# Structural Changes in Networks:

## Estimation and Evidence from Financial Institution Connectedness<sup>\*</sup>

Laura Liu<sup>†</sup> University of Pennsylvania

July 31, 2014

### PRELIMINARY DRAFT

#### Abstract

Financial institution networks potentially feature large structural changes over time, which would affect the systemic risk of the whole system. This paper focuses on the Diebold-Yilmaz connectedness measure obtained via variance decomposition, and provides a fused Lasso method to estimate structural changes in the VAR coefficients. To address the high-dimensionality problem along both cross-sectional and time-series dimensions, the fused Lasso estimator penalizes the VAR coefficients as well as their successive differences. I prove that under reasonably general conditions, the proposed method can consistently detect the unknown number of breaks, the estimated break dates are sufficiently close to the true dates, and the estimated coefficients asymptotically converge to the true values. Monte Carlo simulation evidence is presented, along with an application to stock return volatilities of the major financial institutions traded in the U.S. stock market. Results show that structural changes in the interaction pattern are more responsible for the recent financial crisis, while the effects of unfavorable individual shocks are negligible.

<sup>\*</sup>The author thanks Francesco Bianchi, Evan Chan, Xu Cheng, Francis X. Diebold, Francis J. DiTraglia, Thomas A. Lubik, Christian Matthes, Pierre-Daniel G. Sarte, Frank Schorfheide, Minchul Shin, Molin Zhong, and seminar participants at the University of Pennsylvania and the Federal Reserve Bank of Richmond for helpful comments. All remaining errors are my own.

<sup>&</sup>lt;sup>†</sup>Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, USA. Email: yuliu4@sas.upenn.edu.

## 1 Introduction

Networks play a vital role in many microeconomic and macroeconomic studies, e.g. peer effects, technology adoption, input-output structure, financial crises, etc.<sup>1</sup> In the general context of these examples, agents connect to their neighbors and affect each other through the connections, which generates many interesting features while at the same time complicates the estimation procedure.

Moreover, many networks exhibit large structural changes over time. For example, Diebold and Yilmaz (2014b) employ a rolling window estimation and demonstrate that the connectedness of the U.S. financial firm network varies over time and is driven by major macro and financial events. As shown in Elliott *et al.* (2013), different network structures would lead to different spillover patterns and thus affect the stability of the whole system. However, to the best of our knowledge, there has not been much empirical work rigorously addressing structural changes in networks so far. Aiming to fill this gap in the literature, this paper attempts to provide an econometric tool to estimate the structural changes in networks and examine their effect on the systemic risk in the financial sector.<sup>2</sup>

As our empirical context is the financial institution network, we consider an environment with both large cross-sectional dimension N and large time-series dimension T. The reason is that there are many financial institutions in the system, and high frequency financial data can provide daily realized volatility which would yield more precise estimates for the structural break dates. Furthermore, from some empirical evidence, financial crises often start with a sudden outburst (Hatzius *et al.*, 2010; Guo *et al.*, 2011; Diebold and Yilmaz, 2014b); from economic theory, due to uninsured counter-party risk, a large number of banks would run at the onset of a financial crisis (Zawadowski, 2013), which would immediately increase the total connectedness. To that end, we wish to focus on abrupt changes instead of smooth variations in this framework.

We adopt the connectedness measure proposed by Diebold and Yilmaz in a series of papers<sup>3</sup> – a reduced-form approach based on the VAR approximation and the forecast-error variance decomposition, i.e. the (i, j)-th element in the adjacency matrix denotes the fraction of bank *i*'s *H*-step forecast error variance due to shocks to bank *j*. This measure particularly accommodates the financial institution framework that we are interested in. Please see Section 2.1 for a more detailed comparison with other network measures as well as a discussion on the choice of identification methods and forecasting horizon *H*. Once we estimated the network, it is still hard to eyeball each bank's systemicness and vulnerability from either the network graph or the adjacency matrix due to the large cross-sectional dimension. Hence, we introduce some centrality statistics based on the Bonacich centrality, a commonly used centrality measure in social network analysis. It characterizes the relative importance of a node via its direct effect on its immediate neighbors together with its

<sup>&</sup>lt;sup>1</sup>There is a growing literature in related studies. For further reference, please see the book of Jackson (2010).

<sup>&</sup>lt;sup>2</sup>Ahmed and Xing (2009) in the machine learning literature presents a fused-Lasso flavored technique, named temporally smoothed  $L^1$ -regularized logistic regression (TESLA), for recovering the structure of time-varying networks. However, they only take into account undirected networks with binary observables, and the time-series dimension is considerably small with T being around only 10 to 20 in their applications.

<sup>&</sup>lt;sup>3</sup>For a good summary, please refer to the book of Diebold and Yilmaz (2014a).

indirect effect on other nodes (Jackson, 2010).

Now our endeavor has been narrowed down to a time-varying VAR estimation with one salient challenge being the high dimensionality problem. Without structural changes, there are already  $pN^2$ parameters in the VAR coefficient matrix to be estimated,<sup>4</sup> which could be greater than the number of the observations, especially when the cross-sectional dimension N is large. The Least Absolute Shrinkage and Selection Operator (Lasso) is designed to deal with such situations and has proved to be consistent under the sparsity assumption.<sup>5</sup> Considering structural changes, however, the high dimensionality problem becomes even more severe. Unfortunately, the classical methods for multiple structural changes, such as the Bai-Perron procedure.<sup>6</sup> do not fit the current setting where we have a large number of regressors in each equation. Therefore, we resort to a variant of Lasso, the fused Lasso (Tibshirani et al., 2005), which is designed for problems with ordered features and encourages sparsity of both the coefficients and their successive differences via  $L^1$ -penalization. There are several studies, such as Harchaoui and Levy-Leduc (2010), Qian and Su (2013), and Zhang et al. (2013), trying to extend the fused Lasso to the world of structural breaks. However, the existing literature in this area limits to the scenario with a scalar dependent variable and a small number of regressors, so they only penalize the difference in coefficients without penalizing the coefficients themselves, which would cause some finite-sample difficulties when the cross-sectional dimension N is large.

We take the following steps to tailor the fused Lasso for large-scale VARs. First, we estimate the system equation by equation to reduce the computational burden. Second, unlike Qian and Su (2013), we also penalize the coefficients to recover the degree of freedom in each equation. Intuitively, the consistency argument still applies as long as the penalty terms on the coefficients are much smaller than the penalty terms on the successive differences in the expression for the Karush-Kuhn-Tucker optimality condition. Under reasonably general regularity conditions, the proposed method can consistently detect the unknown number of breaks; the estimated break dates are sufficiently close to the true dates; and the estimated coefficients asymptotically converge to the true values. Monte Carlo simulations show that the proposed fused Lasso method works well with a reasonable computational time for the cases where N is around  $10^2$  and T is over  $10^3$ .

Equipped with all these econometric tools developed for structural changes in financial institution networks, we present a pilot application to 61 major financial institutions traded in the U.S. stock market.<sup>7</sup> As high volatilities are particularly associated with panics and crises, we devote our effort to the volatility connectedness with data being constructed as logged daily realized volatility

<sup>&</sup>lt;sup>4</sup>Without loss of generality, here we consider a VAR(p) with zero mean.

 $<sup>^{5}</sup>$ The Lasso-type estimator is first motivated by Tibshirani (1996); Knight and Fu (2000) give the first systematic asymptotic analysis; Yuan and Lin (2006) introduce the group Lasso which selects a group of regressors in or out of the model together; and Zou (2006) proposes the adaptive Lasso which enjoys the oracle properties. In terms of static network estimation, a recent paper by Barigozzi and Brownlees (2014) decomposes the estimated network into a dynamic component and a contemporaneous component and develops a two step Lasso procedure to estimate each part separately. They also show that their Lasso estimator is consistent for both components under some regularity conditions.

<sup>&</sup>lt;sup>6</sup>Perron (2006) provides a comprehensive survey.

<sup>&</sup>lt;sup>7</sup>51 commercial banks, 4 investment banks, 2 credit card companies, and 4 insurance companies.

(RV) spanning from January 2, 2004 to December 31, 2013. It is worth highlighting several major findings. First, as in the counter-party risk theory (Zawadowski, 2013), the financial institution network tends to be more strongly connected with higher average degree and centrality during the crisis, which would significantly increase the systemic risk. Second, we further allow for time-varying heteroskedasticity of the shocks to assess the relative importance of each channel. Results show that the recent financial crisis can be attributed mostly to changes in the interaction pattern rather than the unfavorable individual shocks.

The contribution of this paper is two-fold. Theoretically, the proposed fused Lasso is a useful econometric tool to detect and estimate structural breaks in large-scale VARs. Empirically, the application to the financial institution network documents some interesting facts regarding the late-2000s financial crisis. The rest of the paper is organized as follows. Section 2 introduces the Diebold-Yilmaz connectedness measure as well as various centrality statistics characterizing the systemicness and the vulnerability of a network system. Section 3 proposes the fused Lasso algorithm and analyzes its large sample properties. Section 4 presents the Monte Carlo simulation experiments, and Section 5 applies our method to the 61 major financial institutions and analyzes the empirical findings. Finally, we conclude and discuss the future research in Section 6.

## 2 Network Model

### 2.1 Connectedness Measure

The connectedness measure is introduced by Diebold and Yilmaz in a series of literature (Diebold and Yilmaz, 2012; Demirer *et al.*, 2014; Diebold and Yilmaz, 2014a,b),<sup>8</sup> which is defined according to the variance decomposition of a VAR. The idea behind this approach is to assess the link strength from agent i to agent j as the share of agent j's forecast error variance attributed to the innovations in agent i. It provides a non-structural, model-free technique to infer a latent network from observed panel data.

There are many competing network measures in the literature. For example, Billio *et al.* (2012) rely on the pairwise Granger causality, Bonaldi *et al.* (2013) resort to the VAR coefficient matrix, and Barigozzi and Brownlees (2014) propose the long run partial correlation network. Compared to the others, the Diebold and Yilmaz connectedness index enjoys several main advantages, particularly in the setting of financial institution networks. First, it ensures that all entries in the adjacency matrix are non-negative, which is required by the definition of a network. Second, it represents a weighted directed network, which is a desirable feature for the linkage across financial institutions. Third, even though we implement a sparse VAR estimation, the resulting variance decomposition matrix is still not sparse, which respects the fact that the financial sector is highly integrated. On the contrary, the measure based on the VAR coefficient matrix misses the first and the third points, while the pairwise Granger causality and the long run partial correlation network miss the second one.

Another branch of literature, such as Aldasoro and Angeloni (2013) and Minoiu and Reyes (2013), among others, directly utilize the balance sheet and the capital flow information, which is conceptually different from the Diebold-Yilmaz framework. For instance, the whole system may be alarmed by a bad shock to bank i, so banks become more reluctant to lend to each other. Consequently, every bank's volatility would potentially increase due to lack of risk sharing across banks, and so would the Diebold-Yilmaz measure from bank i to the others. In contrast, the balance sheet network would be less connected as a result of less borrowing and lending.

Specifically, the Diebold-Yilmaz connectedness measure is built on an N-variable  $VAR(p)^9$ 

$$y_t = \sum_{l=1}^p \Phi_l y_{t-l} + \epsilon_t, \ \epsilon_t \stackrel{iid}{\sim} N\left(0, \Sigma\right) \ , \tag{1}$$

which (under the stationarity condition) can be converted to a vector moving average (VMA) representation

$$y_t = \sum_{l=0}^{\infty} \Theta_l \epsilon_{t-l} \; ,$$

<sup>&</sup>lt;sup>8</sup>Please see the original works for background and technical details.

<sup>&</sup>lt;sup>9</sup>Without loss of generality,  $y_t$  is normalized to be mean zero.

where the coefficient matrices  $\Theta_l$  can be deduced recursively as

$$\Theta_l = \sum_{s=1}^p \Phi_r \Theta_{l-s}$$

with  $\Theta_0$  being the  $N \times N$  identity matrix, and  $\Theta_l = 0_{N \times N}$  for any l < 0.

Note that the VAR shocks are not necessarily orthogonal to each other. To perform the variance decomposition, we first need to pick an identification scheme. Diebold and Yilmaz prefer the general variance decomposition (GVD) (Koop *et al.*, 1996; Pesaran and Shin, 1998) to the traditional Cholesky decomposition as the former is invariant to variable ordering.<sup>10</sup> Since GVD assumes normality of the shocks, we take the log transformation of the realized volatility to satisfy such requirement. The (i, j)-th element of the *H*-step GVD matrix represents the part of bank *i*'s *H*-step forecast error variance contributed by shocks to bank *j*, and can be calculated as

$$G_{ij}^{(H)} = \frac{\sigma_{jj}^{-1} \sum_{h=0}^{H-1} (e'_i \Theta_h \Sigma e_j)^2}{\sum_{h=0}^{H-1} (e'_i \Theta_h \Sigma \Theta'_h e_i)}$$

where  $\Sigma$  is the covariance matrix of the shocks  $\epsilon_t$ ,  $\sigma_{jj}$  is the *j*-th diagonal entry of  $\Sigma$ , which is incorporated for scale adjustment, and  $e_i$  is an  $N \times 1$  selection vector with unity as its *i*-th element and zero elsewhere. However, due to the correlated shocks, the row sums of  $G^{(H)}$  are generally not equal to one. Diebold and Yilmaz further normalize each element in  $G^{(H)}$  by its corresponding row sum. Thus, the network adjacency matrix can be defined as the normalized GVD matrix with the diagonal elements being set to zeros to eliminate self-loops.

Meanwhile, another decision to make is the choice of the forecasting horizon H. As H increases, the network will get more and more connected. Especially, when H = 1, (the unnormalized)  $G^{(H)}$ is just the squared contemporaneous correlation among the shocks; and when  $H \to \infty$ , it becomes the unconditional variance decomposition. In the subsequent simulation and application, we choose H = 10 days following the 10-day Value-at-Risk requirement in the Basel Accord.<sup>11</sup>

In summary, the Diebold-Yilmaz framework focuses on empirical characterization of networks "seek[ing] connectedness measures that are informed by financial and economic theory and that help to inform future theory, but that are not wed to a particular theory" (Diebold and Yilmaz, 2014a). This connectedness measure integrates the effects of variable interactions, common factors, and shock structure, in a forecasting setup. The last one is represented by the covariance matrix  $\Sigma$ . The first two are captured by the VAR coefficients  $\Phi$ , but it is hard to distinguish between variable interactions and common factors in the current reduced-form VAR setting without a structural

 $<sup>^{10}</sup>$ In a robustness check, we implement the Cholesky decomposition based on a moderate number of random orderings (10,000) and then examine the resulting distribution of the connectedness measure. Its overall time-varying pattern resembles the GVD result.

<sup>&</sup>lt;sup>11</sup>We have also tried to compute the connectedness by varying H between 5 and 15 days, and found that it is not sensitive to H within this range.

For the simplicity of notation, we drop the superscript (H) from now on.

model.<sup>12</sup>

### 2.2 Centrality Statistics

For networks with large cross-sectional dimension, there would be too much information contained in the network graph or the adjacency matrix, so we need to summarize it into some statistics to better evaluate the relative importance of the nodes. The simplest choice is the traditional degree centrality. In graph theory, the in-degree is defined as the count of arrows coming into an individual node, i.e. the row sum of the adjacency matrix, while the out-degree as the count of arrows coming out of an individual node, i.e. the column sum of the adjacency matrix. However, the in-degree and the out-degree only capture the local direct effect, without considering a node's interaction with the whole system.

Bearing this in mind, we introduce some centrality statistics based on the Bonacich centrality, a widely used centrality measure in social network analysis. It is a generalization of the out-degree, and characterizes the relative importance of a node via not only its direct effect on its immediate neighbors but also its indirect effect on other nodes (Jackson, 2010).<sup>13</sup> Let  $b_i(\eta)$  denote the Bonacich centrality of node *i* with  $\eta$  being the discounting factor, then the Bonacich centrality vector  $b(\eta) = [b_1(\eta), \dots, b_N(\eta)]$ , which can be expressed as

$$b_{i}(\eta) = \sum_{j=1}^{N} \left( G_{ji} + \eta \left( G^{2} \right)_{ji} + \eta^{2} \left( G^{3} \right)_{ji} + \cdots \right) ,$$
  
$$b(\eta) = \iota'_{N} G \left( I_{N} - \eta G \right)^{-1} .$$

where  $I_N$  is an  $N \times N$  identity matrix and  $\iota_N$  is an  $N \times 1$  vector of ones. Intuitively, the centrality of a node depends on not only how many links it has, but also who it links to. Analogous to Bonaldi *et al.* (2013) which is based on a generalized version of the Katz Centrality, we define our systemicness measure as the Bonacich centrality on G and our vulnerability measure as the Bonacich centrality on G' (i.e. the transpose of G). Essentially, the systemicness of bank *i* captures the total effect on the whole system as a consequence of an exogenous shock to bank *i*, and similarly, the vulnerability of bank *i* represents the total effect on bank *i* when there is an exogenous shock to the whole system.

The discounting factor  $\eta$  determines the importance of the indirect effect. In general, the systemicness and the vulnerability increase with  $\eta$ , while being equivalent to the out-degree and the in-degree when  $\eta = 0$ , respectively. Please see Appendix A.1 for an illustrative example. In the subsequent simulation and application, we anchor on  $\eta = 0.9$  for both systemicness and vulnerability so as to gradually discount the remote links. We can expect that with a smaller  $\eta$ , the systemicness

 $<sup>^{12}</sup>$ For instance, Foerster *et al.* (2011) utilize a multisector growth model together with the input-output table to disentangle the sectoral versus the aggregate shocks.

<sup>&</sup>lt;sup>13</sup>We have also checked other popular centrality measures, like the eigenvalue centrality and the Katz centrality, and found fairly similar results.

and the vulnerability will look more similar to the out-degree and the in-degree, respectively.

To compare and contrast these four statistics, we would bring up several points. First, the means of the in-degree and the out-degree should be exactly the same. By definition, they are both equal to the sum of all entries in the adjacency matrix divided by N. Likewise, the means of the systemicness and vulnerability should also be the same.<sup>14</sup> Second, the Diebold-Yilmaz connectedness index normalizes the in-degree to be less than one but impose no restriction on the out-degree. This asymmetry would generally cause the row-sum based statistics, such as in-degree and vulnerability, to be less dispersed than their column-sum based counterparts, i.e. out-degree and systemicness. Last, from the following simulation and application, we can see a comparable pattern between indegree and vulnerability, and between out-degree and systemicness,<sup>15</sup> while the systemicness and vulnerability exhibit more dispersed distribution and larger jump size.

<sup>&</sup>lt;sup>14</sup>Of course, their distributions are different in general.

<sup>&</sup>lt;sup>15</sup>Indeed, all row-based statistics share similar patterns, including those based on the eigenvalue centrality and the Katz centrality. A parallel observation can be drawn among column-based statistics.

### 3 Fused Lasso

### 3.1 The Setting

Within the Diebold-Yilmaz framework, structural changes in network can be reduced to a timevarying VAR. Accordingly, the expression for the standard VAR in equation (1) (in Section 2.1) can be adapted as

$$y_t = \sum_{l=1}^{p} \Phi^{(l,t)} y_{t-l} + \epsilon_t, \ \epsilon_t \sim N\left(0, \Sigma_t\right) ,$$

where  $\epsilon_t$ 's are independent across t, and we allow for a time-vary feature on both the coefficient matrices  $\Phi^{(l,t)}$ 's and the covariance matrix  $\Sigma_t$ . Note that the potential time-varying heteroskedasticity in  $\Sigma_t$  does not interfere with our fused Lasso estimates of the coefficients (see Section 3.3). In this paper, we mainly focus on the homoskedasticity case with  $\Sigma_t$  being constant, while make a preliminary extension to the heteroskedasticity case in the last part of the application, i.e. Section 5.2.2.

As mentioned in the introduction, we are interested in a setting with a large but fixed crosssectional dimension N, a large time-series dimension T that can potentially go to infinity,<sup>16</sup> abrupt changes rather than smooth variations, and relatively sparse structural breaks such that the number of break dates B is much smaller than the total periods T.<sup>17</sup> The main goal of this exercise is to estimate the number of breaks, the break dates, and the coefficient matrices all at once.

The salient challenge here is the super high-dimensional problem. There are potentially  $pTN^2$  parameters with only T observations, which would incapacitates standard estimation methods, such as OLS. What is worse, the classical procedures for multiple structural changes, like Bai-Perron (BP), are not designed for such context either – even if we implement the BP technique equation by equation, there are pN regressors in each equation, which would eventually exceed the maximum number of regressors tabulated in the Bai-Perron critical value table.<sup>18</sup>

#### 3.2 Fused Lasso

To proceed, we resort to a variant of Lasso, the fused Lasso (Tibshirani *et al.*, 2005). However, the existing literature has several limitations. First, only the case with a scalar dependent variable has been considered, though it can be easily extend to a setting with multiple dependent variables. Second, the usual fused Lasso as in Tibshirani *et al.* (2005) penalizes the change in coefficients across regressors instead of across observations. Third, Harchaoui and Levy-Leduc (2010), Qian and Su (2013), and Zhang *et al.* (2013) are the closest to this research utilizing the fused Lasso to

<sup>&</sup>lt;sup>16</sup>Here we distinguish the asymptotic behaviors between N and T because there is some limitation in our current consistency argument (see Section 3.3). A more general case where both  $N, T \to \infty$  is working in progress.

 $<sup>^{17}\</sup>mathrm{We}$  will be more rigorous about this assumption in Section 3.3.

<sup>&</sup>lt;sup>18</sup>We have also tried BP with N = 6 where the critical value table is still valid. Unfortunately, it performs much worse than the fused Lasso. The results are available from the author upon request.

study structural breaks, but they only consider a small number of regressors and thus only need to penalize the difference in coefficients without the need of penalizing the coefficients themselves. When a large cross-sectional dimension N is involved, such estimator's finite-sample performance would deteriorate sharply.

We propose an extension of Qian and Su (2013) to circumvent these obstacles. First, to prevent the RAM from overflow, we conduct the estimation equation by equation. Note that consistency can still be achieved through the equation-by-equation estimation.<sup>19</sup> Beyond the computational benefit, it also promotes the group Lasso consideration as different equations may have different tuning parameter  $\lambda$ 's, which resonate the empirical observation that some banks may be more vulnerable than the others.

Second, to ensure enough degree of freedom in each equation, we add the penalty terms on the coefficients (i.e.  $\lambda_1^i \sum_{t=1}^T \|\beta_t^i\|_1$  in equation (2) below) back to Qian and Su (2013)'s setup. For each equation  $i = 1, \dots, N$ , let  $\{\tau_b^i\}_{b=1}^{B^i}$  be the set of break dates and set  $\tau_0^i = 0$  and  $\tau_{B^i+1}^i = T$ , then for  $l = 1, \dots, p$ , for  $b = 1, \dots, B^i + 1$ , and for  $t = \tau_{b-1}^i + 1, \dots, \tau_b^i$ ,

$$\Phi_{i\cdot}^{(l,t)} = A_{i\cdot}^{(l,b)}$$

where  $M_i$  denotes the *i*-th row of matrix M, and the set of rows  $\left\{A_{i\cdot}^{(l,b)}\right\}_{i=1}^p$  are distinct across b. Denote  $x_t = [y'_{t-1}, \dots, y'_{t-p}]', \beta_t^i = \left[\Phi_{i\cdot}^{(1,t)}, \dots, \Phi_{i\cdot}^{(p,t)}\right]'$ , and  $\alpha_b^i = \left[A_{i\cdot}^{(1,b)}, \dots, A_{i\cdot}^{(p,b)}\right]'$ . The fused Lasso estimator is defined by the following minimization problem:

$$\left\{\tilde{\beta}_{t}^{i}\right\} = \arg\min_{\left\{\beta_{t}^{i}\right\}} \left\{\frac{1}{T} \sum_{t=1}^{T} \left(y_{it} - \beta_{t}^{i\prime} x_{t}\right)^{2} + \lambda_{1}^{i} \sum_{t=1}^{T} \left\|\beta_{t}^{i}\right\|_{1} + \lambda_{2}^{i} \sum_{t=2}^{T} \left\|\beta_{t}^{i} - \beta_{t-1}^{i}\right\|_{2}\right\},$$
(2)

where  $||x||_1 \equiv \sum_i |x_i|$  being the  $L^1$ -norm and  $||x||_2 \equiv \sqrt{\sum_i x_i^2}$  being the  $L^2$ -norm. There are several points worth stressing:

(i) The objective function is convex, thus the global minimizer can be efficiently calculated.

(ii) As Yuan and Lin (2006) pointed out, Lasso treats all parameters separately and thus encourages sparsity; Ridge treats all parameters coherently and thus dissuades sparsity; the group Lasso divides the parameters into groups and thus encourages sparsity at the group level, e.g. in the above fused Lasso framework, each group is composed of the successive difference of the VAR coefficients at each period t.

(iii) Nonetheless, estimation inefficiency and selection inconsistency in the original Lasso would still be a problem, so we can further refine the algorithm with adaptive weights to enhance the performance of our fused Lasso estimator (Zou, 2006; Wang and Leng, 2008). Denote  $\beta_t^i = \left[\beta_{t,1}^i, \cdots, \beta_{t,Np}^i\right]'$ ,

<sup>&</sup>lt;sup>19</sup>The efficiency might be hampered as the actual regressors in each equation might differ (Hsu *et al.*, 2008). We plan to construct some efficient system Lasso algorithms for large time-varying VARs in a separate project. At least, we can always implement a post-Lasso Seemingly Unrelated Regression (SUR) based on the estimated break dates and the selected regressors to alleviate such issue.

and  $\boldsymbol{\beta}_{t} = [\beta_{t}^{1\prime}, \cdots, \beta_{t}^{N\prime}]'$ . Above  $\{\tilde{\boldsymbol{\beta}}_{t}\}$  can be used to construct the adaptive weights:  $w_{1t,j}^{i} = 1/|\tilde{\beta}_{t,j}^{i}|$  for each individual coefficient, and  $w_{2t} = 1/||\tilde{\boldsymbol{\beta}}_{t} - \tilde{\boldsymbol{\beta}}_{t-1}||_{2}$  for the VAR as a whole.<sup>20</sup> Then, the objective function can be written as

$$\left\{\hat{\beta}_{t}^{i}\right\} = \arg\min_{\left\{\beta_{t}^{i}\right\}} \left\{\frac{1}{T} \sum_{t=1}^{T} \left(y_{it} - \beta_{t}^{i\prime} x_{t}\right)^{2} + \lambda_{1}^{i} \sum_{t=1}^{T} \sum_{j=1}^{Np} w_{1t,j}^{i} \left|\beta_{t,j}^{i}\right| + \lambda_{2}^{i} \sum_{t=2}^{T} w_{2t} \left\|\beta_{t}^{i} - \beta_{t-1}^{i}\right\|_{2}\right\}, \quad (3)$$

and the resulting  $\{\hat{\beta}_t\}$  are our adaptive fused Lasso estimates. From now on, we denote the fused Lasso estimates by a tilde and the adaptive ones by a hat.

(iv) In the Bayesian world, the fused Lasso solution can be viewed as the posterior mode of a Time-Varying Parameter VAR (TVP-VAR) with the coefficients centered around zero and following a random walk, where  $\lambda_1$  governs the prior dispersion and  $\lambda_2$  controls the step size. We plan to explore the Bayesian TVP-VAR approach in future research.<sup>21</sup> There are many potential directions, for example, D'Agostino *et al.* (2011) employ a traditional Gibbs sampling for a small-scale VAR, while Koop and Korobilis (2013) exploit the forgetting factor method and make it possible to handle a large-scale VAR.

(v) In Figure 17 in the Appendix, we plot the results with and without the penalty terms on the coefficients for simulation DGP 5 in Section 4. It can be clearly seen that such penalty terms are indeed indispensable in the current large-N setting. Without them, the estimated results can still pick up the breaks, but paths are quite off due to lack of regularization on the coefficients. Thanks to these additional penalty terms, we are now able to obtain good estimates for both the breaks and the parameters.

### 3.3 Asymptotic Properties

The following asymptotic analysis is built on Qian and Su (2013)'s setup. Once again we examine the system of VAR equation by equation and suppress the index *i* below when there is no confusion. Define  $\mu_{\min}(M)$  and  $\mu_{\max}(M)$  as the smallest and the largest eigenvalues of matrix *M*. Also define the *b*-th interval length  $I_b = \tau_b - \tau_{b-1}$ , the *b*-th jump  $J_b = \alpha_{b+1} - \alpha_b$ , the minimum interval length  $I_{\min} = \min_{1 \le b \le B+1} I_b$ , and the minimum jump size  $J_{\min} = \min_{1 \le b \le B} ||J_b||_2$ . In addition, denote  $\mathbb{D} = \text{diag}\left(\sqrt{I_1}\mathbb{I}_{Np}, \cdots, \sqrt{I_{B+1}}\mathbb{I}_{Np}\right)$  where  $\mathbb{I}_{Np}$  is the  $Np \times Np$  identity matrix,  $\mathbb{V} = \text{diag}\left(\sigma_{i1}^2, \cdots, \sigma_{iT}^2\right)$  where  $\sigma_{it}^2 = \text{Var}\left(\epsilon_{it}\right)$ ,  $\mathbb{X} = \text{diag}\left(\mathbb{X}_1, \cdots, \mathbb{X}_{B+1}\right)$  where  $\mathbb{X}_b = [x_{\tau_{b-1}+1}, \cdots, x_{\tau_b}]', \alpha = [\alpha'_1, \cdots, \alpha'_{B+1}]', \Psi = \text{plim}\frac{1}{T}\mathbb{X}'\mathbb{X}$ , and  $\Phi = \text{plim}\frac{1}{T}\mathbb{X}'\mathbb{V}\mathbb{X}$ .

 $<sup>^{20}</sup>$ We choose separate weights for the coefficients as there is no particular evidence on coefficient grouping, and joint weights for the successive changes as it is more relevant to have common break dates in the network. A computer science paper, Ohlsson *et al.* (2010), has also imposed adaptive weights on the successive differences to attain better estimation results, but there might be some theoretical difficulties to obtain its large sample properties (see Qian and Su (2013)).

<sup>&</sup>lt;sup>21</sup>A desirable feature of the Bayesian approach is that it can easily produce credible intervals for the estimates and the likelihood-ratio statistic for the tuning parameters.

The choice of  $\lambda_1$ 's would be relatively straight-forward since the consistency argument would still hold as long as  $\lambda_1$ 's satisfies Assumption 6 (see Appendix B). Things get more complicated for  $\lambda_2$ 's, which we select via the following information criterion (IC)

$$IC(\lambda_2) = \log\left(\check{\sigma}^2\right) + \rho_T \cdot \#\left(\text{nonzero coefficients in } \{\check{\alpha}_b\}_{b=1}^{B+1}\right) , \qquad (4)$$

where  $\check{\sigma}^2$  and  $\{\check{\alpha}_b\}_{b=1}^{B+1}$  are obtained by the post-Lasso regression. Paralleling the argument in Qian and Su (2013), the IC structure above plays an essential role – without it, we can only achieve that  $P(\hat{B} \geq B) \to 1$  as  $T \to \infty$ , i.e. the number of breaks may be overestimated.

Based on the assumptions in Appendix B, we are able to establish the following propositions regarding the fused Lasso estimation.

Proposition 1. (Fused Lasso)

Under Assumptions 1-6, as  $T \to \infty$ , (i)  $P\left(\tilde{B}=B\right) \to 1$ ; (ii)  $P\left(\max_{1 \le b \le B} |\tilde{\tau}_b - \tau_b| \le T\delta_T\right) \to 1$ ; (iii) Post-Lasso estimates  $\tilde{\mathbb{D}}\left(\tilde{\alpha} - \alpha\right) \xrightarrow{d} N\left(0, \Psi^{-1}\Phi\Psi^{-1}\right)$ .

If some uniformity on the DGP is assumed, we could obtain the limiting distribution of the estimated break dates.

**Proposition 2.** (Fused Lasso - Limiting Distribution of Estimated Break Dates)

Under Assumptions 1-7, as  $T \to \infty$ ,  $(\Delta'_b \Psi_b \Delta_b) I_0^2 (\tilde{\tau}_b - \tau_b) \xrightarrow{d} \arg \max_s Z_b(s)$  for each  $b = 1, \dots, B+1$ , where

$$Z_b\left(s\right) = \begin{cases} \sqrt{\frac{\Delta_b' \Phi_b \Delta_b}{\Delta_b' \Psi_b \Delta_b}} W_{b,1}\left(|s|\right) - \frac{|s|}{2}, & \text{if } s < 0\\ \sqrt{\frac{\Delta_b' \Phi_{b+1} \Delta_b}{\Delta_b' \Psi_b \Delta_b}} W_{b,2}\left(|s|\right) - \frac{\Delta_b' \Psi_{b+1} \Delta_b}{\Delta_b' \Psi_b \Delta_b} \frac{|s|}{2}, & \text{if } s \ge 0 \end{cases}$$

where  $W_{b,j}(s)$ 's are independent Wiener processes across (b, j).

The sketch proof for the propositions can be found in Appendix B. Intuitively, the asymptotic argument in Qian and Su (2013) remains valid as long as the penalty terms on the coefficients are much smaller than the penalty terms on the successive differences in the expression for the Karush-Kuhn-Tucker optimality condition. In the following simulation and application, we can also see that the optimal  $\lambda_1$ 's are much smaller than the corresponding  $\lambda_2$ 's.

There are several caveats in interpreting above asymptotic statements:

(i) Here we obtain the consistency and the asymptotic normality for  $\{\hat{\alpha}_b\}_{b=1}^{B+1}$  but not for  $\{\hat{\beta}_t\}_{t=1}^T$  – the fraction of misplaced observations shrinks to zero, but the number of misplaced observations may goes to infinity at a rate slower than  $T\delta_T$ .

(ii) In the current fused Lasso setup, although we are able to achieve the consistency and obtain the limiting distribution for the estimates of the break dates and the coefficients  $(\{\hat{\alpha}_b\}_{b=1}^{B+1})$ , we cannot guarantee the consistency in terms of variable selection. The above fused Lasso favors larger models which may include some irrelevant regressors in addition to the relevant ones. Fortunately, as the Diebold-Yilmaz connectedness measure is based on forecasting variance decomposition, such limitation in variable selection would not interfere with our result much.

The adaptive version of the fused Lasso would be one promising way to overcome the variable selection problem and thus attain the oracle properties (Zou, 2006; Wang and Leng, 2008). Right now we are working on the asymptotic properties for the adaptive fused Lasso.

### **3.4** Estimation Procedure

To wrap up, our suggested algorithm can be executed as follows.

#### **Algorithm** For each equation $i = 1, \dots, N$ ,

1. The first-stage estimator (fused Lasso):

Let  $\Lambda^i$  denote the set of candidate  $(\lambda_1^i, \lambda_2^i)$ 's. For each  $(\lambda_1^i, \lambda_2^i) \in \Lambda^i$ ,

- (a) Solve the optimization problem in equation (2) to obtain the estimated break dates,  $\{\tilde{\tau}_b^i\}_{b=1}^{\tilde{B}^i}$ , and the selected (non-zero) regressors in  $\{\tilde{\alpha}_b^i\}_{b=1}^{\tilde{B}^i+1}$ .
- (b) Perform the post-Lasso regression based on the estimated break dates and the selected regressors to obtain the coefficient estimates,  $\{\tilde{\alpha}_b^i\}_{b=1}^{\tilde{B}^i+1}$ , and the average RSS,  $(\tilde{\sigma}^i)^2 = \frac{RSS}{T}$ .
- (c) Calculate the information criterion in equation (4).

Then, select the pair of  $(\hat{\lambda}_1^i, \hat{\lambda}_2^i)$  that minimizes the above IC.

- 2. Use the corresponding first-stage post-Lasso estimates,  $\left\{\tilde{\boldsymbol{\beta}}_{t}\right\}_{t=1}^{T}$ , to construct the adaptive weights for the second stage:  $w_{1t,j}^{i} = 1/\left|\tilde{\beta}_{t,j}^{i}\right|$  and  $w_{2t} = 1/\left\|\tilde{\boldsymbol{\beta}}_{t} \tilde{\boldsymbol{\beta}}_{t-1}\right\|_{2}$ .

## 4 Simulation

### 4.1 DGPs

In this experiment, we consider the following 5 environments with the VAR order p = 1, the crosssectional dimension N = 100, and the time-series dimension T = 1500 (days). For each DGP, we simulate  $n_{sim} = 100$  times with the same parameters but different shocks.

- DGP 1: No break at all.
- DGP 2: Sharp breaks with a single crisis:

The crisis regime is t = 201 - 350 with higher connectedness.

• DGP 3: Sharp breaks with multiple crises:

There are three crisis regimes, t = 201 - 350, 551 - 700, and 901 - 1050, with higher connectedness.

• DGP 4: Smooth variation:

 $\Phi_t$  follows a random walk.

• DGP 5: Smooth variation with multiple crises:

Here we combine DGPs 3 and 4. There are three crisis regimes, t = 201 - 350, 551 - 700, and 901 - 1050, with higher connectedness. Beyond the jumps at the break dates,  $\Phi_t$  follows a random walk with the step sizes being much smaller than the sizes of the leaps.

We use DGP 1 to test the severity of false discovery. DGPs 2 and 3 capture a somewhat ideal case with clean sharp breaks.<sup>22</sup> We impose a higher connectedness during the crisis regimes according to the economic theory in Zawadowski (2013), as well as the empirical findings in Diebold and Yilmaz (2014b) and our application in Section 5.2.1. As the proposed fused Lasso technique is designed for abrupt changes, we try the worst scenario in DGP 4 with only smooth variation to see how bad it can get under complete misspecification. DGP 5 is the most realistic case and thus our benchmark, where the connectedness changes mildly over time except a few occasional jumps between the crisis and the non-crisis regimes. Note that the random walk steps are much larger in DGP 4 than in DGP 5, because in DGP 4, we intend to generate sizable fluctuations solely from the random walk variation.

For all DGPs, the equation-by-equation signal-to-noise ratios are quite dispersed and most of them are less than 1 (see Table 4 in the Appendix). The reason is that under the sparsity assumption,

 $<sup>^{22}</sup>$ Note that our approach is an unsupervised learning method treating each sub-interval in a completely separate manner, so what only matters for estimation is the minimum sub-interval length rather than the percentage of time in crisis. 150 days would be a sensible choice of crisis length concerning the empirical observations in Hatzius *et al.* (2010), Guo *et al.* (2011), Diebold and Yilmaz (2014b), and our application in Section 5.2.1.

a large number of equations have only a few non-zero coefficients and even no non-zero coefficients. These small signal-to-noise ratios would add further difficulties to our estimation.

#### 4.2 Computational Details

As both the RSS and the penalty terms are convex, we can employ some common convex solvers, e.g. CVX (Grant and Boyd, 2008, 2014),<sup>23</sup> to tackle our optimization problem. We construct a 10-by-10 grid on the tuning parameters  $(\lambda_1, \lambda_2)$ ,<sup>24</sup> and use  $\rho_T = T^{-1/2}$  to pick the optimal pair. This choice of  $\rho_T$  is suggested by Assumption 4 in the theoretical derivation. There is no consent in the literature which information criterion works the best - Wang *et al.* (2007) and Zou *et al.* (2007) recommend *BIC*, while Flynn *et al.* (2013) find *AIC<sub>c</sub>* superior. To check for robustness, we have also tried the model selection via *AIC<sub>c</sub>* and *BIC*. They produce only a negligible difference from the current estimates.<sup>25</sup> The descriptive statistics for the distribution of the selected  $(\lambda_1, \lambda_2)$ 's are given in Table 5 in the Appendix. We can see that  $\lambda_1$ 's are roughly two orders of magnitude smaller than  $\lambda_2$ 's as suggested in our theoretical result, and the chosen  $(\lambda_1, \lambda_2)$ 's are comparable across different DGPs.

To proceed, we compare our method with the rolling window approach (Diebold and Yilmaz, 2012; Demirer *et al.*, 2014; Diebold and Yilmaz, 2014a,b).<sup>26</sup> Resembling the choice of tuning parameters in the fused Lasso, we need to select the window width for the rolling window approach to control the smoothness of the estimated time paths of the VAR coefficients. As there is no systematic way to pick the window width, here we take a shortcut – for each DGP, we choose from the three candidate widths, 100, 150, and 200 by minimizing the root-mean-square error (RMSE) of the

$$\min_{\{\beta_t\}} \left\{ \frac{1}{T} \sum_{t=1}^{T} \left( y_t - \beta'_t x_t \right)^2 + \lambda_1 \sum_{t=1}^{T} \|\beta_t\|_2 \right\}$$

contain only zeros; and similarly, take the reference tuning parameter  $\lambda_2^m$  such that for any  $\lambda_2 \ge \lambda_2^m$ , the resulting  $\ddot{\beta}_t$  from

$$\min_{\{\beta_t\}} \left\{ \frac{1}{T} \sum_{t=1}^{T} \left( y_t - \beta_t' x_t \right)^2 + \lambda_2 \sum_{t=2}^{T} \|\beta_t - \beta_{t-1}\|_2 \right\}$$

are constant over time. The minimum  $(\lambda_1^m, \lambda_2^m)$  satisfying the above criteria can be obtained through closely examining the Karush-Kuhn-Tucker optimality condition (Ohlsson *et al.*, 2010).

Another way to pin down the benchmark tuning parameters is based on the rate-optimal penalty level derived in Belloni *et al.* (2011) with a conservative plug-in for the standard deviation of the shocks.

Then, for penalty terms j = 1, 2, we place the 10 ticks between  $10^{-4} \lambda_j^m$  and  $\lambda_j^m$  spaced equally on a log scale.

 $^{25}$ We didn't attempt to implement the cross-validation approach because in our setting, it is hard to split the data into training and validation sets as we have network structure on cross-sectional dimension together with structural changes over time.

<sup>26</sup>In the simulation and the subsequent application, we adopt a symmetric two-sided window rather than the onesided window as in their studies, since the former would eliminate the phase shift and thus make our competitor, the rolling window approach, more competitive.

 $<sup>^{23}</sup>$ CVX may be computationally suboptimal, since it is a general-purpose convex solver without taking full advantage of the special features of the fused Lasso. We would explore more about it in our future research.

<sup>&</sup>lt;sup>24</sup>In the spirit of Qian and Su (2013), for each equation  $i = 1, \dots, N$  (the index *i* has been dropped below), we first take the reference tuning parameter  $\lambda_1^m$  such that for any  $\lambda_1 \ge \lambda_1^m$ , the resulting  $\hat{\beta}_t$  from

estimated time-varying graph  $\hat{G}_t$  (averaged over elements, time, and simulations). The asterisks in Table 1 indicate the selected rolling window widths. Note that this criterion requires the knowledge of the true  $G_t$ , and thus is infeasible in real data analysis. In each rolling sample, we implement the equation-by-equation adaptive Lasso estimation to handle the large cross-sectional dimension N. The tuning parameter  $\lambda$  is picked via *BIC*.

#### 4.3 Results

In Figures 1 to 5, we show a typical view for each DGP based on the information from one simulation. The four rows of graphs demonstrate the time paths of the four centrality statistics respectively. The left panels are derived from the true DGP, while the central panels from the fused Lasso estimates, and the right panels from the rolling window estimates under the optimal window width. For all DGPs, the fused Lasso estimates successfully track the time-varying centrality statistics of the true networks. From DGP 1, false discovery seems not a concern; from DGPs 2, 3, and 5, it clearly identifies the break dates, but exhibits modest underestimation of the peaks; even for DGP 4 with only smooth evolution, it can still retain the general shape of the time paths. Besides, by examining relative positions of the means and the medians in the graphs, we can see that the fused Lasso provides good approximation to the distributions of the centrality measures. In contrast, although the simple rolling window approach is able to detect the peaks and troughs in like manner, the estimated paths are much noisier (except for DGP 4) and miss the exact break dates.

Table 1 presents RMSEs for the estimated coefficients  $\hat{\Phi}_t$ , the implied adjacency matrix  $\hat{G}_t$ , and the four centrality statistics, under different DGPs and different estimation methods (averaged over elements, time, and simulations). We can clearly see that the fused Lasso yields noticeably smaller RMSEs than the rolling window method. Within the four statistics, the in-degree and the vulnerability enjoy relatively smaller errors since the Diebold-Yilmaz connectedness measure normalized the in-degree, i.e. the row sum of the adjacency matrix, to be less than one.

Table 2 assesses the performance of the fused Lasso in break detection. The first column,  $P(Correct \ Detect)$ , calculates the correct detection rate regarding whether a break exists or not; and the second column, hd/T, provides the average Hausdorff distance<sup>27</sup> between the estimated and the true sets of break dates normalized by the total number of periods T, conditional on correct detection of the breaks. Both measures are averaged over equations and simulations. Similar to what we found in the figures, the fused Lasso does a good job in finding the break dates. It produces a considerably low false-discovery rate for DGP 1. The estimated break dates are quite close to the true jumps for DGPs 2, 3 and 5, where the results for DGP 5 are slightly worse due to the small

$$hd(A,B) = \max\left\{\sup_{a\in A}\inf_{b\in B}|a-b|, \sup_{b\in B}\inf_{a\in A}|a-b|\right\}.$$

 $<sup>^{27}</sup>$ The Hausdorff distance is a distance measure between two subsets A and B of a metric space. In the current situation where both A and B are sets of integers, the Hausdorff distance is defined as

random-walk evolution. There is no surprise that DGP 4 suffers from a high false-discovery rate – the random-walk step sizes for DGP 4 are larger than those for DGP 5, which may misguide the fused Lasso to conceive the long steps as structural changes.



Figure 1: DGP 1: No break - A Typical View



Figure 2: DGP 2: Sharp Breaks with a Single Crisis - A Typical View

\* The blue vertical lines indicate the true break dates.



Figure 3: DGP 3: Sharp Breaks with Multiple Crises - A Typical View

\* The blue vertical lines indicate the true break dates.



Figure 4: DGP 4: Smooth Variation - A Typical View

21



Figure 5: DGP 5: Smooth Variation with Multiple Crises - A Typical View

\* The blue vertical lines indicate the true break dates.

DCD	Mothod		Â.	Ĉ	In	Out	Crea	Val
DGL	method		$\Psi_t$	Gt	10.	Out.	Sys.	vul.
	Fused Las	SO	0.02	0.01	0.01	0.04	0.07	0.03
1		$\mathrm{width}=100$	0.11	0.04	0.33	0.37	0.69	0.65
1	Rolling	$\mathrm{width} = 150$	0.07	0.02	0.27	0.31	0.58	0.52
		width = $200^*$	0.06	0.02	0.19	0.25	0.48	0.39
	Fused Las	SO	0.05	0.02	0.05	0.13	0.35	0.18
0		width $= 100$	0.12	0.04	0.28	0.37	0.81	0.60
Ζ	Rolling	width = $150^*$	0.10	0.04	0.20	0.29	0.64	0.44
		$\mathrm{width}=200$	0.10	0.04	0.35	0.45	0.94	0.73
	Fused Las	0.07	0.02	0.06	0.16	0.45	0.25	
n		width $= 100$	0.13	0.05	0.29	0.42	0.95	0.67
0	Rolling	width = $150^*$	0.13	0.06	0.20	0.32	0.74	0.48
		$\mathrm{width}=200$	0.14	0.06	0.35	0.49	1.06	0.79
	Fused Lasso		0.08	0.03	0.09	0.19	0.26	0.14
4		width $= 100$	0.13	0.04	0.26	0.32	0.50	0.43
4	Rolling	$\mathrm{width}=150$	0.10	0.03	0.21	0.27	0.41	0.35
		width = $200^*$	0.09	0.03	0.16	0.24	0.37	0.28
۳	Fused Las	0.07	0.03	0.06	0.16	0.45	0.26	
		width $= 100$	0.13	0.05	0.29	0.40	0.89	0.65
0	Rolling	width = $150^*$	0.13	0.05	0.20	0.30	0.68	0.45
	width $= 200$		0.14	0.06	0.36	0.47	1.01	0.78

Table 1: Simulation RMSE

\* In.: In-degree; Out.: Out-degree; Sys.: Systemicness; Vul.: Vulnerability.

\* The asterisks indicate the selected rolling window widths.

DGP	P(Correct Detect)	$\rm hd/T$
1	99.8%	-
2	99.3%	0.24%
3	99.5%	0.39%
4	8.5%	-
5	92.0%	1.68%

\* **P(Correct Detect)**: the correct detection rate regarding whether a break exists or not (averaged over equations and simulations).

\* hd/T: the average Hausdorff distance between the estimated and the true sets of break dates normalized by the total number of periods T (averaged over equations and simulations).

## 5 Application

### 5.1 Data

In this application, we are particularly interested in the volatility connectedness among financial institutions, for the reason that volatility is a widely used measure of risk and that a high volatility provides a good indicator of financial crises. As volatility is not directly observable, we try to infer it via the realized volatility measure (Andersen *et al.*, 2010). Sheppard *et al.* (2013) analyze a wide range of realized measures of asset price variation and find that it is hard to beat the standard 5-minute realized volatility. We obtain the intraday high-frequency data from the Trade and Quote (TAQ) database,<sup>28</sup> and compute the bank-specific daily realized stock return volatility as the sum of squared intraday returns at a given sampling frequency, e.g. once every 5 minutes. For each trading day, we only consider the transactions executed between 9:30 am and 4:00 pm, as there would be much fewer and more irregular transactions beyond this period. This leaves us with 78 intraday returns each day to construct the daily realized volatility. In addition, as the volatility tends to be right-skewed, we further perform a log transformation and a standardization on the bank-specific realized volatility.<sup>29</sup> Now, for each financial institution, the resulting data approximate a standard normal distribution and hence comply with the requirement for GVD (Koop *et al.*, 1996; Pesaran and Shin, 1998).

Originally, we had in our mind 110 major financial institutions traded in the U.S. stock market. For each stock, we first filter out the "short days" with transactions recorded for less than 60% of the trading hours (Sheppard *et al.*, 2013); we then exclude all transactions which are labeled as unfavorable "Correction Indicator" or irregular "Sale Condition".<sup>30</sup> After all these trimming, we only keep the financial institutions with valid realized volatility for more than 95% of the trading days, and then impute the remaining missing data via linear interpolation.<sup>31</sup> In the end, we are left with 61 financial institutions, composed of 51 commercial banks, 4 investment banks, 2 credit card companies, and 4 insurance companies.<sup>32</sup> The panel spans from January 2, 2004 to December 31, 2013 with 2517 days in total. We omit the periods before 2004 as there were fewer banks available then, and include the most recent data as we also want to examine the aftermath of the late-2000s financial crisis.

Table 6 in the appendix presents the list of financial institutions in our sample, along with their business type, market capitalization, and some summary statistics for the standardized log daily

 $<sup>^{28}\</sup>mathrm{TAQ}$  is a thorough data source for all intraday transactions of all stocks listed on the NYSE, NYSE-AMEX, NASDAQ-NMS and NASDAQ-SmallCap.

<sup>&</sup>lt;sup>29</sup>Note that shrinkage regressions are in general not scale invariant, so typically, the regressors are standardized beforehand.

<sup>&</sup>lt;sup>30</sup>Please refer to the TAQ User's Guide for details.

 $<sup>^{31}</sup>$ The remaining missing data would be less than 5% of the total observations. Besides, the imputation procedure can be further justified by the persistence of high-frequency volatility.

<sup>&</sup>lt;sup>32</sup>Unfortunately, there is no mortgage company surviving all the criteria. Especially, the full time series for Fannie Mae and Freddie Mac are not available from the current database.

realized volatility. The skewness is positive but small, and the kurtosis is around three, which confirms our previous argument that the distribution for the resulting data would be close to a standard normal.

### 5.2 Discussions

#### 5.2.1 Structural Breaks

The following analysis is accomplished under a VAR(1) approximation, and the residues do not exhibit much serial correlation.

Figure 6 shows the time paths of the four centrality statistics. First, those from the fused Lasso follows a neat step pattern with the major break dates in general coinciding with crucial financial events, which backs the validity of our method. The outburst of the recent crisis is marked on September 11, 2008, immediately after the US government seized Fannie Mae and Freddie Mac, and just before Merrill Lynch was purchased by Bank of America, Lehman Brothers filed for bankruptcy court protection, and AIG accepted the federal bailout. The crisis significantly attenuated after June 5, 2009, which coincides with the NBER ending quarter of the Great recession, and corresponds to the recovery of the financial sector as ten big banks participating in the Capital Purchase Program (CPP) have met the requirements for repayment. Second, resonated with the counter-party risk theory in Zawadowski (2013) and previous empirical findings in Diebold and Yilmaz (2014b), the network gets more strongly connected with the distributions of degrees and centrality statistics shifting up altogether during the crisis. Accordingly, there are more banks turning influential and vulnerable during the crisis, which would aggravate the spread of volatility risk.

### 5.2.2 Heteroskedasticity

With the estimated time-varying financial institution VAR at hand, we can examine the relative importance of different components of the variance decomposition measure. As pointed out in Section 2.1, the Diebold-Yilmaz connectedness measure takes into account variable interactions, common factors, shock sizes, and shock correlation. The first two are captured by the VAR coefficients  $\Phi$ ,<sup>33</sup> the third one is represented by diagonal elements of the shock covariance matrix { $\Sigma_{ii}$ }, and the last one is given by the shock correlation matrix { $\rho_{ij} = \Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}}$ }. Above Section 5.2.1 and Figure 6 are constrained to the homoskedastic case featuring time-varying  $\Phi$  but fixed  $\Sigma$ . In this subsection, we attempt to extend our discussion to the heteroskedastic case and assess the effects of shock sizes and shock correlation on the connectedness measure.

We first draw the estimation residues for each financial institution (with the absolute value taken) in Figures 18 to 20 - there tends to be more large shocks during the crisis periods, which suggests that the third component, shock sizes, may also play a role. Considering that the coefficient estimates are

<sup>&</sup>lt;sup>33</sup>Unfortunately, it is hard to separately identify them without a structural model.

consistent, we can also employ the fused Lasso on the outer-product of the VAR residues to capture potential time-varying heteroskedasticity of the shocks in this preliminary exploration.<sup>34</sup> We can compare the following four heteroskedastic setups to the homoskedastic version (Figure 6):

- varying both  $\Phi$  and  $\Sigma$  (Figure 7),
- fixing  $\Phi$  while varying  $\Sigma$  (Figure 8),
- fixing  $\Phi$  and  $\{\rho_{ij}\}$  while varying  $\{\Sigma_{ii}\}$  (Figure 9),
- and fixing  $\Phi$  and  $\{\Sigma_{ii}\}$  while varying  $\{\rho_{ij}\}$  (Figure 10).

It can be seen that: First, the shock sizes, i.e.  $\{\Sigma_{ii}\}$ , exert almost no effect on the network structural changes. Second, the shock correlation, i.e.  $\{\rho_{ij}\}$ , induces an inverse U-shaped evolution path peaking at the recent crisis for all centrality statistics. Third, the recent financial crisis is mainly driven by variable interactions (and/or common factors) and shock correlation, i.e.  $\Phi$  and  $\{\rho_{ij}\}$ . Last, the hump from July 2011 to January 2012 in Figure 7 appears to be purely due to the variation of shock correlation, i.e.  $\{\rho_{ij}\}$ , which may arise from the intensive Fed policies during that period, e.g. Operation Twist and bank stress tests. Furthermore, Table 3 tabulates the relative differences of the estimated graphs with respect to the simplest case with fixed  $\Phi$  and  $\Sigma$ , and provides quantitative evidence for above observations that structural changes in the interaction pattern play a more decisive role in the recent financial crisis, while unfavorable individual shocks exhibit a merely negligible effect.

#### 5.2.3 Connectedness Graphs

Moreover, we also present the typical graphs for the financial institution connectedness before, during, and after the crisis in Figures 11 to 13, based on the fused Lasso estimates with both  $\Phi$ and  $\Sigma$  varying over time.<sup>35</sup> Once again, the links are noticeably stronger during the crisis with AIG being the most influential "trouble-maker".<sup>36</sup> This makes sense as AIG was a major supplier of the credit default swaps (CDSs) and a big investor in sub-prime lending.

#### 5.2.4 Policy Implications

Therefore, the estimated model can be used to guide real-time crisis monitoring and facilitate policymaking. First, an increase in the average degree or centrality of the financial institution connectedness would be a good indicator for the start of a financial crises. In addition, the regulators may wish to pay closer attention to the financial institutions with higher centrality, like AIG, as they are

<sup>&</sup>lt;sup>34</sup>An additional constraint is imposed to ensure positive-definiteness of the convariance matrices.

 $<sup>^{35}</sup>$ The connectedness graphs are drawn with the Cytoscape package (Saito *et al.*, 2012).

<sup>&</sup>lt;sup>36</sup>Note that acquired or bankrupted financial institutions, such as Lehman Brother, are not in our sample.

more pivotal for the spread of the crisis,<sup>37</sup> and may even impose taxes on the financial institutions according to their centrality (Acharya *et al.*, 2009) to provide a more compelling incentive for the financial institutions to maintain a healthier financial structure.

### 5.2.5 Limitations

Despite all these exciting findings, it is worth mentioning a couple of limitations of our approach in the empirical context. First, the fused Lasso VAR proposed in this paper calls for a balanced panel, so we are not able to incorporate the financial institutions acquired or bankrupted during the crisis, like Bear Stearns and Lehman Brother. Second, as stated in the assumptions in Appendix B, there are some resolution restrictions on the interval length. If two jumps are too close to each other, it would be hard to distinguish them and provide reliable estimates.

 $<sup>^{37}</sup>$  On July 8, 2013, The Financial Stability Oversight Council voted to designate AIG as systemically important, which "will be subject to consolidated supervision by the Federal Reserve and enhanced prudential standards". (http://www.treasury.gov/initiatives/fsoc/designations/Pages/default.aspx)



Figure 6: Application: Fix  $\Sigma$ , Vary  $\Phi$ 



\* The blue vertical lines indicate the estimated major break dates: 9/11/2008 and 6/5/2009, from left to right.



Figure 7: Application: Vary Both  $\Phi$  and  $\Sigma$ 



\* The blue vertical lines indicate the estimated major break dates: 9/11/2008 and 6/5/2009, from left to right.



Figure 8: Application: Fix  $\Phi$ , Vary  $\Sigma$ 



\* The blue vertical lines indicate the estimated major break dates: 9/11/2008 and 6/5/2009, from left to right.



Figure 9: Application: Fix  $\Phi$  and  $\{\rho_{ij}\}$ , Vary  $\{\Sigma_{ii}\}$ 



\* The blue vertical lines indicate the estimated major break dates: 9/11/2008 and 6/5/2009, from left to right.



Figure 10: Application: Fix  $\Phi$  and  $\{\Sigma_{ii}\}$ , Vary  $\{\rho_{ij}\}$ ,



\* The blue vertical lines indicate the estimated major break dates: 9/11/2008 and 6/5/2009, from left to right.

	Fu	sed La	SSO	Rolling		
	$\hat{G}_t$	Sys.	Vul.	$\hat{G}_t$	Sys.	Vul.
Vary Both $\Phi$ and $\Sigma$	0.36	0.29	0.15	0.39	0.31	0.13
Fix $\Sigma$ , Vary $\Phi$	0.12	0.15	0.14	0.31	0.27	0.08
Fix $\Phi$ , Vary $\Sigma$	0.21	0.20	0.07	0.24	0.23	0.08
Fix $\Phi$ and $\{\rho_{ij}\}$ , Vary $\{\Sigma_{ii}\}$	0.03	0.03	0.01	0.03	0.03	0.00
Fix $\Phi$ and $\{\Sigma_{ii}\}$ , Vary $\{\rho_{ij}\}$	0.21	0.21	0.08	0.24	0.23	0.08

Table 3: Difference w.r.t. "Fixed  $\Phi$  and  $\Sigma$ "

\* Sys.: Systemicness; Vul.: Vulnerability. \* The relative difference is defined as  $\frac{\operatorname{avg}(|A-A_0|)}{\operatorname{avg}(|A_0|)}$  for  $A = \hat{G}_t$ , Sys., and Vul. (averaged over elements and time), and the reference point  $A_0$  is from the simplest setup where both  $\Phi$  and  $\Sigma$  are fixed over time.



Figure 11: The Connectedness Graph: 9/1/2006 (before the Crisis)

\* Node size shows the market capitalization. Node color indicates the bank business type - orange for a commercial bank; green for an investment bank; blue for a credit card company; and yellow for an insurance company.

 $\ast$  Link width and grayscale illustrate the strength of the link.

\* For clearer display, we only plot the links whose strengths are above the 99th percentile.





\* Node size shows the market capitalization. Node color indicates the bank business type - orange for a commercial bank; green for an investment bank; blue for a credit card company; and yellow for an insurance company.

 $\ast$  Link width and grayscale illustrate the strength of the link.

\* For clearer display, we only plot the links whose strengths are above the 99th percentile.



Figure 13: The Connectedness Graph: 10/1/2013 (after the Crisis)

\* Node size shows the market capitalization. Node color indicates the bank business type - orange for a commercial bank; green for an investment bank; blue for a credit card company; and yellow for an insurance company.

 $\ast$  Link width and grayscale illustrate the strength of the link.

\* For clearer display, we only plot the links whose strengths are above the 99th percentile.

## 6 Concluding Remarks

This paper proposes a fused Lasso method which performs well in detecting structural changes in network connectedness. An application to the major financial institutions traded in the U.S. stock market gives an insightful interpretation regarding the nature of the recent financial crisis. We are planning to extend the current analysis in the following several dimensions:

First, the model selection steps introduce some discontinuities in the estimator which would generate huge difficulties for regular inference methods, while ignoring the model selection steps would lead to over-optimistic inference. To that end, we plan to assess the estimation accuracy and construct the confidence intervals via either the local quadratic approximation (LQA) as in Fan and Li (2001) or the Bootstrap aggregating (Bagging) as in Efron (2012).

Second, we would also try a Bayesian version of the fused Lasso where the frequentist estimates can be viewed as the posterior mode. Park and Casella (2008) propose a fully Bayesian version of the original Lasso, while some recent works (Kang and Guo, 2009; Kyung *et al.*, 2010; Liu *et al.*, 2010a) discuss various extensions concerning adaptive Lasso, fused Lasso, and VAR. To determine the structural changes in networks, our endeavor would be to develop a Bayesian fused Lasso algorithm in the VAR setting built on above pieces while taking care of the high-dimensionality problem. We can construct a Gibbs sampler by exploring the hierarchical structure of the model. Accordingly, the tuning parameters  $(\lambda_1, \lambda_2)$  can be selected via the empirical Bayesian method or the hyper-prior method. Compared to the frequentist method, a desirable feature of the Bayesian estimation is that it can easily produce credible intervals for the estimates and even the likelihood-ratio statistic for the tuning parameters  $(\lambda_1, \lambda_2)$ , which would facilitate tests and inferences.

Third, the current program takes about 10 hours to solve the case where N is around  $10^2$  and T is over  $10^3$  on an Intel(R) Core(TM) i7-3770 CPU @ 3.40 GHz with 12.0 GB RAM. We can imagine that it would be much more burdensome if we push N to the order of magnitude over  $10^3$ , though it is a favorable situation for some network studies. To deal with this issue, we are considering exploring faster algorithms exploiting the specific features of the fused Lasso (Liu *et al.*, 2010b; Yu *et al.*, 2013), switching the programming language from MATLAB to C, and utilizing high-performance computing clusters.

Fourth, it would be desirable but might be difficult to prove the oracle properties of the adaptive Lasso and extend the asymptotic argument to the case where  $N \to \infty$ . We conjecture that satisfactory results can be achieved if N goes to infinity at a rate slower than T. It is still working in progress.

Fifth, there are many potential applications for the proposed method, e.g. spatial economics analysis (Arbia, 2006; Anselin, 2010; de Souza, 2012), input-output structure and propagation of productivity shocks (Acemoglu *et al.*, 2012; Atalay, 2014), etc. It is also worthwhile to explore these areas through the lenses of structural changes in networks.

Last, we also plan to check several variations of the Diebold-Yilmaz connectedness framework.

For example:

(a) We currently perform the shrinkage/selection on the VAR coefficients. Hence, the resulting variance decomposition matrix is not sparse, which provides a good approximation to the highly integrated financial system. An alternative setup, which may be useful in another context, would be imposing some sparsity constraints on the network adjacency matrix, i.e. the variance decomposition matrix. The main challenge here is that the optimization problem may not be convex anymore, which aggravates the computational burden and cannot guarantee the global minimum.

(b) We focus on the volatility connectedness as high volatilities are often associated with panics and crises, but the traditional financial theory is main built on the correlation among returns instead of volatilities, such as the beta in the capital asset pricing model (CAPM). In this sense, it is worthwhile to estimate the return connectedness as well and compare it with the volatility connectedness.<sup>38</sup>

 $<sup>^{38}</sup>$ The linkage in return connectedness may be feeble due to the low persistence of the returns.

## References

- ACEMOGLU, D., CARVALHO, V. M., OZDAGLAR, A. and TAHBAZ-SALEHI, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, **80** (5), 1977–2016.
- ACHARYA, V. V., PEDERSEN, L. H., PHILIPPON, T. and RICHARDSON, M. (2009). Restoring Financial Stability: How to Repair a Failed System. Wiley Finance series, Hoboken, N.J.: Wiley, book section Regulating Systemic Risk, pp. 283–303.
- AHMED, A. and XING, E. P. (2009). Recovering time-varying networks of dependencies in social and biological studies. In *Proceedings of the National Academy of Sciences*, National Acad Sciences, vol. 106, pp. 11878–11883.
- ALDASORO, I. and ANGELONI, I. (2013). Input-output-based measures of systemic importance.
- ANDERSEN, T. G., BOLLERSLEV, T. and DIEBOLD, F. X. (2010). Handbook of Financial Econometrics, Amsterdam: North-Holland, book section Parametric and Nonparametric Volatility Measurement, pp. 67–138.
- ANSELIN, L. (2010). Thirty years of spatial econometrics. Papers in Regional Science, 89 (1), 3–25.
- ARBIA, G. (2006). Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence. Berlin and New York: Springer, book.
- ATALAY, E. (2014). How important are sectoral shocks?
- BAI, J. and PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, pp. 47–78.
- and (2003a). Computation and analysis of multiple structural change models. *Journal of* Applied Econometrics, **18** (1), 1–22.
- and (2003b). Critical values for multiple structural change tests. *Econometrics Journal*, **6** (1), 72–78.
- and (2006). Econometric Theory and Practice: Frontiers of Analysis and Applied Research, Cambridge and New York: Cambridge University Press, book section Multiple Structural Change Models: A Simulation Analysis, pp. 212–37.
- BARIGOZZI, M. and BROWNLEES, C. T. (2014). Nets: Network estimation for time series.
- BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2011). Inference for high-dimensional sparse econometric models. arXiv preprint arXiv:1201.0220.

- BERTSEKAS, D. P., NEDI, A. and OZDAGLAR, A. E. (2003). *Convex analysis and optimization*. Athena Scientific.
- BILLIO, M., GETMANSKY, M., LO, A. W. and PELIZZON, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104 3, 535–59.
- BONALDI, P., HORTACSU, A. and KASTL, J. (2013). An empirical analysis of systemic risk in the euro-zone.
- BREIDT, F. J. and CARRIQUIRY, A. L. (1996). Modelling and Prediction Honoring Seymour Geisser, Springer, book section Improved quasi-maximum likelihood estimation for stochastic volatility models, pp. 228–247.
- BRUNNERMEIER, M. K. (2009). Deciphering the liquidity and credit crunch 2007-2008. Journal of Economic Perspectives, 23 (1), 77–100.
- BUHLMANN, P. and VAN DE GEER, S. (2011). Statistics for high-dimensional data: methods, theory and applications. Springer.
- D'AGOSTINO, A., GAMBETTI, L. and GIANNONE, D. (2011). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*.
- DE SOUZA, P. C. L. (2012). Estimating networks: Lasso for spatial weights.
- DEMIRER, M., DIEBOLD, F. X. and YILMAZ, K. (2014). Volatility connectedness and the global bank network.
- DIEBOLD, F. X. and YILMAZ, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, **28** (1), 57–66.
- and (2014a). Financial and Macroeconomic Connectedness: A Network Approach to Measurement and Monitoring.
- and (2014b). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, p. 36 pages.
- EFRON, B. (2012). Model selection, estimation, and bootstrap smoothing.
- ELLIOTT, M., GOLUB, B. and JACKSON, M. O. (2013). Financial networks and contagion.
- EROL, S. and VOHRA, R. (2014). Network formation and systemic risk.
- FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96** (456), 1348–1360.

- FLYNN, C. J., HURVICH, C. M. and SIMONOFF, J. S. (2013). Efficiency for regularization parameter selection in penalized likelihood estimation of misspecified models. *Journal of the American Statistical Association*, **108** (503), 1031–43.
- FOERSTER, A. T., SARTE, P.-D. G. and WATSON, M. W. (2011). Sectoral versus aggregate shocks: A structural factor analysis of industrial production. *Journal of Political Economy*, **119** (1).
- GIUDICI, P. and SPELTA, A. (2013). Graphical network models for international financial flows.
- GRANT, M. and BOYD, S. (2008). Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd and H. Kimura (eds.), *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, Springer-Verlag Limited, pp. 95–110.
- and (2014). CVX: Matlab software for disciplined convex programming, version 2.1.
- GUO, F., CHEN, C. R. and HUANG, Y. S. (2011). Markets contagion during financial crisis: A regime-switching approach. *International Review of Economics and Finance*, **20** (1), 95–109.
- HARCHAOUI, Z. and LEVY-LEDUC, C. (2010). Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association*, **105** (492), 1480–93.
- HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The elements of statistical learning*, vol. 2. Springer.
- HATZIUS, J., HOOPER, P., MISHKIN, F. S., SCHOENHOLTZ, K. L. and WATSON, M. W. (2010). Financial conditions indexes: A fresh look after the financial crisis. *National Bureau of Economic Research, Inc, NBER Working Papers.*
- HSU, N.-J., HUNG, H.-L. and CHANG, Y.-M. (2008). Subset selection for vector autoregressive processes using lasso. *Computational Statistics & Data Analysis*, **52** (7), 3645 3657.
- JACKSON, M. O. (2010). *Social and Economic Networks*. Princeton and Oxford: Princeton University Press, book.
- KANG, J. and GUO, J. (2009). Self-adaptive lasso and its bayesian estimation.
- KNIGHT, K. and FU, W. (2000). Asymptotics for lasso-type estimators. *Annals of statistics*, pp. 1356–1378.
- KOOP, G. and KOROBILIS, D. (2013). Large time-varying parameter vars. *Journal of Econometrics*, **177** (2), 185–198.
- —, PESARAN, M. H. and POTTER, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74 (1), 119–147.

- KYUNG, M., GILL, J., GHOSH, M. and CASELLA, G. (2010). Penalized regression, standard errors, and bayesian lassos. *Bayesian Analysis*, **5** (2), 369–411.
- LIU, F., CHAKRABORTY, S., LI, F. and LIU, Y. (2010a). Bayesian regularization via the graph laplacian prior.
- LIU, J., YUAN, L. and YE, J. (2010b). An efficient algorithm for a class of fused lasso problems. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp. 323–332.
- LOWENSTEIN, R. (2010). The End of Wall Street. Penguin.
- MEDEIROS, M. C. and MENDES, E. F. (2012). L1-estimation of high-dimensional time-series models with flexible innovations.
- MINOIU, C. and REYES, J. A. (2013). A network analysis of global banking: 1978–2010. *Journal of Financial Stability*, **9** (2), 168–184.
- OHLSSON, H., LJUNG, L. and BOYD, S. (2010). Segmentation of arx-models using sum-of-norms regularization. *Automatica*, **46** (6), 1107–1111.
- PARK, T. and CASELLA, G. (2008). The bayesian lasso. Journal of the American Statistical Association, 103 (482), 681–86.
- PERRON, P. (2006). Dealing with structural breaks. Palgrave handbook of econometrics, 1, 278–352.
- PESARAN, H. H. and SHIN, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics letters*, 58 (1), 17–29.
- QIAN, J. and SU, L. (2013). Shrinkage estimation of regression models with multiple structural changes.
- SAITO, R., SMOOT, M. E., ONO, K., RUSCHEINSKI, J., WANG, P. L., LOTIA, S., PICO, A. R., BADER, G. D. and IDEKER, T. (2012). A travel guide to cytoscape plugins. *Nat Methods*, 9 (11), 1069–76.
- SHEPPARD, K., LIU, L. and PATTON, A. J. (2013). Does anything beat 5-minute rv? a comparison of realized measures across multiple asset classes.
- SONG, Y. (2010). Modelling regime switching and structural breaks with an infinite dimension markov switching model. p. 49 pages.
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), pp. 267–288.

- (2011). Regression shrinkage and selection via the lasso: a retrospective. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73 (3), 273–282.
- —, SAUNDERS, M., ROSSET, S., ZHU, J. and KNIGHT, K. (2005). Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67 (1), 91–108.
- WANG, H. and LENG, C. (2008). A note on adaptive group lasso. *Computational Statistics & Data Analysis*, **52** (12), 5277–5286.
- ---, LI, R. and TSAI, C.-L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*, **94** (3), 553–568.
- YU, D., WON, J.-H., LEE, T., LIM, J. and YOON, S. (2013). High-dimensional fused lasso regression using majorization-minimization and parallel processing. *Journal of Computational and Graphical Statistics*.
- YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68 (1), 49–67.
- ZAWADOWSKI, A. (2013). Entangled financial systems. *Review of Financial Studies*, **26** (5), 1291–1323.
- ZHANG, B., GENG, J. and LAI, L. (2013). Detecting changes in regression models via sparse group lasso.
- ZOU, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, **101** (476), 1418–29.
- ---, HASTIE, T. and TIBSHIRANI, R. (2007). On the "degrees of freedom" of the lasso. *The Annals of Statistics*, **35** (5), 2173–2192.

## A Discussion

#### A.1 Centrality Measures

In this illustration, we consider a simple weighted directed network plotted in Figure 14, with the following adjacency matrix

$$G = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0.5 & 1 & 0 & 0 \end{bmatrix}$$

Figures 15 and 16 demonstrate how systemicness and vulnerability vary with the discounting factor  $\eta$ .

Recall that the systemicness of node *i* measures how an exogenous shock to node *i* affects the whole system. With  $\eta = 0$ , we only care about node *i*'s direct effect on its immediate neighbors, namely, the out-degree. With  $0 < \eta < 1$ , we also take into account its indirect effect on other nodes while discounting remote linkages. A larger  $\eta$  implies less discount, and thus larger systemicness. For example, although nodes 1 and 2 have same level of out-degree (i.e. direct effect), as long as  $\eta > 0$ , node 1 turns out to be more systemic than node 2. This can be explained by their relative positions in the network – in terms of the total effect, node 1 can potentially cast an impact on the whole system, while node 2 only has a potential impact on nodes 3 and 4. From this point of view, systemicness provides a more reasonable measure than the traditional out-degree in assessing the importance of a node. Moreover, node 3 has zero systemicness, as it affects no one else; and the systemicness statistics of both node 3 and node 4 do not change over  $\eta$ , as these two nodes have no further impact beyond their direct effect.

Similar argument holds for vulnerability. In-degree is a special case of vulnerability with  $\eta = 0$ ; and a larger  $\eta$  would result in a larger vulnerability measure. Note that for a small  $\eta$ , node 4 is more vulnerable than node 3; situation reverses for a large  $\eta$ . This is once again due to their relative position in the network – in terms of the direct effect, node 4 is directly exposed to both nodes 1 and 2, while node 3 directly exposed to node 4 only; in terms of the total effect, node 3 can potentially be affected by everyone else in the system, while node 4 only by nodes 1 and 2. Figure 14: A Simple Weighted Directed Network



Figure 15: Systemicness





Figure 16: Vulnerability

## **B** Assumptions and Proofs

#### Assumption 1. (DGP)

(i)  $(x_t, \epsilon_{it})$  follows a strong mixing process with the mixing coefficient  $\alpha(\tau) \leq c_{\alpha}\rho^{\tau}$  for some  $c_{\alpha} > 0$  and  $\rho \in (0, 1)$ .  $E(x_t \epsilon_{it}) = 0$  for each t.

(ii) Either one of the following two conditions holds:

- (a)  $\sup_{t\geq 1} E \|x_t\|_2^{4q} < \infty$  and  $\sup_{t\geq 1} E \|\epsilon_{it}\|_2^{4q} < \infty$  for some q > 2;

-  $(b) \sup_{t\geq 1} E\left[\exp\left(C_{xx} \|x_t\|_2^{2\gamma}\right)\right] \leq C_{xx} < \infty \text{ and } \sup_{t\geq 1} E\left[\exp\left(C_{x\epsilon} \|x_t\epsilon_{it}\|_2^{\gamma}\right)\right] \leq C_{x\epsilon} < \infty \text{ for some constants } C_{xx} \text{ and } C_{x\epsilon}, \text{ and some } \gamma \in (0, \infty].$ 

The assumption on DGP is quite general, which includes ARMA, ARCH, etc. and allows for both lagged dependent variables in  $x_t$  and heteroskedasticity in  $\epsilon_{it}$ . Under Assumption 1 (ii.b), it is possible to estimate the break dates at an almost optimal convergence rate  $(T^{-1})$  up to a log factor.

### Assumption 2. $(\delta_T)$

(i) The positive sequence  $\{\delta_T\}$  satisfies  $\sqrt{T}\delta_T \to 0$  as  $T \to \infty$ . (ii) - (a) If Assumption 1 (ii.a) holds, we need  $T\delta_T \ge c_v T^{1/q}$  for some  $c_v > 0$ .

- (b) If Assumption 1 (ii.b) holds, we need  $T\delta_T \ge c_v (\log T)^{(2+\gamma)/\gamma}$  for some  $c_v > 0$ .

(iii) There exist two positive constants  $\underline{c}_{xx}$  and  $\overline{c}_{xx}$  such that

$$\underline{c}_{xx} \leq \inf_{1 \leq s < r \leq T+1, r-s \geq T\delta_T} \mu_{\min} \left( \frac{1}{r-s} \sum_{t=s}^{r-1} E\left(x_t x_t'\right) \right)$$
$$\leq \sup_{1 \leq s < r \leq T+1, r-s \geq T\delta_T} \mu_{\max} \left( \frac{1}{r-s} \sum_{t=s}^{r-1} E\left(x_t x_t'\right) \right) \leq \overline{c}_{xx}$$

 $\delta_T$  governs the speed at which the estimated break dates converge to the true dates. Assumption 2 (iii) is a weaker version of stationarity. As we have lagged dependent variables in  $x_t$  together with structural changes, the usual stationarity assumption cannot be satisfied.

Assumption 3. (Interval Length and Jump Size)

$$(i) I_{\min} = O(T).$$

(ii)  $J_{\min} \to J^* \ge 0$  and  $J_{\min}^2 / \max\left\{ (\log T)^{c_{\delta}} / (T\delta_T), T^{-1/2} \right\} \to \infty$  as  $T \to \infty$  where  $c_{\delta} = 6$  if Assumption 1 (ii.a) holds, and  $c_{\delta} = 1$  if Assumption 1 (ii.b) holds.

(iii)  $\lambda_2/(J_{\min}\delta_T) \to 0 \text{ as } T \to \infty.$ 

Intuitively, we want  $I_{\min}$  and  $J_{\min}$  to be large enough for easier identification of the jumps and better estimation of the parameters; and we also want  $\lambda_2$  to be small enough, so it would not introduce too much bias.

Assumption 4. (Information Criterion)  $(1 + J_{\min}^{-2}) \rho_T \to 0 \text{ and } \delta_T^{-1} \rho_T \to \infty \text{ as } T \to \infty.$  The existence of such  $\rho_T$  is guaranteed by Assumptions 2 (i) and 3 (ii).

#### Assumption 5. (Breaks)

The true number of breaks is bounded by a finite number  $B_{\text{max}}$ .

This assumption follows the classical structural break literature, such as Bai and Perron (1998, 2003a,b, 2006).

#### Assumption 6. $(\lambda_1)$

 $T\lambda_1/\lambda_2 \to 0 \text{ as } T \to \infty.$ 

Intuitively, when  $\lambda_1$  is much smaller than  $\lambda_2$ , the penalty terms on the coefficients are much smaller than the penalty terms on the successive differences in the expression for the Karush-Kuhn-Tucker optimality condition, which in turn ensures the asymptotic argument in Qian and Su (2013).

#### Assumption 7. (Uniformity)

- (i)  $J_b = J_0 \Delta_b$ , where
- $\Delta_b$  is independent of T,

- and the positive scalar  $J_0$  satisfies  $J_0 \to 0$  and  $T^{\frac{1}{2}-\vartheta}J_0 \to \infty$  as  $T \to \infty$ , for some  $\vartheta \in (0, 1/2)$ . (ii) For  $b = 1, \dots, B+1$ , as  $T \to \infty$ ,  $I_b^{-1} \sum_{t=\tau_{b-1}+1}^{\tau_{b-1}+\lfloor sI_b \rfloor} E(x_t x'_t) \to s\Psi_b$  and  $I_b^{-1} \sum_{t=\tau_{b-1}+1}^{\tau_{b-1}+\lfloor sI_b \rfloor} \sigma_{it}^2 E(x_t x'_t) \to s\Phi_b$  uniformly in s, where  $\Psi_b$  and  $\Phi_b$  are the b-th diagonal matrices of  $\Psi$  and  $\Phi$ , respectively.

*Proof.* (Propositions 1 and 2 - Fused Lasso)

(i) Reparametrize the objective function (equation 2): Let

$$z_t = \left[\underbrace{x'_t, \cdots, x'_t}_{t \text{ replica}}, 0, \cdots, 0\right]'$$
$$d_{(NTp\times 1)} = \left[d'_1, d'_2, \cdots, d'_T\right]',$$

where  $d_1 = \beta_1$ ,  $d_2 = \beta_2 - \beta_1$ ,  $\cdots$ ,  $d_T = \beta_T - \beta_{T-1}$ . Then, we can express

$$\beta_t = \sum_{s=1}^t d_s ,$$
  
$$y_{it} = \beta'_t z_t + \epsilon_{it} = d' z_t + \epsilon_{it} .$$

Note that the coefficient vector d does not change over time as we stack up all the variations together in one long vector. Accordingly, our objective function becomes

$$\tilde{d} = \arg\min_{\{d\}} \left\{ \frac{1}{T} \sum_{t=1}^{T} \left( y_{it} - d'z_t \right)^2 + \lambda_1 \sum_{t=1}^{T} \left\| \sum_{s=1}^{t} d_s \right\|_1 + \lambda_2 \sum_{t=2}^{T} \|d_t\|_2 \right\} .$$

#### (ii) "First-order condition":

As the regularization terms are not differentiable at zero, we consider subdifferentials as in convex analysis (Bertsekas *et al.*, 2003). For  $t = 1, \dots, T$ , a necessary and sufficient condition for a global minimizer is

$$-\frac{2}{T}\sum_{r=t}^{T}\left(y_r - \left(\sum_{s=1}^{r}\tilde{d}_s\right)'x_r\right)x_r + \lambda_1\sum_{r=t}^{T}v_r + \lambda_2u_t = 0_{Np\times 1},$$

where

$$\begin{aligned} v_{r,j} &= \operatorname{sgn}\left(\sum_{s=1}^{r} \tilde{d}_{s,j}\right) = \operatorname{sgn}\left(\tilde{\beta}_{r,j}\right), \text{ if } \tilde{\beta}_{r,j} \neq 0; \text{ and } ||v_{r,j}| \leq 1, \text{ if } \tilde{\beta}_{r,j} = 0; \\ u_t &= \frac{\tilde{d}_t}{\left\|\tilde{d}_t\right\|_2}, \text{ if } \left\|\tilde{d}_t\right\|_2 \neq 0; \text{ and } ||u_t||_2 \leq 1, \text{ if } \left\|\tilde{d}_t\right\|_2 = 0. \end{aligned}$$

Then, we can bound the FOC of RSS as

$$\frac{1}{T} \left\| \sum_{r=t}^{T} \left( y_r - \left( \sum_{s=1}^{r} \tilde{d}_s \right)' x_r \right) x_r \right\|_2 = \frac{1}{T} \left\| \sum_{r=t}^{T} \left( y_r - \tilde{\beta}'_r x_r \right) x_r \right\|_2$$
$$= \frac{1}{2} \left\| \lambda_1 \sum_{r=t}^{T} v_r + \lambda_2 u_t \right\|_2$$
$$\leq \frac{1}{2} \left( \lambda_1 T \sqrt{Np} + \lambda_2 \right) .$$

Based on Assumption 6 (ii), we can neglect the first term on the right hand side as N is fixed in the current setting, which leads to

$$\frac{1}{T} \left\| \sum_{r=t}^{T} \left( y_r - \tilde{\beta}'_r x_r \right) x_r \right\|_2 \le \frac{\lambda_2}{2} \left( 1 + o\left( 1 \right) \right) \,.$$

Thus, the proofs of Theorems 3.1 and 3.2 in Qian and Su (2013) still hold.

### (iii) Post-Lasso average RSS

Theorems 3.3 and 3.4 in Qian and Su (2013) determine the correct number of breaks based on post-Lasso estimation. Let  $\sigma_T^2 = \frac{1}{T} \sum_{t=1}^T \epsilon_{it}^2$ ,  $\sigma^2(\vec{\tau}_B, \vec{v}_B)$  denote the average RSS for the OLS estimates given the set of break dates  $\vec{\tau}_B$  and variable selection  $\vec{v}_B$ , and  $\sigma^2(\vec{\tau}_B)$  denote the average RSS for the OLS estimates given the set of break dates  $\vec{\tau}_B$  without variable selection. In an analogous manner, we would like to show:

(i) 
$$\frac{T}{I_{\min}J_{\min}^{2}} \left( \sigma^{2} \left( \vec{\tau}_{\tilde{B}}, \vec{v}_{\tilde{B}} \right) - \sigma_{T}^{2} \right) \geq c + o_{p} (1), \text{ for any } \left( \vec{\tau}_{\tilde{B}}, \vec{v}_{\tilde{B}} \right) \text{ with } \tilde{B} < B;$$
  
(ii) 
$$\delta_{T}^{-1} \left| \sigma^{2} \left( \vec{\tau}_{\tilde{B}}, \vec{v}_{\tilde{B}} \right) - \sigma_{T}^{2} \right| = O_{p} (1), \text{ for any } \left( \vec{\tau}_{\tilde{B}}, \vec{v}_{\tilde{B}} \right) \text{ with } \tilde{B} \geq B.$$

Note that  $\sigma^2\left(\vec{\tau}_{\tilde{B}}, \vec{v}_{\tilde{B}}\right) \geq \sigma^2\left(\vec{\tau}_{\tilde{B}}\right)$ , which entails claim (i) and one direction of claim (ii):  $\delta_T^{-1}\left(\sigma^2\left(\vec{\tau}_{\tilde{B}}, \vec{v}_{\tilde{B}}\right) - \sigma_T^2\right) > -O_p\left(1\right)$ , for any  $\left(\vec{\tau}_{\tilde{B}}, \vec{v}_{\tilde{B}}\right)$  with  $\tilde{B} \geq B$ . Moreover, Assumptions 2 (i), 3 (ii), 3 (iii), and 6 jointly

imply that  $T^{3/2}\lambda_1 \to 0$  and  $T^{1/2}\lambda_2$  as  $T \to \infty$ . Then, from the KKT condition with respect to  $\beta_{t,j}$ ,

$$-\frac{2}{T}\left(y_t - \tilde{\beta}'_t x_t\right) x_{t,j} + \lambda_1 v_{t,j} + \lambda_2 \left(u_{t,j} - u_{t+1,j}\right) = 0 ,$$

where

$$v_{t,j} = \operatorname{sgn}\left(\tilde{\beta}_{t,j}\right), \text{ if } \tilde{\beta}_{t,j} \neq 0; \text{ and } |v_{t,j}| \leq 1, \text{ if } \tilde{\beta}_{t,j} = 0;$$
$$u_{t,j} = \frac{\tilde{d}_{t,j}}{\left\|\tilde{d}_t\right\|_2}, \text{ if } \left\|\tilde{d}_t\right\|_2 \neq 0; \text{ and } \|u_t\|_2 \leq 1, \text{ if } \left\|\tilde{d}_t\right\|_2 = 0.$$

Hence, all relevant regressors are included asymptotically, which guarantees the other direction of claim (ii):  $\delta_T^{-1} \left( \sigma^2 \left( \vec{\tau}_{\tilde{B}}, \vec{v}_{\tilde{B}} \right) - \sigma_T^2 \right) < O_p (1)$ , for any  $\left( \vec{\tau}_{\tilde{B}}, \vec{v}_{\tilde{B}} \right)$  with  $\tilde{B} \ge B$ .

Therefore, we can obtain proposition 1 based on Theorems 3.1 - 3.4 and 3.6 in Qian and Su (2013), and proposition 2 based on Theorem 3.5 in Qian and Su (2013).

# C Tables and Graphs

## C.1 Simulation

Table	4: Descriptive	Statistics	IOT	Signal-to-Ivoise	Ratio

**a**.

ът ·

ъ

1 /

. . . .

DGP	Mean	Med	25%	75%	Std	Skew	Kurt
1	6.03	0.69	0.00	2.09	14.41	2.60	8.26
2	5.67	0.74	0.03	2.05	13.22	2.60	8.30
3	4.97	0.92	0.09	2.46	10.90	2.58	8.22
4	2.03	0.02	0.00	1.11	5.85	3.96	19.67
5	5.12	0.90	0.10	2.36	11.61	2.91	10.86

\* The signal-to-noise ratio for each variable i is approximated by

$$SNR_i \approx \frac{Var\left(\Phi_{i\cdot}^{(1,t)}Y_{t-1}\right)}{Var\left(\epsilon_{it}\right)}$$

where  $\Phi_{i}^{(1,t)}$  denotes the *i*-th row of  $\Phi^{(1,t)}$ , the true time-varying coefficient matrix. \* Number of simulations  $n_{sim} = 100$ . Cross-sectional dimension N = 100. For each DGP, we have  $10^4$  SNRs in total.

	$\lambda_1$								
DGP	Mean (×10 <sup>-4</sup> )	Med (×10 <sup>-4</sup> )	$25\% (\times 10^{-4})$	$75\% (\times 10^{-4})$	Std ( $\times 10^{-4}$ )	Skewness	Kurtosis		
1	3.52	2.36	0.79	3.60	4.24	2.48	5.68		
2	2.36	2.12	0.69	3.05	2.42	3.27	15.92		
3	2.66	2.25	0.81	3.30	2.61	2.98	12.36		
4	1.54	1.63	0.69	1.85	1.13	3.98	27.14		
5	2.36	2.26	0.74	3.12	2.25	4.28	29.49		
			$\lambda_2$						
DGP	Mean (×10 <sup>-2</sup> )	Med $(\times 10^{-2})$	$25\% (\times 10^{-2})$	$75\% (\times 10^{-2})$	Std $(\times 10^{-2})$	Skewness	Kurtosis		
1	2.08	2.18	1.75	2.40	0.58	-0.37	0.61		
2	2.07	2.10	1.45	2.40	0.89	1.83	8.33		
3	2.06	2.04	1.65	2.32	0.87	3.01	17.00		
4	2.29	1.29	0.75	1.83	4.42	5.94	41.76		
5	2.59	2.13	1.50	2.53	3.40	7.68	68.45		

Table 5: Descriptive Statistics for Selected Tuning Parameters  $(\lambda_1, \lambda_2)$ 

\* Number of simulations  $n_{sim} = 100$ . Cross-sectional dimension N = 100. For each DGP, there are  $10^4$  pairs of  $(\lambda_1, \lambda_2)$  in total, as we allow for different tuning parameters for different equations.



Figure 17: DGP 5: With/Without Penalty Terms on the Coefficients

\* The blue vertical lines indicate the true break dates.

# C.2 Application

Ticker	Company Name	Type	Mkt-Cap	Med	25%	75%	Skew	Kurt
AF	Astoria Financial Corporation	CB	1.24	-0.17	-0.75	0.65	0.63	3.10
ASBC	Associated Banc-Corp	CB	2.70	-0.15	-0.72	0.61	0.61	3.18
BAC	Bank of America Corporation	CB	152.58	-0.11	-0.77	0.56	0.78	3.50
BBT	BB&T Corporation	CB	26.46	-0.20	-0.72	0.56	0.89	3.50
BK	The Bank of New York Mellon Corp.	CB	38.81	-0.23	-0.67	0.46	1.28	5.24
BOH	Bank of Hawaii Corporation	CB	2.42	-0.22	-0.71	0.58	0.88	3.45
BPOP	Popular, Inc.	CB	3.00	-0.08	-0.81	0.78	0.25	2.30
BXS	BancorpSouth, Inc.	CB	2.16	-0.15	-0.69	0.53	0.79	3.75
С	Citigroup Inc.	CB	141.07	-0.17	-0.77	0.59	0.90	3.84
CBSH	Commerce Bancshares, Inc.	CB	4.08	-0.12	-0.66	0.56	0.57	3.32
CBU	Community Bank System Inc.	CB	1.44	-0.15	-0.72	0.59	0.70	3.59
$\operatorname{CFR}$	Cullen/Frost Bankers, Inc.	CB	4.51	-0.22	-0.69	0.52	0.91	3.55
CMA	Comerica Incorporated	CB	8.36	-0.21	-0.74	0.61	0.82	3.25
CYN	City National Corporation	CB	3.76	-0.15	-0.75	0.61	0.65	3.05
EWBC	East West Bancorp, Inc.	CB	4.68	-0.23	-0.71	0.58	0.73	3.15
FBC	Flagstar Bancorp Inc.	CB	0.94	-0.03	-0.89	0.77	0.30	2.26
FCF	First Commonwealth Financial Corp.	CB	0.77	-0.16	-0.73	0.65	0.56	3.03
FITB	Fifth Third Bancorp	CB	17.04	-0.22	-0.75	0.57	0.97	3.83
FMER	FirstMerit Corporation	CB	3.12	-0.19	-0.70	0.53	0.89	4.07
FNB	F.N.B. Corporation	CB	1.99	-0.18	-0.71	0.57	0.70	3.50
FNFG	First Niagara Financial Group Inc.	CB	2.95	-0.16	-0.68	0.54	0.87	4.42
FULT	Fulton Financial Corporation	CB	2.18	-0.15	-0.72	0.65	0.66	3.48
HBAN	Huntington Bancshares Incorporated	CB	7.45	-0.16	-0.75	0.61	0.77	3.39
HCBK	Hudson City Bancorp, Inc.	CB	4.77	-0.15	-0.72	0.65	0.56	3.20
ING	ING Groep N.V.	CB	50.30	-0.13	-0.74	0.55	0.83	3.71
JPM	JPMorgan Chase & Co.	CB	201.74	-0.20	-0.73	0.52	0.99	3.89
KEY	KeyCorp.	CB	11.58	-0.16	-0.76	0.60	0.77	3.36
NTRS	Northern Trust Corporation	CB	14.14	-0.23	-0.68	0.45	1.23	4.84
ONB	Old National Bancorp.	CB	1.34	-0.15	-0.69	0.58	0.67	3.53
PBCT	People's United Financial Inc.	CB	4.22	-0.17	-0.69	0.54	0.66	3.63

Table 6: Descriptive Statistics for Standardized Log Daily RV

**Type**: Bank business types - CB stands for a commercial bank; IB for an investment bank; CC for a credit card company; and INS for an insurance company.

**Mkt-Cap**: Market capitalization, as of 12/31/2013, unit: billion\$.

Ticker	Company Name	Type	Mkt-Cap	Med	25%	75%	Skew	Kurt
BPFH	Boston Private Financial Holdings, Inc.	CB	0.96	-0.19	-0.74	0.66	0.60	2.94
PFS	Provident Financial Services, Inc.	CB	0.95	-0.17	-0.73	0.62	0.74	3.44
PNC	The PNC Financial Services Group, Inc.	CB	44.22	-0.24	-0.74	0.58	0.93	3.54
$\operatorname{RF}$	<b>Regions Financial Corporation</b>	CB	13.62	-0.14	-0.78	0.63	0.63	3.02
RJF	Raymond James Financial, Inc.	CB	6.79	-0.24	-0.69	0.50	1.01	3.91
RY	Royal Bank of Canada	CB	96.62	-0.15	-0.72	0.56	0.81	3.83
SNV	Synovus Financial Corporation	CB	3.04	0.07	-0.83	0.76	0.03	2.25
STI	SunTrust Banks, Inc.	CB	19.76	-0.15	-0.78	0.62	0.73	3.15
STSA	Sterling Financial Corporation	CB	2.07	-0.31	-0.76	0.75	0.66	2.58
STT	State Street Corporation	CB	27.20	-0.25	-0.69	0.47	1.31	5.12
SUSQ	Susquehanna Bancshares, Inc.	CB	1.82	-0.18	-0.73	0.67	0.57	2.95
TCB	TCF Financial Corporation	CB	2.50	-0.15	-0.72	0.57	0.72	3.39
TD	The Toronto-Dominion Bank	CB	87.31	-0.15	-0.73	0.55	0.78	3.62
TRMK	Trustmark Corporation	CB	1.55	-0.19	-0.70	0.59	0.85	3.85
UMPQ	Umpqua Holdings Corporation	CB	3.49	-0.21	-0.71	0.57	0.79	3.57
USB	U.S. Bancorp	CB	74.22	-0.17	-0.71	0.51	0.95	3.97
VLY	Valley National Bancorp	CB	1.91	-0.14	-0.69	0.53	0.67	3.47
WBS	Webster Financial Corp.	CB	2.62	-0.18	-0.77	0.62	0.70	3.09
WFC	Wells Fargo & Company	CB	258.51	-0.20	-0.77	0.59	0.91	3.49
WTFC	Wintrust Financial Corporation	CB	2.01	-0.21	-0.75	0.65	0.68	3.06
ZION	Zions Bancorporation	CB	5.18	-0.17	-0.77	0.64	0.66	2.92
BCS	Barclays PLC	IB	66.45	-0.20	-0.70	0.50	1.06	4.35
$\operatorname{GS}$	The Goldman Sachs Group, Inc.	IB	72.71	-0.23	-0.67	0.42	1.28	5.22
NMR	Nomura Holdings, Inc.	IB	22.08	-0.15	-0.66	0.46	1.30	6.60
UBS	UBS AG	IB	74.69	-0.17	-0.76	0.59	0.81	3.33
AXP	American Express Company	CC	92.75	-0.23	-0.74	0.55	0.89	3.45
COF	Capital One Financial Corporation	$\mathbf{C}\mathbf{C}$	43.26	-0.23	-0.74	0.54	0.95	3.52
AIG	American International Group, Inc.	INS	75.95	-0.17	-0.72	0.56	0.90	3.95
BRKB	Berkshire Hathaway Inc.	INS	0.21	-0.16	-0.67	0.51	1.06	6.03
MET	MetLife, Inc.	INS	55.65	-0.19	-0.65	0.44	1.21	4.91
PRU	Prudential Financial, Inc.	INS	36.78	-0.19	-0.67	0.44	1.19	4.79

**Type**: Bank business types - CB stands for a commercial bank; IB for an investment bank; CC for a credit card company; and INS for an insurance company.

Mkt-Cap: Market capitalization, as of 12/31/2013, unit: billion\$.

	Mean (×10 <sup>-2</sup> )	Med (× $10^{-2}$ )	$25\% (\times 10^{-2})$	$75\% (\times 10^{-2})$	Std $(\times 10^{-2})$	Skewness	Kurtosis
$\lambda_1$	1.45	0.44	0.37	3.21	1.79	1.33	0.18
$\lambda_2$	42.28	41.00	35.85	48.74	11.61	0.57	0.63

Table 7: Application: Descriptive Statistics for Selected Tuning Parameters  $(\lambda_1, \lambda_2)$ 

\* There are N = 61 pairs of  $(\lambda_1, \lambda_2)$  in total, as we allow for different tuning parameters for different equations.



Figure 18: Time Paths of the Residues (Absolute Value) - 1



Figure 19: Time Paths of the Residues (Absolute Value) - 2



Figure 20: Time Paths of the Residues (Absolute Value) - 3

