized regression models. The former strategy is prone to large false positive selection, while the latter suffers from shrinkage bias and lack of formal inference machinery. These issues are accentuated in small sample, low signal-to-noise settings common in financial data. Building up on recent advances in high-dimensional inference, we propose debiased Lasso Penalized Vector Autoregression (DLVAR), a method for building financial networks that addresses these limitations. Our empirical analysis highlights the importance of de-biasing in a way that increases power of the algorithm in finite samples. We also provide formal inference guarantees of Granger causality tests in high-dimension to justify our method. We apply DLVAR to the stock returns of U.S. large financial institutions covering the period 1990-2021 and illustrate its usefulness in detecting systemically risky periods and institutions, especially during the Great Financial Crisis of 2008-09 and the most recent Covid-19 related market shock.

1. Introduction. After the U.S. financial crisis of 2007-09, there has been increasing interest, amongst policy makers, industry stakeholders and academics, in data-driven methods of modeling the financial system as a network of interacting agents (financial institutions). A key motivation comes from the emphasis on forming macroprudential regulatory policy, wherein the focus is in measuring the risk of not only a single institution, but the combined effect of the potential failure of a handful of institutions on the entire financial system. While information on key channels of risk transmission across institutions (e.g. interbank lending (Allen and Gale, 2000), common asset holding (Bluhm and Krahnen, 2014), fire sales of assets (Acharya et al., 2017)) are often not easy to measure or publicly available, the analysis of co-movements of firm health characteristics (e.g., stock returns, volatilities, leverage or other publicly available financial variables) through a network model (with nodes corresponding to firms and edges encoding some statistical measure of association between them) can provide interesting and useful insights into the dynamic interactions in a highly interlinked financial market. As chronicled in a number of recent empirical papers (Billio et al., 2012; Diebold and Yilmaz, 2014; Demirer et al., 2018; Ahelegbey, Billio and Casarin, 2016a), these data-driven network models tend to become dense around periods of severe financial stress to the system, and also the highly connected nodes in the network correspond to systemically important firms which are deemed as “too-connected-to-fail” in the event of a crisis. Understanding complex network structure among financial markets is also important for policy making. According to a recent annual report from the Financial Stability Oversight Council (FSOC), “regulators should also assess the complex linkages among markets,
examine factors that could cause stress to propagate across markets, and consider potential ways to mitigate these risks" (FSOC, 2019).

Existing statistical and econometric methods to model financial networks using publicly available time series data on individual firms’ characteristics can be broadly divided into two classes. Methods in the first class take a pairwise approach to build a network, wherein a network edge between two firms is present, if some measure of pairwise association (e.g. correlation, mutual information, lead-lag or Granger causality relationship) calculated from their bivariate time series data exceeds a prespecified threshold. Arguably the most popular method in this class is the one proposed in Billio et al. (2012), where an edge from firm $A$ to firm $B$ is present, if the time series of monthly stock returns of $A$ Granger causes the series of firm $B$. Pairwise methods are robust and easy to implement, even when the sample size is small (e.g. Billio et al. (2012) uses data on $T = 36$ months to form a network among $p = 100$ firms), and provides formal inference machinery to select statistically significant network connections. However, the inherent limitation of pairwise methods in distinguishing between direct and indirect effects (a lead-lag relationship between firms $A$ and $B$ can be mediated by firm $C$) renders these methods prone to produce many false positive edges in a network.

The second class of methods, which we will refer to as “system-wide” henceforth, avoids this pitfall by jointly modeling the time series of all the firms in the network and assigns an edge between two firms only if there is a strong association between their time series conditional on all the other time series of firms in the system. Examples of such methods include Barigozzi and Brownlees (2019); Diebold and Yılmaz (2014) and Ahelegbey, Billio and Casarin (2016b) among others. In the context of Granger causality based associations, a popular workhorse model in this class of methods is vector autoregression (VAR) (Lütkepohl, 2005). While Ahelegbey, Billio and Casarin (2016a,b) use off-diagonal entries of the VAR transition matrices to form network edges, Diebold and Yılmaz (2014) uses forecast error variance decomposition obtained from the VAR transition matrices to build network.

These system-wide approaches inherently adjust for direct vs. indirect associations between firms. However, since large VAR models are overparameterized for small to moderate data sets, regularization in the form of sparsity-inducing prior distributions (Ahelegbey, Billio and Casarin, 2016a,b) or penalties (Demirer et al., 2018) are used to fit them. These penalized VAR methods suffer from well-known limitations in the literature of penalized regression: (i) they do not provide tools for frequentist statistical inference, viz. confidence intervals and hypothesis tests for network edges, valid under high-dimensional asymptotics; and (ii) Granger causal effects suffer from shrinkage bias. When the sample size is large enough, these problems can be alleviated by bootstrap-based inference. However, for small sample sizes the issues remain unresolved. Instead of analysis of monthly return data, the above papers have focused primarily on the analysis of daily volatility data, wherein larger segments of stationary observations are available. To the best of our knowledge, there has not been an empirical investigation on the original pairwise monthly data analysis in Billio et al. (2012) using high-dimensional methods. Building financial networks based on small to moderate sample sizes is also of independent interest since financial networks are expected to change frequently in a dynamic environment.

In this paper, we build upon debiased Lasso penalized linear regressions and propose debiased Lasso VAR (DLVAR), a method for constructing financial networks that addresses the above limitations. We investigate properties of DLVAR using double-asymptotic analysis ($p \to \infty$ as $n \to \infty$) and simulation experiments, and conduct an extensive empirical analysis of financial networks based on monthly stock returns of US financial firms as in Billio et al. (2012).

Our first observation is that the choice of approximate inverse covariance matrix when debiasing the Lasso estimates, plays a crucial role in small sample, low-signal settings of monthly returns data. The node-wise regression based approximator proposed in van de Geer et al. (2014) and also adopted in Zheng and Raskutti (2018) and Krampe, Kreiss and Paparoditis (2018) for VAR models, despite its asymptotic optimality, was not powerful enough to recover interesting network patterns in our empirical analysis. However,
the approximator proposed in Javanmard and Montanari (2014), which explicitly maximizes power in finite samples by solving a separate convex optimization problem, provides compelling results and identifies systemically risky institutions.

The convex optimization problem used in Javanmard and Montanari (2014) introduces additional challenges in developing high-dimensional asymptotic theory for time series data. As a theoretical contribution of this work, we provide a formal high-dimensional theory justifying the validity of our proposed debiased Lasso VAR (DLVAR) algorithm. Our asymptotic analysis requires a combination of spectral theory and concentration inequalities for high-dimensional time series, and a martingale central limit theorem to establish this result.

In our empirical analysis, we estimate financial networks using DLVAR on stock returns data of three important sectors of the financial 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. Following the analysis framework of Billio et al. (2012) to compare DLVAR with pairwise VAR, we estimate networks of Granger causal effects on a 3-year rolling basis for the period from 1992 to 2012, study historical evolution of network density and identify highly connected firms in and around periods of systemic stress.

A number of network connectivity measures, including degree (number of firms connected to each other), and closeness (shortest path length from one firm to another), 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. Thus, the constructed network provides insightful information on the buildup of systemic risk in the financial system. As expected, DLVAR provides sparser networks than pairwise VAR by eliminating indirect associations. Yet, it exhibits good sensitivity for identifying interesting and interpretable signals in the data, which was not the case for the debiased VAR procedure based on van de Geer et al. (2014).

In particular, we can rank financial institutions that are relatively more important in the network at any given point in time based on the firm degree and closeness measures. More importantly, we show that the important firms are more clearly separated from others, than the separation observed in pairwise VAR networks. We find that AIG becomes one of the most important nodes in our network before and during the 2008 financial crisis. This finding is consistent with both empirical work and anecdotal evidence that highlights the central role of AIG in the 2008 financial crisis (Mishkin, 2011; Calomiris and Khan, 2015). AIG had taken large exposure to the U.S. subprime housing market through mortgage-backed-securities (MBS) and collateralized-debt-obligations (CDO), as an insurer of these risks. As defaults and foreclosures began to increase in the subprime mortgage market, AIG experienced huge losses and the company was eventually bailed out by the U.S. Treasury. Our estimation is consistent with these developments in the financial markets at the time.

We further provide the ranking of financial institutions at different points in time during our sample period, which can be insightful to industry regulators and other policy decision makers in identifying 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 Brothers. The result makes economic sense since several financial institutions had bought insurance protection from AIG for their exposure to the MBS and CDO products, and Lehman Brothers’ failure directly impacted AIG’s ability to meet those obligations. Our results also show that AIG and its immediate neighbors in DLVAR networks received a large proportion of total bailout money distributed by the Troubled Asset Relief Program (TARP). Overall, these findings show that our estimates are consistent with economic developments during the subprime mortgage crisis.

Finally, we use DLVAR to monitor the evolution of financial networks in the period following the financial crisis and until May 2021. In this analysis, DLVAR identified a sharp increase in network connectivity in March 2020, shortly after the COVID-19 shock in the US market. This increase in linkage is consistent
with the market’s reaction to a number of measures taken by the Federal Reserve Bank in March 2020 to mitigate the COVID-19 shock.

In summary, our analysis provides empirical evidence that suitable bias correction strategies in the literature of high-dimensional statistics can be applied to high-dimensional time series models like VAR for building sparse and interpretable models of financial networks. We also provide asymptotic theory in support of our algorithm under a high-dimensional setting.

2. Model and Method. We start by describing the VAR model and its use in the literature of learning financial networks from publicly available firm level data. Subsequently, we discuss our proposed estimation and inference of Granger causal effects using debiased Lasso VAR (DLVAR), and provide theoretical guarantees under a high-dimensional setting. Finally, we describe our methods for constructing financial networks using DLVAR.

2.1. VAR Models for Learning Financial Networks. We model the joint process of stock returns ¹ of p firms \( X_t = (X_{1t}, \ldots, X_{pt})^\top \) using a \( p \)-dimensional stable, Gaussian VAR(1) model.

\[
(2.1) \quad X_t = AX_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_{\varepsilon}), \quad \Sigma_{\varepsilon} = \text{diag} (\sigma_1^2, \ldots, \sigma_p^2), \quad \sigma_j^2 > 0 \text{ for all } j = 1, 2, \ldots, p.
\]

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 \( V = \{1, 2, \ldots, p\} \) 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 from firm \( j \) to firm \( i \). For ease of presentation, in our empirical analysis we report the undirected, unweighted skeleton of the network \( G \), denoted by \( S(G) \), where there is an edge \( i - j \) between institutions \( i \) and \( j \) if \( A_{ij} \neq 0 \) or \( A_{ji} \neq 0 \).

In this work, we build networks using VAR(1) model and contrast the results with pairwise VAR(1) networks of Billio et al. (2012). While the model can be generalized to VAR(\( d \)) for \( d > 1 \), how to best aggregate information from different lags to form a single interpretable network, is not well-understood. Cumulative impulse response functions (Lütkepohl, 2005) or forecast error variance decomposition (Diebold and Yilmaz, 2014) are potential directions, but their asymptotic properties in high-dimension are not well-understood. However, asymptotic theory of DLVAR can be extended to more general VAR(\( d \)), with \( d > 1 \). The error distribution can also be relaxed to subGaussian or subexponential, in which case the convergence rates change. We discuss these extensions in Appendix B.

Similar to common debiasing procedures in high-dimensional statistics, our DLVAR method works in two steps. In the first step, we fit a Lasso penalized VAR. In the second step, we correct the bias of Lasso VAR and provide asymptotically valid confidence intervals for the entries in the VAR transition matrix. Next, we discuss the methods and theoretical validity of these two steps separately.

2.2. Estimation of Large VAR Models using Lasso. 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 \( \{X_1, X_2, \ldots, X_{n+1}\} \) from the stationary VAR process (2.1), one forms autoregressive design

\[
(2.2) \quad \begin{bmatrix} (X_{n+1})^\top \\ (X_2)^\top \\ \vdots \\ (X_1)^\top \\ \varepsilon_{n+1}^\top \\ \varepsilon_2^\top \\ \vdots \\ \varepsilon_1^\top \\ E \end{bmatrix} = \begin{bmatrix} (X_n)^\top \\ \vdots \\ (X_1)^\top \\ \varepsilon_{n+1}^\top \\ \varepsilon_2^\top \\ \vdots \\ \varepsilon_1^\top \end{bmatrix} A^\top + \begin{bmatrix} (\varepsilon_{n+1})^\top \\ \vdots \\ (\varepsilon_2)^\top \end{bmatrix}
\]

¹In the data analysis, we use the residuals of a GARCH model fitted to the univariate series of returns following Billio et al. (2012). Other suitable transformations can be applied to adjust for heteroskedasticity. Our statistical methodology is general and can be applied on other characteristics of the institutions; e.g., volatilities (after log transformation), leverage ratios etc.
The OLS estimator of the VAR transition matrix $A$ is then obtained by conducting $p$ separate, equation-by-equation OLS regressions to estimate the rows of $A$. To be precise,

$$
\hat{A}_{i:}^{OLS} = \arg\min_{\beta \in \mathbb{R}^p} \frac{1}{n} \| Y_{i:} - X\beta \|^2, \quad \text{for } i = 1, \ldots, p.
$$

In classical, low-dimensional asymptotics ($p$ fixed, $n \to \infty$), $\hat{A}^{OLS}$ is a consistent estimator 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.\(^2\)

In a high-dimensional setting with $n < p$, equation-by-equation estimation (2.3) with OLS is no longer possible. Even for $p < n$, the overall estimation error $\| \hat{A}^{OLS} - A \|^2_F$ is of order $O_p(p^2/n)$, which implies that 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 the assumption that the transition matrices are sparse, the following $\ell_1$ regularized estimation problem offers a solution. Specifically, the equation-by-equation estimator of Lasso VAR is defined as

$$
\hat{A}_{i:} = \arg\min_{\beta \in \mathbb{R}^p} \frac{1}{n} \| Y_{i:} - X\beta \|^2 + \lambda_i \| \beta \|_1, \quad i = 1, \ldots, p,
$$

where $\| \beta \|_1 := \sum_{j=1}^p |\beta_j|$ is the $\ell_1$-norm penalty which encourages sparsity by shrinking smaller coefficients to zero. The tuning parameter $\lambda_i$ controls the sparsity of the $i^{th}$ row of the transition matrix $\hat{A}_{i:}$.

### 2.3. Choice of tuning parameters

In practice, choosing the “best” tuning parameters $\lambda_i$ is cumbersome and depends on the context of the problem. AIC, BIC or Cross-validation (CV) guided choice of $\lambda$ are commonly used, although they are known to perform poorly in high-dimensional ($n \ll p$) problems. In such cases, we adopt a theory-driven, plug-in estimate rather than cross-validation or data-driven strategies. The theoretical choice of $\lambda_i \propto \sigma_i \sqrt{\log p/n}$ (Bühlmann and Van De Geer, 2011) requires knowledge of the error standard deviation $\sigma_i = \sqrt{\text{Var}(\varepsilon_{ij})}$, which is seldom available in practice. Following Javanmard and Montanari (2014), we use the scaled lasso algorithm in Sun and Zhang (2012) to obtain an estimate of $\hat{\sigma}_i$, and choose $\lambda_i = \hat{\sigma}_i \sqrt{\log p/n}$.

### 2.4. Consistency of Lasso VAR in high-dimension

Basu and Michailidis (2015) established that the Lasso VAR estimators are consistent in high-dimensional settings. Here we re-state their main results using our notation, and state some error bounds which will be used to provide formal inference guarantees for debiased Lasso VAR, which is the main theoretical contribution of this work.

Under a double asymptotic framework where both $p, n \to \infty$, $p = O(n^\gamma)$ for any $\gamma > 0$, 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 $\lambda = O_p \left( \sqrt{\log p/n} \right)$, as long as the underlying Gaussian VAR 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 consistently, which is the same as if a priori knowledge of the positions of the $s$ non-zero edges was available and we were only estimating the $s$ free parameters of edge strengths. Thus, for problems where $s \log p \ll p^2$, Lasso VAR can achieve comparable estimation accuracy than OLS with much smaller sample size.

\(^2\)Note that this is different from the approach of Billio et al. (2012), that fits separate bivariate VAR models for different pairs of firms $(i, j)$, $1 \leq i, j \leq p$. 
We define \( p, n \) of parameters on the convergence rates, we will need some additional notations. So we define \( n \rightarrow \infty \) to match the known error bounds in the classical statistics, asymptotic theory of VAR (dimension \( p \) fixed and the model parameters remain unchanged as \( n \rightarrow \infty \)) is established under a stability assumption \( \det(\mathcal{A}(z)) \neq 0 \), i.e. all the eigenvalues of \( \mathcal{A}(z) \) are non-zero for all \( |z| = 1 \) so that \( \mathcal{A}^{-1}(z) \) exists for all \( z \in \mathbb{C}, |z| = 1 \) (Lütkepohl, 2005). We also work under the same stability assumption. However, in a high-dimensional regime where \( p \rightarrow \infty \) as \( n \rightarrow \infty \), the model parameters \( A \) are allowed to change with \( n, p \). In order to capture the effect of these parameters on the convergence rates, we will need some additional notations. So we define

\[
\|\mathcal{A}\| := \max_{|z|=1} \|\mathcal{A}(z)\|, \quad \|\mathcal{A}^{-1}\| := \max_{|z|=1} \|\mathcal{A}^{-1}(z)\|
\]

Note that the stability assumption guarantees that \( \|\mathcal{A}^{-1}\| < \infty \), hence the condition number \( \kappa(\mathcal{A}) := \|\mathcal{A}\||\mathcal{A}^{-1}\| < \infty \). For ease of presentation, we also assume \( \|\mathcal{A}\| > 1 \). Our results hold with \( \|\mathcal{A}\| \) replaced by \( \max \{\|\mathcal{A}\|, 1\} \) in the general case.

**Error Variance:** We define \( \sigma^2_{\text{max}} := \max_{1 \leq j \leq p} \sigma^2_j, \sigma^2_{\text{min}} := \min_{1 \leq j \leq p} \sigma^2_j \). Note that \( \sigma^2_{\text{min}} > 0 \), so that \( \kappa(\Sigma_e) := \sigma^2_{\text{max}}/\sigma^2_{\text{min}} < \infty \). As before, the condition number \( \kappa(\Sigma_e) \) will be used to describe how the rate of convergence of our estimator to the true parameter depends on the error variance \( \Sigma_e \).

**Sparsity:** We assume there are \( s \) non-zero entries in the transition matrix \( A \), i.e. \( \|A\|_0 = s \). In addition, let the inverse covariance matrix of \( X_t \) be \( \Theta := \Sigma^{-1} \) [we omit the subscript \( X \) for ease of presentation]. We denote the \( \ell_1 \)-norm of \( \text{vec}(\Theta) \) as \( K_\Theta := \sum_{1 \leq i, j \leq p} |\Theta_{ij}| \).

The parameter \( s \) controls the sparsity of the true network. The expression \( K_\Theta \) controls a weak form of sparsity of \( \Theta \), i.e. parsimony with respect to the \( \ell_1 \) norm. While we do not need \( \Theta \) to be exactly sparse as in van de Geer et al. (2014), an upper bound on the growth of \( K_\Theta \) is needed to ensure that the de-biasing procedure works asymptotically.

Formally, we develop our asymptotic theory under the following assumption:

**Assumption 2.1.** We consider an asymptotic regime where \( p, n \rightarrow \infty \), with

\[
\kappa^2(\Sigma_e)\kappa^4(\mathcal{A})\|\mathcal{A}\|^2 \max\{K_\Theta^2, s\} \log p/\sqrt{n} \rightarrow 0.
\]

Assumption 2.1 allows the dimension \( p \) to grow with \( n \), as long as \( s \), the number of non-zero parameters in \( A \), and \( K_\Theta^2 \) grow as \( o(\sqrt{n}) \). In particular, if we assume the eigenvalues of \( \Sigma_e \) are bounded away from zero and infinity, the factor \( \kappa(\Sigma_e) \) will be bounded above. If we further assume that the modulus of eigenvalues of \( \mathcal{A}(z) \) are bounded away from zero and infinity uniformly over \( \{z : |z| = 1\} \), the factors \( \kappa(\mathcal{A}) \) and \( \|\mathcal{A}\| \) will be bounded above. Under these two assumptions, the terms involving \( \mathcal{A} \) and \( \Sigma_e \) will not appear in our convergence analysis, and the error bounds in terms of \( s, n \) and \( p \) will match the known error bounds in the literature of high-dimensional regression with i.i.d. data.

Under these assumptions, we state the following result (whose proof follows directly from Basu and Michailidis (2015)) on the consistency of Lasso VAR with an appropriately chosen penalty parameter \( \lambda \).

**Proposition 2.2.** Consider a random realization of \( n \) consecutive observations \( \{X_1, X_2, \ldots, X_{n+1}\} \) from the VAR(1) model (2.1), and consider a Lasso VAR estimator \( \hat{A} \) with a penalty parameter \( \lambda = O_p\left(\sigma^2_{\text{max}}\kappa^2(\mathcal{A})\sqrt{\log p/n}\right) \). Then under Assumption 2.1, for all \( 1 \leq i \leq p \), we have

\[
\|\hat{A}_i - A_i\|_1 = O_p\left(\kappa(\Sigma_e)\kappa^2(\mathcal{A})\|\mathcal{A}\|^2 s \sqrt{\log p/n}\right),
\]
\[ \| \tilde{A}_i - A_i \| = O_P \left( \kappa(\Sigma_\epsilon) \kappa^2(A) \| A \|_2^2 \sqrt{\frac{s \log p}{n}} \right). \]

A proof is given in Appendix A for completeness.

2.5. Bias correction of Lasso VAR estimates. Despite its consistency, Lasso VAR has two limitations when 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) introduced methodology to bias-correct the Lasso estimates and obtain valid confidence intervals. Bias correction of nonlinear estimates is a common technique in classical statistics (Cordeiro and McCullagh, 1991; Cordeiro and Vasconcellos, 1997). One way to motivate such a debiasing procedure is to view the bias correction step as an approximate Newton-Raphson iterate, since \( (1/n) X^\top (Y - X \hat{A}) \) and \( \hat{\Sigma} \) can be viewed as the approximate gradient and Hessian terms, respectively, of the squared error loss function evaluated at the current iterate \( \hat{A}_i \). The classic method of Fisher scoring uses a similar one-step update to obtain a consistent estimate that reduces its variance and makes it statistically efficient (Le Cam, 1956).

For high-dimensional regression problems, the main issue with this approach is that the Hessian is no longer invertible, so one needs to construct a pseudo-inverse of the expected Hessian matrix \( \Sigma \) carefully, so that order of bias remains small asymptotically. 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. The main difference between these procedures is that van de Geer et al. (2014) constructs a pseudo-inverse of Hessian \( M \) using nodewise regression, which is asymptotically a good estimator of \( \Theta \) and is analytically tractable. Javanmard and Montanari (2014) makes the observation that there is no reason to use nodewise regression to find \( M \) as long as \( \| M \hat{\Sigma} - I \|_\infty \) is small, and any such \( M \) will asymptotically correct the bias and allow for inference. Based on this observation, the authors propose to select \( M \) over the feasible set in a way that the length of confidence intervals of individual regression parameters is minimized. This approach explicitly ensures that the power of the algorithm in finite samples is increased. For a more detailed discussion of the intuition behind bias correction, we refer the readers to the comprehensive review article by Dezeure et al. (2015).

We start by constructing the matrix \( M \), which can be viewed as an approximate inverse of the sample covariance matrix \( \hat{\Sigma} = X^\top X/n \). Given two tuning parameters \( \mu, \mu_1 > 0 \), the \( j \)th row of the matrix \( M \), \( 1 \leq j \leq p \), is obtained by solving the following convex program

\[
\min_{m \in \mathbb{R}^p} \quad m^\top \hat{\Sigma} m \\
\text{subject to} \quad \| \hat{\Sigma} m - e_j \|_\infty \leq \mu, \| m \|_1 \leq \mu_1,
\]

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 identity.

As discussed next, the constraint \( \| \hat{\Sigma} m - e_j \|_\infty \) helps ensure that the bias correction is appropriate, and the expression \( m^\top \hat{\Sigma} m \) appears in the standard errors of the debiased Lasso coefficients. Therefore, this optimization explicitly constructs an \( M \) that minimizes the standard error and maximizes power of significance testing.

We introduce the additional side constraint on \( \| m \|_1 \) to make our asymptotic theory tractable, and do not enforce this in practice. This is equivalent to choosing a large \( \mu_1 \) so that this side constraint is not binding and the algorithm is identical to the one proposed in Javanmard and Montanari (2014). Addition of similar side constraints for theoretical analysis is common in the literature of high-dimensional statistics (Loh and
where $\Phi\left(\lambda\right)$ with a large $\mu_1$ is different from the popular inverse covariance estimator CLIME (Cai, Liu and Luo, 2011), where $\|m\|_1$ is explicitly minimized to enforce sparsity in the inverse covariance matrix.

Given the new matrix $M$, the DLVAR estimator is given by

$$A_i^\top = \hat{A}_i^\top + \frac{1}{n}MX^\top(Y_{i,1} - X\hat{A}_i^\top), \quad i \in \{1, \ldots, p\},$$

where $\hat{A}$ is the Lasso penalized VAR estimator.

The estimation error of the debiased Lasso, after rescaling by $\sqrt{n}$, allows the following decomposition, for every $i = 1, \ldots, p$:

$$\sqrt{n}\left(\hat{A}_i^\top - A_i^\top\right) = \frac{1}{\sqrt{n}}MX^\top E_{i,i} + \Delta_i,$$

where $\Delta_i = -\sqrt{n}\left(M\hat{\Sigma} - I\right)(\hat{A}_i^\top - A_i^\top)$.

Suppose the tuning parameters $\lambda$ and $\mu$ are chosen to be of the order $\sqrt{\log p/n}$. In a double asymptotic regime, $p, n \to \infty$ mentioned above, Proposition 2.2 establishes that $\|\hat{A}_i^\top - A_i^\top\|_1$ is $O_P\left(s\sqrt{\log p/n}\right)$. This, together with (2.5), implies that $\|\Delta_i\|_\infty$ is of the order $O_P\left(s\log p/\sqrt{n}\right)$. Hence, the bias term $\Delta$ is asymptotically negligible when $s\log p = o(\sqrt{n})$, and it is possible to conduct inference using only the asymptotic distribution of the individual coordinates of the first term.

### 2.6. Asymptotic distribution of DLVAR.

Next, we derive the asymptotic distribution of individual DLVAR Granger causal estimators under a double-asymptotic regime, $p, n \to \infty$. To this end, note that

$$\sqrt{n}\left(\hat{A}_{ij} - A_{ij}\right) = \frac{1}{\sqrt{n}}MX^\top E_{ij} + \Delta_{ij},$$

for any $j = 1, \ldots, p$. We focus on the asymptotic distribution of the first term.

**Proposition 2.3.** Consider the debiased Lasso estimator (2.6) with a choice of $\lambda$ as described in Proposition 2.2, $\mu = O_P\left(\kappa(\Sigma)\kappa(\lambda)\sqrt{\log p/n}\right)$, $\mu_1 \geq K_{0\theta}$, and $\mu_1 = O_P\left(K_{0\theta}\right)$ as $n, p \to \infty$. Then, for any $1 \leq i, j \leq p$, under Assumption 2.1, we have

$$\sqrt{n}(\hat{A}_{ij} - A_{ij}) = Z_{ij} + O_P(1),$$

where $Z_{ij} \xrightarrow{d} N(0, \sigma_i^2\Theta_{ij})$.

In addition, $m_j^\top\hat{\Sigma}m_j \xrightarrow{p} \Theta_{jj}$ and $\hat{\sigma}_i^2 := \sum_{t=1}^n(X_{it} - X_{i,t-1}\hat{A}_i^\top)^2/n \xrightarrow{p} \sigma_i^2$, so that

$$\frac{\sqrt{n}(\hat{A}_{ij} - A_{ij})}{\hat{\sigma}_i\sqrt{m_j^\top\hat{\Sigma}m_j}} \xrightarrow{d} N(0, 1).$$

The proof is deferred to Appendix A.

### 2.7. Inference with debiased Lasso.

Leveraging the asymptotic negligibility of the bias term in (2.7), we construct $p$-values for testing the statistical significance of the individual edges $A_{ij}$. Formally, for every $i, j \in \{1, \ldots, p\}$, the $p$-value for the test

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

is given by

$$P_{ij} = 2\left[1 - \Phi\left(\frac{\sqrt{n}|\hat{A}_{ij}|}{\hat{\sigma}_i[M\hat{\Sigma}M^\top]_{jj}}\right)\right],$$

where $\Phi(.)$ is the standard normal cdf, and $\hat{\sigma}_i$ is a consistent estimator of error standard deviation $\sigma_i$. 

Wainwright, 2012).
2.8. Network construction with DLVAR. 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 2. An edge is present from node $j$ to node $i$, if $A_{ij}$ is significant at a pre-specified threshold $\alpha > 0$.

The choice of the significance threshold $\alpha$ 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). The False Discovery Rate corresponds to the expected proportion of falsely rejected hypotheses over the total number of rejected hypotheses. Thus, a 10% false discovery rate would imply that, on average, 1 out of 10 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). Finite sample properties of FDR correction have been investigated in the literature largely in the context of regression with i.i.d. data. In order to justify its use for recovering network structure from a VAR transition matrix, we further illustrate the performance of DLVAR under different choices of FDR on simulated from VAR time series in Section 3.

The topology of a weighted, directed network with edges declared as statistically significant at a level $\alpha$ (after correcting for multiple testing), or its undirected, unweighted skeleton $S(G)$, can be explored by standard visualization software or by calculating network centrality measures described in section 4.

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

3. Numerical Experiments with DLVAR. We illustrate the performance of DLVAR on simulated data from VAR models, where the transition matrices encode hub-structures. We also have a simulation experiment with the transition matrix obtained from the real data analysis. For different choices of $n$ and $p$ and FDR threshold, we report precision, recall and F1 score of DLVAR on these simulated data sets.

Precision measures the proportion of true positives among the total number of edges detected by the algorithm. Higher precision implies better control of false positives. Recall, on the other hand, measures the proportion of true edges that are detected by the algorithm. Higher recall implies better control of false negatives. The F1 score is the harmonic mean of precision and recall. It provides a measure of the overall model selection accuracy.
Our simulation results show that the FDR based thresholds help recover the underlying network structure with good accuracy in finite sample settings as long as the underlying VAR model is sparse with sufficiently strong signal. For networks of size $p = 75$ with more than 5000 potential parameters in the transition matrix, we find that DLVAR with FDR thresholds of 1% and 5% can provide F1 scores close to 90% with a sample size as small as $n = 150$.

3.1. Hub Network. We focus on VAR(1) model where the transition matrix corresponds to a network with hubs of 5 nodes, with node 3 affecting nodes 1, 2, 4, 5, node 8 affecting nodes 6, 7, 9, 10, and so on. A sample network topology for $p = 25$ is shown in Figure 1. Such a network topology is embedded in a VAR transition matrix $A$ where $A_{j3} = 0.7$ for $j = 1, 2, 4, 5$, $A_{j8} = 0.7$ for $j = 6, 7, 9, 10$, and so on. The diagonal entries of $A$ are set at 0.6, and the error standard deviations $\sigma_j$ are set at 0.1. For different values of $p = 25, 50, 75$, we simulate time series of different lengths $n = 50, 100, 150$ from a $p$-dimensional Gaussian VAR(1) model, and apply DLVAR with different levels of FDR thresholds.

The precision, recall and F1 score, averaged over 20 replications, are reported in Table 1. Standard deviations over the three metrics over different replicates are displayed inside parentheses. As expected, we observe that for a higher threshold on FDR, the precision (ability to control false positive) is lower, but the recall (ability to detect true positives) is higher. Even though DLVAR estimates potentially a large number of VAR parameters ($625, 2500$ and $5625$ in the three settings), the F1 scores are close to 90% when $n = 150$.

3.2. Data-inspired Network Topology. Next, we simulate VAR time series using a network whose topology matches the graph structure obtained by DLVAR on the real data for the time window September 2005 - August 2008, where AIG emerged as a central node. Figure 1 displays the network topology. We simulated from a Gaussian VAR(1) model where the transition matrix has 0.3 in the diagonals, and 0.8 in its off-diagonal positions that correspond to a network edge.

The results, averaged over 20 replicates, are reported in Table 2. The standard deviations over different replicates are displayed in parentheses. As before, we see that precision is higher for smaller choices of $\alpha$. Further, the F1 scores increase with larger sample size $n$. For $n = 150$, the average F1 scores are close to 90% even though the total number of potential VAR parameters is over 5000.
4. Empirical Results for Financial Firms’ Stock Returns. We use DLVAR to detect the Network Granger Causality structure on firms from three sectors - banks (BA), primary broker-dealers (PB) and insurance companies (INS). Firms from these three sectors, in addition to some hedge funds, were studied by Billio et al. (2012). We excluded the hedge funds from our analysis because the data were not publicly available.

4.1. Data Description and Summary Statistics. We use monthly returns data from January, 1990 to December, 2012 for the above three financial sectors 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 12 and 13 show the mean and standard deviation (in %) of monthly stock returns across different sectors in each 3-year rolling window. During the financial crisis of 2007-09, the average returns started to drop in 2007 and began a recovery in early 2009. Their standard deviations (volatility) also increased consistently during this period. During the Russian financial crisis of 1998, the average returns did not experience significant drop but the standard deviations increased, and remained high throughout the Dotcom bubble (1998-2000) period. 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.

4.2. Network Estimation and Corresponding Connectivity Measures. In order to estimate our network, we consider the Generalized AutoRegressive Conditional Heteroscedaticity (GARCH(1,1)) as our baseline

---

Since we work with the largest 25 firms in each sector in each time window, there are no incomplete records in the database of their returns. The only cases where missing data appear are when a firm cease to exist in the database. For instance, Lehman Brothers declared bankruptcy in September 2008, so we only have this firm in our analysis for the 36-month time windows that ended in August 2008 or earlier.
model for returns of individual firms. This allows us to remove any effect of heteroskedasticity from contaminating our DLVAR measure. 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 \((n = 36 \text{ time points}, p = 75 \text{ firms})\), 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.

\[
R_{i,t} = \mu_i + \sigma_{i,t} \epsilon_{i,t}, \quad \epsilon_{i,t} \sim N(0,1) \\
\sigma_{i,t}^2 = \alpha_0 + \alpha_1 \epsilon_{i,t-1}^2 + \beta_1 \sigma_{i,t-1}^2
\]

We estimate the GARCH(1,1) parameters \(\mu_i, \sigma_{i,t}, \alpha_0, \alpha_1 \text{ and } \beta_1\) for each of the 75 institutions in every time window. Then we fit DLVAR to the estimated GARCH residuals, namely \(\hat{\epsilon}_{i,t} = R_{i,t} - \hat{\mu}_i \hat{\sigma}_{i,t}\) in every window. Our 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 DLVAR model. We also used the debiasing procedure of van de Geer et al. (2014) implemented in the R package hdi to de-bias VAR models, but most of the resulting networks (at the same FDR threshold) across different rolling windows exhibited no statistically significant edges. Therefore, we selected to contrast only the results of DLVAR with the original pairwise VAR of Billio et al. (2012).

4.3. Comparison of Pairwise VAR and DLVAR. 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.

Figures 2 and 16 depict the networks estimated using the pairwise VAR and the DLVAR models for two periods related to two different financial crises, October 1995 - September 1998 and August 2006 - July 2009, respectively. Both types of network plots show high connectivity during these crisis periods. As expected, the pairwise VAR model estimates far denser networks. In comparison, the DLVAR networks are sparse and in the latter crisis period, also identifies AIG and Goldman Sachs as key central nodes. Compared to pairwise VAR, DLVAR allows us to identify highly interconnected firms, such as AIG and Goldman Sachs in a cleaner manner by providing a stronger separation between them and the rest.

4.4. Time Series of Summary Statistics. Next 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.

\[
\text{Degree of node } i = \text{deg}(i) = \text{number of edges adjacent to node } i \\
\text{Closeness of node } i = \frac{1}{\sum_{i \neq j} d(i,j)}
\]

where \(d(i,j) = \text{shortest path length between node } i \text{ and } j\), 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.

\(^{4}\text{FDR threshold of 20\% over different rolling windows corresponded to p-value thresholds between 0.2\% and 1.6\%, which are more conservative than a 5\% significance level used in Billio et al. (2012). See Figure 15 for a comparison.}\)
Fig 2: The DLVAR network is sparser than the pairwise VAR network (window: August 2006 - July 2009). 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).
Figures 3 and 4 depict average degree and closeness, respectively, of our DLVAR 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. Specifically, 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, Goldsmith-Pinkham and Vickery (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. Our inter-sectoral connectivity analysis (see section 4.6 and Figure 8) also shows increased connectivity between insurance sectors and the other two sectors during this period. Two other peaks in Figures 3 and 4, appearing a few months after the Dot com bubble (around 2000 May) and Lehman Brothers collapse (around February 2009), also coincide with similar increase in connectivity between the insurance and the other two sectors.

The time series results provide further evidence that network measures extracted from DLVAR estimated measures are sensible and aid in detecting large systemic events. To contrast them with those obtained from a pairwise VAR network model, Figure 5 depicts 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 5 provides the results. Both models are able to detect the 2008-09 financial crisis; however, the proposed DLVAR model does a much better job around the Russian/LTCM default. It is comforting to see the sharp spike in DLVAR model-based connectivity in periods leading up to both important events during our sample period.
Fig 4: Average closeness of DLVAR 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.

Fig 5: Evolution of average degree of return networks, scaled by their historical average (over 1990 – 2012), for DLVAR and pairwise VAR. Around LTCM crisis and Russian effective default, connectivity of DLVAR networks increased sharply compared to a pairwise VAR network.

As mentioned earlier, a key feature of our model is to separate out relatively stronger connections from the weaker ones and better identify firms that are systemically more important than the others in a stressful situation. We identify such important institutions using our model. Figure 6 shows the list of important firms based on our connectivity measures, and Table 4 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., (degree - average node degree)/(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
Fig 6: Firms with highest number of connections in DLVAR 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.

Fig 7: 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.

see the increasing connectivity of Goldman Sachs from March, 2009 onwards. These estimates line up well with 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 7 reproduces the figure based on pairwise 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.
4.5. Results around the Lehman Brothers Failure Event. We exploit the failure of Lehman Brothers in September 2008 as a shock to the financial system, and use this event to shed light on the usefulness of our DLVAR obtained network in detecting interconnected firms. Recall that on September 10, 2008 Lehman Brothers put itself up for sale, but could not secure a buyer. The U.S. Treasury refused to step in and provide assistance and thus 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 possibility of a government led bailout. We expect firms connected to Lehman to experience large negative returns during this period. That is indeed the case based on our network estimate. Lehman has two direct connections in the network – AIG and Cigna. As shown in Table 3, 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 until 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 illustrates, 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 presence of confounding indirect effects.
Fig 8: Comparison of within- and between-sectoral connectivities for the Insurance sector in estimated DLV AR 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.

4.6. Inter-sectoral Connectivity. DLVAR networks can be used to study both within and across sector connectivity. Ever since the Great Depression of the early 1930s, there has been a number of policy interventions in the banking industry that are primarily motivated by concerns about connections across the banking, broker-dealer, and insurance sectors. A prominent example is the passage of the Glass-Stegall Act in 1933 that prohibited commercial banks from engaging in investment banking activities, such as underwriting of securities or investing in certain class of securities using their own capital or their client’s money. Some of the key provisions of the Act were repealed during our estimation period through the enactment of the Gramm-Leach-Bliley (GLB) Act of 1999. The GLB Act removed barriers between the commercial banks, broker-dealers and the 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, as its details were made public. 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-Stegall 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 linkages between the insurance sector and the other two sectors in Figure 8. The figure illustrates that the insurance sector became more connected with both the broker-dealer and banking sectors in the 1998-1999 period. These results show that our network topology is consistent with the intended consequence of the repeal of the Glass-Stegall Act that increased the connectivity across sectors. Overall our results are consistent with broad changes in the markets and regulations.

4.7. Analysis of Troubled Asset Relief Program (TARP) data. In order to demonstrate potential usage of our DLVAR networks in regulatory decision-making, we conducted a joint analysis of our stock returns data and the data on TARP disbursement. The goal of this analysis is to assess the usefulness of our model in detecting firms that were rescued by TARP intervention. To conduct this analysis, we take the sample of all large firms that appeared at least once in our 36-month rolling window analyses, between January 2005 and December 2008. This results in a sample of 83 firms - 26 banks (BA), 28 insurance firms (INS) and 29 broker-dealers (PB). We take their monthly returns data over the 36-month window from January 2005 to December 2008 to estimate the network structure using our method.
Fig 9: DLVAR Network of 83 large firms, calculated using monthly stock returns in 2005 – 2008. There are strong outgoing edges from AIG and Goldman Sachs to broker-dealers and insurance firms who received bailout packages through TARP after financial crisis (firm names appended by “_1”), but not to TARP recipient banks.

Information on distribution of TARP money were obtained from the Propublica website (https://projects.propublica.org/bailout/list). We manually match the names of 83 firms in our samples to this database, and find that 24 firms have received bailout packages through TARP program. There are 4 broker-dealers and 4 insurance firms in this list, while the remaining 16 are banks.

We build a single DLVAR network from this data set comprising of 83 firms and 48 time points - see Figure 9. In this network, AIG has an outdegree of 23, while Goldman Sachs and Chubb Corp an outdegree of 3, and in addition there are 10 firms with outdegree of 1 or 2. In short, our network provides strong evidence of the central role of AIG in Granger-causing stock returns of other firms of the market. Goldman
Sachs’ connections include amongst others, State Street Corp, Interactive Data Corp and Hartford Financial Services Group, while Chubb Corp’s were AllState, Manulife and Unum Group.

Interestingly, all three broker-dealer firms other than Goldman Sachs who received TARP bailout packages had strong incoming edges from AIG in our network, and so did two out of four TARP recipient insurance firms in our sample (Hartford and Lincoln). However, our network did not predict correctly the banks who received TARP money. Only two of the TARP recipient banks (viz. State Street Corp and Bank of New York Mellon Corp) had strong incoming edges from AIG.

Together, AIG and its recipients (a total of 10 TARP recipient firms with an incoming edge from AIG, with raw p-value $< 0.01$) received $263$ billion, which constitutes $58\%$ of the total bailout money received by $24$ firms in our sample.

In sum, these findings show that our model provides useful economic information. AIG had a central role in the subprime mortgage crisis as the main insurer of risks in the MBS and CDO market. The company received almost $180$ billion of bailout assistance from the U.S. Treasury immediately after the collapse of Lehman Brothers in 2008. Our model correctly picks up the centrality of AIG in the crisis. The model does a relatively poor job in picking up TARP recipients, which is not surprising. One of the objectives of the TARP was to restore public’s confidence in the U.S. financial system. In such a situation, regulators face a tension between targeting bailout assistance to only the most needy institutions and revealing information about the poor financial health of a firm to the market. Sometime it is optimal to provide assistance to a mix of healthy and distressed firms (Philippon and Schnabl, 2013) to avoid panic. Therefore, a richer empirical model, incorporating the regulator’s objectives and private information, is needed to detect the bailout recipients.


In view of the empirically compelling results of DLVAR networks during the financial crisis, we perform a separate analysis to use DLVAR to monitor the evolution of U.S. financial networks during the post-crisis period starting in January 2010 and ending in May 2021. As before, we divide the entire time period into 36-month rolling windows, and study the joint time series of monthly returns of the largest 25 firms from each of the three sectors BA, PB and INS.

In the aftermath of the financial crisis, financial sectors came under increased scrutiny by the regulators. During 2010-2013, several capital requirement and stress testing regulations were passed under the overall framework of the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010. At the same time,
the Federal Reserve Bank injected unprecedented amount of liquidity in the financial sector, making the role of the Fed even more central to the financial markets. Financial markets recovered gradually from the adverse shocks of the financial crisis, and the U.S. financial sector did not experience any major shock till the COVID-19 crisis.

Evolution of mean and standard deviations of monthly returns for different sectors are displayed in Figure 14. The period of the COVID-19 shock in early 2020, followed by a period of significant changes, characterize in these plots a sharp, but transient drop in mean returns, and a sharp, but sustained increase in volatility.

As before, univariate GARCH-fitted residuals are used for model fitting. We used DLVAR with a 20% FDR threshold to build networks in each window. Evolution of the average degree connectivity of networks in this period is shown in Figure 10.

The most prominent peak in degree centrality of the DLVAR network occurs in March 2020. Despite a drop in connectivity around August 2020, the connectivity stayed at an all time high since the financial crisis period. Unlike the financial crisis of 2007 – 09, the connectivity does not increase gradually before the COVID-19 shock. Rather it shows increased linkage during the recovery period starting in March 2020. During this period, the Federal Reserve Bank introduced a number of measures to support the financial
markets, including the purchase of a large amount of securities, to inject liquidity and restore their smooth operation. The increased connectivity of DLVAR networks during this period aligns well with higher co-movement and synchronized activity in the market in response to the Fed policy.

A DLVAR network for the period April 2018 - March 2021 is shown in Figure 11. We observe increased connectivity in the broker-dealer sector, with firms such as Goldman Sachs and E-Trade Financial Corps at the center of the network. Some insurance firms also exhibit high connectivity during this period.

Figure 11 also shows two other periods of moderate increase in connectivity in the DLVAR networks, one around the December 2014 - July 2015, and the other in the December 2018 - July 2019 periods. While there were no major systemic events during those periods, these peaks appear shortly after the Federal Reserve’s policies to stop the expansion and explicit tapering of its balance sheet and investment in the financial markets (see https://www.federalreserve.gov/monetarypolicy/bst_recenttrends_accessible.htm for the growth of total assets of the Federal Reserve). These moves were closely watched by the financial market. To what extent these peaks can be explained as a response to the change in Federal policy, is beyond the scope of this work and is left for future investigation.

5. Conclusion. Statistical methods for building financial networks using large VAR models have so far focused primarily on sparsity inducing priors or penalties. In this work we propose DLVAR, a bias-corrected version of Lasso VAR estimator, that comes with formal statistical inference machinery and produces interpretable networks which can be used to monitor systemic risk build-up in the financial system, and identify risky institutions, as additional analyses related to the Lehman Brothers failure, recipients of TARP money and inter-sectoral connectivity suggest. Our analysis also highlights the importance of choosing a de-biasing step which explicitly aims to maximize power of hypothesis tests in finite sample. Alternative approaches, despite their asymptotic optimality, do not seem to be powerful enough to extract these relationships in our empirical analysis of financial data where signal-to-noise ratio is expected to be low. Our study suggests that high-dimensional approaches with appropriate bias correction are promising candidates for building financial networks because of their ability to substantially reduce false positives in pairwise network models, and yet retain and highlight strong connections among financial institutions.
APPENDIX A PROOFS OF PROPOSITIONS FOR ASYMPTOTICS OF DLVAR

The proofs in this section use asymptotic versions of the non-asymptotic propositions 4.1, 4.2 and 4.3, and the concentration inequalities of Proposition 2.4 in Basu and Michailidis (2015). Note that \( \mu_{\text{max}}(A) \) and \( \mu_{\text{min}}(A) \) in their notation is exactly \( \|A\|^2 \) and \( \|A^{-1}\|^{-2} \) in our notation, \( \Lambda_{\text{max}}(\Sigma) = \sigma^2_{\text{max}}, \) and \( \Lambda_{\text{min}}(\Sigma) = \sigma^2_{\text{min}}. \) Also, note that the spectral density \( f_X(\theta) \) of the VAR(1) process has, for any \( \theta \in [-\pi, \pi], \) the closed form expression \( f_X(\theta) = A^{-1}(e^{i\theta})\Sigma \left(A^{-1}(e^{i\theta})\right)^* \). This implies

\[
\|f_X\| := \max_{\theta \in [-\pi,\pi]} \|f_X(\theta)\| \leq \sigma_{\text{max}}^2 \|\|A^{-1}\||. 
\]

Note that \( \|f_X\| \) is same as \( \mathcal{M}(f_X) \) in the notation of Basu and Michailidis (2015).

For completeness, we provide a version of the key concentration inequality in Proposition 2.4 of Basu and Michailidis (2015) that will be used often in our proofs.

**Proposition 2.1 (Concentration Inequality for Sample Gram Matrix).** Consider a data matrix \( X = [X_1 : \ldots : X_n]^\top \) consisting of \( n \) consecutive observations from a \( p \)-dimensional stable centered Gaussian time series \( \{X_t\}_{t \in \mathbb{Z}}. \) Define the sample Gram matrix \( S = XX^\top/n \) and denote the autocovariance matrix of \( \{X_t\}_t \) by \( \Gamma_X(\ell) = \text{Cov}(X_t, X_{t+\ell}), \) for \( \ell \in \mathbb{Z}. \) Then there exists a constant \( c > 0 \) such that for any vectors \( u, v \in \mathbb{R}^p \) with \( \|u\| \leq 1, \|v\| \leq 1, \) and any \( \eta > 0, \) we have

\[
P \left( \|u^\top(S - \Gamma_X(0))v\| > 6\pi\|f_X\|\eta \right) \leq 6 \exp \left( -cn \min \{\eta, \eta^2\} \right). 
\]

**Proof of Proposition 2.2.** According to Proposition 4.3 of Basu and Michailidis (2015), we can choose \( \lambda = Q(A, \Sigma)\sqrt{\log p/n}, \) where

\[
Q(A, \Sigma) = c_0 \sigma_{\text{max}}^2 \left[ 1 + \|A^{-1}\|^2 \left( 1 + \|A\|^2 \right) \right] < c_0 \sigma_{\text{max}}^2 \left[ 1 + 2\kappa^2(A) \right] = O_P \left( \sigma_{\text{max}}^2 \kappa^2(A) \right).
\]

Note that in the notations of Proposition 4.2 in their paper, \( \omega = c_3 \kappa(\Sigma) \kappa^2(A), \) \( \alpha = \sigma_{\text{min}}^2/\|A\|^2, \) \( \tau = \alpha \max\{\omega^2, 1\} \log p/n, \) so that \( s\tau/\alpha = \max\{\omega^2, 1\} (s \log p/n) \to 0 \) in the proposed asymptotic regime. So the condition \( s\tau/\alpha \leq 32 \) required in that Proposition holds asymptotically.

**Proof of Proposition 2.3.** In order to simplify notations, we use \( \varepsilon \) for \( E_{i; i} = [\varepsilon_{i,n+1}, \varepsilon_{i,n}, \ldots, \varepsilon_{i,2}]^\top \) throughout this proof.

Note that for any \( i, j \in \{1, \ldots, p\}, \) we have

\[
\sqrt{n} \left( A_{ij} - A_{ij} \right) = \frac{1}{\sqrt{n}}m_j^\top X^\top \varepsilon - \sqrt{n} \left( \Sigma m_j - e_j \right)^\top \left( \hat{A}_{ij}^\top - A_{ij}^\top \right).
\]

First, we focus on bounding the second term

\[
\sqrt{n} \left( \Sigma m_j - e_j \right)^\top \left( \hat{A}_{ij}^\top - A_{ij}^\top \right) \leq \sqrt{n} \left\| \Sigma m_j - e_j \right\| \| \hat{A}_{ij} - A_{ij} \|_1
\]

\[
= \sqrt{n} O_P \left( \kappa(\Sigma) \kappa^2(A) \sqrt{\log p/n} \right) O_P \left( \kappa(\Sigma) \kappa^2(A) \|A\|_2^2 \sqrt{\log p/n} \right)
\]

\[
= O_P \left( \kappa(\Sigma) \kappa^4(A) \|A\|_2^2 \sqrt{\log p/n} \right).
\]
Next, we focus on the first term
\[
\frac{1}{\sqrt{n}} e_j^\top M X^\top \varepsilon = \frac{1}{\sqrt{n}} e_j^\top M \Theta X^\top \varepsilon \quad \text{(since } \Sigma \Theta = I) \\
= \frac{1}{\sqrt{n}} e_j^\top (\Sigma - \hat{\Sigma}) \Theta X^\top \varepsilon + \frac{1}{\sqrt{n}} e_j^\top (M \hat{\Sigma} - I) \Theta X^\top \varepsilon + \frac{1}{\sqrt{n}} e_j^\top \Theta X^\top \varepsilon \\
\]
\[\text{(A.2)}\]

The first two terms in this decomposition are of the form \(u^\top \Theta v\), and can be bounded above by \(\|u\|_\infty \|v\|_\infty M\). The third term is a martingale, and we will apply martingale central limit theorem to show its asymptotic normality. We start by deriving upper bounds on \(|\hat{\Sigma} - \Sigma|\).

To control term I, note that \(\|m_j\|_1\) is bounded above by \(\sigma_\max K\). Together, this leads to

\[
\|\hat{\Sigma} - \Sigma\|_\infty \leq \|\hat{\Sigma} - \Sigma\|_\max \|m_j\|_1 = O_p \left( \sigma_\max K \|A\|^{-1} \|\sqrt{\frac{\log p}{n}}\right), \\
\]

\[
\frac{1}{n} \|X^\top \varepsilon\|_\infty = O_p \left( \sigma_\max^2 \kappa^2(A) \sqrt{\frac{\log p}{n}} \right). \\
\]

This implies term I is \(O_p \left( \sigma_\max^4 \kappa^2(A) \|\sqrt{\frac{\log p}{n}}\right). \)

Similarly, to control term II, note that \(\|\hat{\Sigma}m_j - e_j\|_\infty = O_p \left( \kappa(\Sigma\varepsilon) \kappa^2(A) \sqrt{\frac{\log p}{n}} \right)\) by our choice of \(\mu\).

So, term II is at most \(O_p \left( \sigma_\max^2 \kappa(\Sigma\varepsilon) \kappa^2(A) K \sqrt{\frac{\log p}{n}} \right)\).

Now, we consider the third term converges in distribution to a normal random variable asymptotically. We use a martingale central limit theorem argument from Hall and Heyde (1980).

Consider \(\varepsilon_j^\top \Theta X^\top \varepsilon = \sum_{i=1}^n (\Theta_j^\top X_i) \varepsilon_{i,t+1} = \sum_{t=1}^n Z_t = S_n\), say. Let \(\mathcal{F}_n = \{\ldots, X_{n-1}, X_n\}\) denote a filtration for \(n \geq 1\). Note that

\[
\mathbb{E}[S_n - S_{n-1} | \mathcal{F}_n] = 0, \quad \text{for all } n \geq 1, \\
\]

so that \(\{S_n\}\) is a martingale with respect to the filtration \(\{\mathcal{F}_n\}\). Define

\[
V_n^2 := \sum_{t=1}^n \mathbb{E}[Z_t^2 | \mathcal{F}_{t-1}] = \sum_{t=1}^n (\Theta_j^\top X_t)^2 \sigma^2 = n \sigma^2 \Theta_j^\top \hat{\Sigma} \Theta_j, \\
\]

\[
s_n^2 := \mathbb{E}V_n^2 = n \mathbb{E} \left[ (\Theta_j^\top X_t)^2 \right] \sigma^2 = n \sigma^2 \Theta_j^\top \Sigma \Theta_j = n \sigma^2 \Theta_{jj}. \\
\]

According to the discussion in Section 1.7 of Hall and Heyde (1980) on the original result in Brown (1971), under the following two conditions

1. \(s_n^{-2} V_n^2 \xrightarrow{p} 1\),
2. \(s_n^{-2} \sum_{t=1}^n \mathbb{E}[Z_t^2 1(|Z_t| > \delta s_n)] \rightarrow 0\);
we have $S_n/s_n \xrightarrow{d} N(0,1)$, i.e. $(1/\sqrt{n})\sum_{t=1}^n \frac{\Theta_i^T X_t \varepsilon_{t+1}}{\sigma/\sqrt{\Theta_j}} \xrightarrow{d} N(0,1)$.

To verify condition 1, note that Proposition 2.4 in Basu and Michailidis (2015) implies that there exists a universal constant $c > 0$ such that for any $\eta > 0$, we have

$$P\left[\left\| \Theta_j^T \hat{\Sigma} \Theta_j - \Theta_j^T \Sigma \Theta_j \right\| > \sigma_j^2 \right] \leq 2 \exp\left[-cn \min\{\eta, \eta^2\}\right].$$

Therefore, setting $\eta = \sqrt{\log p/n}$, we see that in an asymptotic regime where $p \to \infty$, $n \to \infty$, $\log p/n \to 0$,

$$|s_n^{-2}V_n^2 - 1| = \Theta_j^T \hat{\Sigma} \Theta_j - \Theta_j^T \Sigma \Theta_j = 0.$$

To verify condition 2, note that $W_{1t} := \Theta_j^T X_t / \sqrt{\Theta_j \Theta_j} \sim N(0, 1)$, $W_{2t} := \varepsilon_{t+1} / \sigma \sim N(0, 1)$, and $Z_t = \Theta_j \sigma W_{1t} W_{2t}$. With these notations, and $s_n^2 = \sigma_j^2 \Theta_j$, we have

$$s_n^{-2} \sum_{t=1}^n E [Z_t^2 1(\left\| Z_t \right\| > \delta s_n)] = \frac{1}{n} \sum_{t=1}^n E [W_{1t}^2 W_{2t}^2 1(\left\| W_{1t} W_{2t} \right\| > \delta \sqrt{n})] = E [W_{11}^2 W_{21} 1(\left\| W_{11} W_{21} \right\| > \delta \sqrt{n})] \to 0 \text{ as } n \to \infty,$$

by dominated convergence theorem, since $E[W_{11}^2 W_{21}] < \infty$.

The last part of this proof shows that the denominator $[m_j^T \hat{\Sigma} m_j]^{1/2}$ converges to the correct scaling factor asymptotically, for each $j$. To this end, note that $\hat{\Sigma}$ lies inside the constraint of the convex program with high probability, i.e. $\left\| \hat{\Sigma} - \Theta_j \right\| = O_P\left(\kappa(\Sigma) \kappa^2(A) \sqrt{\log p/n}\right)$. This holds since $E[\varepsilon_k^T \hat{\Sigma} \Theta_j] = 1$ for any $k = 1, \ldots, p$, and Proposition 2.4 in Basu and Michailidis (2015) implies that there exists a universal constant $c > 0$ such that for any $n > \log p$,

$$P\left[\left\| e_k^T \hat{\Sigma} \right\| > 3\sigma_j^2 \left\| \Theta_j \right\| \sqrt{\log p/n}\right] \leq 6 \exp\left[-c \log p\right] \to 0.$$

Now note that $\left\| \Theta_j \right\| \leq \left\| \Theta \right\| = \Lambda_{\min}(\Sigma)$. Using the lower bound in Proposition 2.3 of Basu and Michailidis (2015), we have $\left\| \Theta_j \right\| \leq \left\| A \right\| / \sigma_{\min}^2$. Taking a union bound over all $k = 1, \ldots, p$, we obtain the result.

As a result, this means that $M \neq 1$ with high probability. Now, for any $j = 1, \ldots, p$,

$$\left\| m_j^T \hat{\Sigma} m_j - \Theta_j \right\| = \left\| m_j^T (\hat{\Sigma} m_j - \hat{\Sigma} \Theta_j) + \Theta_j^T (\hat{\Sigma} m_j - e_j) \right\| \leq \left\| m_j \right\|_1 \left(\left\| \hat{\Sigma} m_j - e_j \right\|_\infty + \left\| \hat{\Sigma} \Theta_j - e_j \right\|_\infty \right) + \left\| \Theta_j \right\|_1 \left\| \hat{\Sigma} m_j - e_j \right\|_\infty \leq O_P\left(\kappa(\Sigma) \kappa^2(A) \Lambda_{\min} \sqrt{\log p/n}\right).$$

In particular, this implies $m_j^T \hat{\Sigma} m_j \xrightarrow{P} \Theta_j$, and $\sqrt{n}(\hat{A}_{ij} - A_{ij})/\sigma_i \sqrt{m_j^T \hat{\Sigma} m_j} \xrightarrow{d} N(0, 1)$, for any consistent estimator $\hat{\sigma}_i$ of $\sigma_i$, in the high-dimensional asymptotic regime described in Assumption 2.1.

In order to prove that $\hat{\sigma}_i^2 \xrightarrow{P} \sigma_i^2$, note that

$$\frac{1}{n} \sum_{t=1}^n (X_{it} - X_{i-1} \hat{A}_i)^2 = \frac{1}{n} \left\| X \hat{A}_i^T - X A_i^T \right\|^2 = \frac{1}{n} \left\| X \hat{A}_i^T - X A_i^T \right\|^2 - \frac{2}{n} \left(\varepsilon, X(\hat{A}_i^T - A_i^T)\right) + \frac{1}{n} \left\| \varepsilon \right\|^2.$$

Term I in the above expression is $O_P\left(s \log p/n\right)$. Term II is bounded above by $2 \left\| X^T \varepsilon \right\|_\infty \left\| \hat{A}_i - A_i \right\|_1$, which is $O_P\left(s \log p/n\right)$. Finally, term III converges in probability to $\sigma_i^2$, by the law of large numbers. Putting these together and using Assumption 2.1, the result follows.
APPENDIX B  EXTENDING DLVAR BEYOND GAUSSIAN VAR(1)

In this section, we discuss how to generalize the asymptotic properties of DLVAR for VAR(d) models with \( d > 1 \), and for two types of nonGaussian error structures. A close inspection of the proofs of Propositions 2.2 and 2.3 reveals that the distributions of \( X_t \) and \( \varepsilon_t \) affect the proofs in two ways that needs to be addressed for the extension. First, the key concentration inequality of Proposition 4.2 in Basu and Michailidis (2015), restated in Proposition A.1, need to be adjusted appropriately. Second, the martingale central limit theorem argument needs finite second moment assumptions on \( X_t \) and \( \varepsilon_t \) to apply the dominated convergence theorem. As long as these two steps are addressed, the rest of the proof is deterministic and follows immediately.

So we discuss alternate concentration inequalities and the validity of the second moment assumptions for the two generalized settings.

B.1 Extension from VAR(1) to VAR(d), for \( d > 1 \). Consider the Gaussian VAR(d) model \( X_t = \sum_{\ell=1}^{d} A_\ell X_{t-\ell} + \varepsilon_t \), for some fixed \( d > 1 \) that does not increase with \( n \) and \( p \). Define the characteristic polynomial \( \mathcal{A}(z):= I - \sum_{\ell=1}^{d} A_\ell z^\ell \), for \( z \in \mathbb{C} \). We define \( A := [A_1 : A_2 : \ldots : A_d] \) to be the transition matrix, and denote the overall sparsity using \( s = \|A\|_0 \). The VAR(d) model can be expressed as a VAR(1) model

\[
\begin{bmatrix}
X_t \\
X_{t-1} \\
X_{t-2} \\
\vdots \\
X_{t-d+1}
\end{bmatrix}
= \begin{bmatrix}
A_1 & A_2 & \ldots & A_{d-1} & A_d \\
I & 0 & \ldots & 0 & 0 \\
0 & I & \ldots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \ldots & I & 0
\end{bmatrix}
\begin{bmatrix}
X_{t-1} \\
X_{t-2} \\
X_{t-3} \\
\vdots \\
X_{t-d}
\end{bmatrix}
+ \begin{bmatrix}
\varepsilon_t \\
\varepsilon_{t-1} \\
\varepsilon_{t-2} \\
\vdots \\
\varepsilon_{t-d+1}
\end{bmatrix}
\]

We use \( \Theta \) to denote the inverse of the variance-covariance matrix \( \text{Var}(\bar{X}_t) \), and use \( X := [\bar{X}_n : \ldots : \bar{X}_1]^T \) to denote the updated predictor matrix used in Lasso VAR. With these notations, we use Proposition B.2 in Wilms et al. (2021) to derive the upper bound \( \|f_X\| \leq d\|f_X\| \). The key concentration inequality from Proposition 4.2 in Basu and Michailidis (2015), restated in Proposition A.1 above, then takes the form

\[
P\left[ u^\top (S - \Gamma_X(0)) v > 6\pi d\|f_X\| \eta \right] \leq 6 \exp \left[ -cn \min\{\eta, \eta^2\} \right].
\]

For Gaussian VAR(d), \( \bar{X}_t \) and \( \bar{\varepsilon}_t \) are also Gaussian, so the finite moment assumptions are satisfied. Therefore, in the asymptotic regime defined in Assumption 2.1 with the new definition of \( \Theta \), the rest of the proof of Proposition 2.3 goes through with some additional factors depending on \( d \) in the convergence rates. Since \( d \) does not grow with \( n \) and \( p \), the asymptotic normality of DLVAR remains valid in the regime of Assumption 2.1.

B.2 Extension to subGaussian and subexponential VAR. Asymptotic guarantees for DLVAR can be extended to the case where the errors \( \varepsilon_t \) are subGaussian or subexponential random variables. The additional technical challenges can be remedied using recent advances in the theory of high-dimensional time series. Theoretical properties of VAR for such error distributions have been studied extensively by several authors (Wong, Li and Tewari, 2020; Wu and Wu, 2016; Zheng and Raskutti, 2018). Compared to Gaussian VAR, however, all of these results require some additional decay conditions, such as decay on \( \beta \)-mixing coefficients, spectral norm of transition matrix or a physical dependence parameter. Here we build on our earlier work in Sun et al. (2018), where we make a summability assumption on the entries of the VMA(\( \infty \)) coefficients of the VAR process.

\[
X_t = \sum_{\ell=0}^{\infty} B_\ell \varepsilon_{t-\ell}, \text{ with } \sum_{\ell=0}^{\infty} |B_{\ell,rs}| < \infty \text{ for every } r, s \in \{1, \ldots, p\}.
\]
We consider two different families of error distributions for the components \( \varepsilon_{tr}, 1 \leq r \leq p \):

- (subGaussian): there exists some \( \sigma > 0 \) such that for all \( \eta > 0 \), \( \mathbb{P}(|\varepsilon_{tr}| > \eta) \leq 2 \exp \left(-\eta^2/2\sigma^2\right) \).
- (subexponential): there exist constants \( a, b > 0 \) such that for all \( \eta > 0 \), \( \mathbb{P}(|\varepsilon_{tr}| \geq \eta) \leq a \exp(-b\eta) \).

Both types of random variables have finite fourth moment, which is sufficient for applying the dominated convergence theorem in the proof of Proposition 2.3.

We modify the choice of tuning parameter \( \mu \) in DLVAR by changing \( \sqrt{\log p/n} \) to \((\log p)^\zeta/\sqrt{n}\), where \( \zeta := 1/2 \) for subGaussian (same choice as for Gaussian), and \( \zeta := 4 \) for subexponential.

In addition, in the proofs we replace the key concentration inequality in Proposition 4.2 of Basu and Michailidis (2015), restated in Proposition A.1, by a similar inequality in Proposition 4.2 of Sun et al. (2018) for nonGaussian linear processes. Formally, this inequality (A.1) implies the following:

There are positive constants \( c_i \) not depending on the transition matrix \( A \) such that

\[
\mathbb{P} \left( \left| \mathbf{u}^\top (S - \Gamma(0)) \mathbf{v} \right| > 6\pi n \|f_X\| \right) \leq 2 \exp \left(-cn \min\{\eta, \eta^2\}\right) \quad \text{for subGaussian},
\]

and

\[
\leq c_1 \exp \left(-c_2(\sqrt{n}\eta)^4\right) \quad \text{for subexponential}.
\]

This in turn implies, for any \( R > 0 \), choosing \( \eta = (R \log p)^\zeta/\sqrt{n} \) where \( \zeta = 1/2 \) for subGaussian and \( \zeta = 4 \) for subexponential, we have

\[
\mathbb{P} \left( \left| \mathbf{u}^\top (S - \Gamma_X(0)) \mathbf{v} \right| > 6\pi n \|f_X\| (R \log p)^\zeta/\sqrt{n} \right) \leq c_1 \exp \left(-c_2 R \log p\right).
\]

With the modified choice of \( \mu \), we can then show that

\[
\|\Delta_i\|_\infty = \sqrt{n} O_p \left( (\log p)^\zeta \sqrt{n} \right) O_p \left( s (\log p)^\zeta \sqrt{n} \right) = O_p \left( s (\log p)^{2\zeta} \sqrt{n} \right).
\]

The rest of the argument in our proof of Proposition 2.3 follows verbatim. Thus, in the asymptotic regime

\[
K_2^2(n, \varepsilon)\kappa^2(A)\|A\|^2 \max\{K_2^2, s\} (\log p)^{2\zeta} / \sqrt{n} \to 0,
\]

the asymptotic normality of DLVAR remains valid.

**APPENDIX C DETAILS OF COMPUTATION**

All analyses in this paper were conducted using the statistical software \texttt{R}. We calculated the debiased Lasso using the \texttt{R} codes available on the webpage of the authors of Javanmard and Montanari (2014) at http://web.stanford.edu/~montanar/ssllasso/ with the default choices of tuning parameters. We used the \texttt{R} function \texttt{p.adjust()} to implement the Benjamini-Hochberg procedure for multiple testing corrections. In the empirical analyses, univariate GARCH models were fitted using the \texttt{R} package \texttt{fGarch}.

As Assumption 2.1 and Proposition 2.3 show, consistent estimation is possible in an asymptotic regime where \( K_\Theta = \sum_{r,s} |\Theta_{rs}| \) remains small as \( n \to \infty \), and we choose \( \mu_1 = O(K_\Theta) \). Since \( K_\Theta \) is unknown in practice and we are not tuning over \( \mu_1 \), we performed a robustness check for our empirical analysis in Section 4.1. We investigated the range of values of \( \mu_1 \) for which our results remain the same. We find that for any \( \mu_1 \geq 6 \), the side constraint is not binding and our empirical findings remain the same.
Fig 12: Average monthly return of firms used in the empirical analysis of Section 4 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.

Fig 13: Standard deviation of monthly returns of firms used in the empirical analysis of Section 4 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.
Fig 14: Average (top) and standard deviation (bottom) of monthly return of firms over 3-year rolling windows spanning 2010-2021. 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.

Fig 15: Time series of raw p-value cutoffs corresponding to a 20% FDR cutoff used in the empirical analysis. The cutoffs range from 0.02% to 1.6% over different time intervals, showing a 20% FDR cutoff on FDR is more conservative than a 5% threshold on raw p-values used in Billio et al. (2012)
Fig 16: The DLVAR network is sparser than the pairwise VAR network (window: October 1995 - September 1998).
Table 4
Firm Names, Sectors and Ticker Symbols

<table>
<thead>
<tr>
<th>Ticker</th>
<th>Sector</th>
<th>Firm Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>PB</td>
<td>ALLIANCEBERNSTEIN HOLDING L P</td>
</tr>
<tr>
<td>AIG</td>
<td>INS</td>
<td>AMERICAN INTERNATIONAL GROUP INC</td>
</tr>
<tr>
<td>AMTD</td>
<td>PB</td>
<td>T D AMERITRADE HOLDING CORP</td>
</tr>
<tr>
<td>AOC</td>
<td>INS</td>
<td>AON CORP</td>
</tr>
<tr>
<td>AXP</td>
<td>BA</td>
<td>AMERICAN EXPRESS CO</td>
</tr>
<tr>
<td>BAC</td>
<td>BA</td>
<td>BANK OF AMERICA CORP</td>
</tr>
<tr>
<td>BEN</td>
<td>PB</td>
<td>FRANKLIN RESOURCES INC</td>
</tr>
<tr>
<td>BK</td>
<td>BA</td>
<td>BANK OF NEW YORK MELLON CORP</td>
</tr>
<tr>
<td>BMO</td>
<td>BA</td>
<td>BANK MONTREAL QUE</td>
</tr>
<tr>
<td>BNS</td>
<td>BA</td>
<td>BANK OF NOVA SCOTIA</td>
</tr>
<tr>
<td>BX</td>
<td>PB</td>
<td>BLACKSTONE GROUP L P</td>
</tr>
<tr>
<td>CB</td>
<td>INS</td>
<td>CHUBB CORP</td>
</tr>
<tr>
<td>CFC</td>
<td>BA</td>
<td>COUNTRYWIDE FINANCIAL CORP</td>
</tr>
<tr>
<td>CI</td>
<td>INS</td>
<td>C I G N A CORP</td>
</tr>
<tr>
<td>CME</td>
<td>PB</td>
<td>C M E GROUP INC</td>
</tr>
<tr>
<td>COF</td>
<td>BA</td>
<td>CAPITAL ONE FINANCIAL CORP</td>
</tr>
<tr>
<td>FII</td>
<td>PB</td>
<td>FEDERATED INVESTORS INC PA</td>
</tr>
<tr>
<td>FRE</td>
<td>BA</td>
<td>FEDERAL HOME LOAN MORTGAGE CORP</td>
</tr>
<tr>
<td>GNW</td>
<td>INS</td>
<td>GENWORTH FINANCIAL INC</td>
</tr>
<tr>
<td>GS</td>
<td>PB</td>
<td>GOLDMAN SACHS GROUP INC</td>
</tr>
<tr>
<td>HIG</td>
<td>INS</td>
<td>HARTFORD FINANCIAL SVCS GRP INC</td>
</tr>
<tr>
<td>HUM</td>
<td>INS</td>
<td>HUMANA INC</td>
</tr>
<tr>
<td>IVZ</td>
<td>PB</td>
<td>INVESCO LTD</td>
</tr>
<tr>
<td>JNS</td>
<td>PB</td>
<td>JANUS CAP GROUP INC</td>
</tr>
<tr>
<td>LEH</td>
<td>PB</td>
<td>LEHMAN BROTHERS HOLDINGS INC</td>
</tr>
<tr>
<td>LM</td>
<td>PB</td>
<td>LEGG MASON INC</td>
</tr>
<tr>
<td>MBI</td>
<td>INS</td>
<td>M B I A INC</td>
</tr>
<tr>
<td>MS</td>
<td>PB</td>
<td>MORGAN STANLEY DEAN WITTER &amp; CO</td>
</tr>
<tr>
<td>NCC</td>
<td>BA</td>
<td>NATIONAL CITY CORP</td>
</tr>
<tr>
<td>NDAQ</td>
<td>PB</td>
<td>NASDAQ O M X GROUP INC</td>
</tr>
<tr>
<td>STI</td>
<td>BA</td>
<td>SUNTRUST BANKS INC</td>
</tr>
<tr>
<td>TD</td>
<td>BA</td>
<td>TORONTO DOMINION BANK ONT</td>
</tr>
<tr>
<td>TROW</td>
<td>PB</td>
<td>T ROWE PRICE GROUP INC</td>
</tr>
<tr>
<td>UNH</td>
<td>INS</td>
<td>UNITEDHEALTH GROUP INC</td>
</tr>
<tr>
<td>WB</td>
<td>BA</td>
<td>WACHOVIA CORP 2ND NEW</td>
</tr>
<tr>
<td>WDR</td>
<td>PB</td>
<td>WADDELL &amp; REED FINANCIAL INC</td>
</tr>
</tbody>
</table>
Acknowledgements. We would like to thank Viral Acharya, Francis Diebold, Rob Engle, Giampaolo Gabbi, Andreas Hagemann, Indrajit Mitra, Stefan Nagel, Uday Rajan, Toni Whited, and seminar participants at the 2017 Conference on “Banks, Systemic Risk, Measurement and Mitigation” at Sapienza University, the Bank of International Settlement, the Cornell ORIE Colloquium, the DSMM conference in San Francisco, the FDIC, the International Indian Statistical Association (IISA) Conference, the International Society for Non-Parametric Statistics (ISNPS), the NBER Summer Institute (2017), the Office of Financial Research (OFR), the University of Michigan Finance brown bag, and the Michigan Finance-Economics conference for helpful comments on the paper. All remaining errors are our own. SB was supported in part by NSF grant DMS-1812128 and NIH awards R01GM135926, R21NS120227. GM was supported in part by NSF grants IIS-1632730, DMS-1830175 and DMS-1821220.

REFERENCES


