CORVINUS UNIVERSITY of BUDAPEST
INSTITUTE OF FINANCE

Doctoral School of Economics, Business and Informatics

Thesis Summary

on

Network Analysis of the Financial Sector: A Comprehensive Perspective with Adaptive Joint LASSO Method

Milán Csaba Badics

Ph.D dissertation

Supervisor:
Zsuzsa Réka Huszár Ph.D.
Associate Professor

Budapest, 2023

Corvinus University of Budapest

Doctoral School of Economics, Business and Informatics

Thesis Summary

on

Network Analysis of the Financial Sector: A Comprehensive Perspective with Adaptive Joint LASSO Method

Milán Csaba Badics

Ph.D dissertation

Supervisor:
Zsuzsa Réka Huszár Ph.D.
Associate Professor

Contents

List of Figures

List of Tables

1 Chapter 1. Motivation

It is somewhat surprising in hindsight that network analysis of financial systems has become widely recognized as a critical regulatory issue only after the 2008 Global Financial Crisis (GFC). Pairwise and group connections among financial institutions (FIs) can both stabilize or destabilize the system. While on the one hand, FIs' interconnectivity aids flexibility to investment and financing in the economy, on the other hand, these connections may contribute to risk propagation during crisis periods. Furthermore, high connectedness (e.g., strong liquidity dependency) between FIs promotes the sudden transformation of the network architecture (Elliott et al., 2014; Acemoglu et al., 2015).

Pairwise and groupwise spillovers in the financial network play a crucial role in systemic risk assessment. Furthermore, the strength of connections is non-constant, and significantly varies over time, sharply increasing during crisis periods, as documented in a GFC study (Diebold and Yılmaz, 2014). In the network the FIs' connections are especially critical with the liquidity provider or strong network participants because many smaller institutions rely on the large institutions for financing, which is especially critical during turbulent times. For this reason, the regulators need to monitor and analyze the structural changes in financial networks and identify the systematically important financial institutions (SIFIs).

Motivated by the GFC, the financial network literature exploded, with studies attempting to better understand or model the connections between FIs in the US and globally. Acemoglu et al. (2015) and Elliott et al. (2014) derived financial system connectedness from the crossholdings of shares, debt, or obligations, and investments. These linkages can enhance diversification, but such linkages can also lead to cascading defaults and failures in a system. Consequently, identifying and monitoring financial linkages is essential for understanding system vulnerability as a result of large-scale contagions during turmoil.

Besides that, Acemoglu et al. (2015) showed that the connectedness of the financial network enhances stability in the system until the magnitude and the number of shocks hitting the network remain low. However, if the level of financial distress exceeds a certain threshold, the structure of the financial network dictates the extent of contagion. In other words, higher financial connectedness makes the network more sensitive and more prone to contagion of shocks. Thus, not only the analysis of SIFIs but the inclusion of the whole system is necessary. Institutions on the periphery may move to the center of the financial network owing to a significant shock that restructures the linkages.

The thesis focuses on this important topic from various angles to answer the following questions:

- How can machine learning techniques help to improve financial network studies, the analysis of high dimensional time series?
- How can structural changes in financial networks efficiently examined using event analysis framework?

2 Chapter 2. Methods

2.1 Diebold-Yilmaz framework

Networks based on time series are often represented in graphs, where the nodes (which represent time series) and edges (linkages between the series) are graphically displayed. One type of network is the weighted network, which allows for weights on the edges to represent stronger or weaker linkages between the nodes (time series). Direct networks allow for asymmetric linkages. In the econometrics and finance literature Diebold and Yilmaz (2009) were the first to show in a seminal paper that the forecast error variance decomposition (FEVD) of an estimated Vector autoregressive (VAR) model of the network can be interpreted as a weighted directed graph. Diebold and Yilmaz (2009), Diebold and Yilmaz (2012), and Diebold and Yılmaz (2014) suggest a unified framework (DY framework) for measuring linkages or spillovers between the time series. The framework is extremely popular in recent years, the three seminal papers citation is more than 2200 just in 2023.

The most important merits of the DY framework are the following (Demirer et al., 2018):

1. The method does not require additional restrictions beyond those imposed for VAR(p) model estimation and identification.

2. It provides both the direction and magnitude of the measures. The estimated network will be directed.

3. The VAR(p) estimation step, impulse response functions (IRF) and FEVDs are commonly used in economics and econometrics. The estimated networks are easily interpretable.

4. The framework allows one to track spillovers between the time series at all levels of the network, from pairwise connections (micro level) to system-wide connections (macro level).

5. It allows for static and dynamic usage, which is relevant in financial applications where events can cause abrupt changes in the network.

6. It's a generalization of the Granger-causality based networks.

The framework is based on the concept that, for every time series of the network, we can calculate the forecast error variance based on the estimated VAR(p) model coefficient and covariance matrix. This variance is related to its own and other time series shocks. Due to the VAR(p) model identification, the shares of own and other time series' shocks can be calculated. In the last step of the process, the forecast error variance decompositions can be summarized in a spillover table, which we refer to hereafter as the DY spillover table.

In the next few paragraph I present a framework in details. The first step of the framework is to specify and estimate a stationary VAR(p) model for the J time series of the network based on the following equation:

$$Y_t = \sum_{i=1}^{p} \beta_i Y_{t-i} + \epsilon_t \tag{1}$$

where Y_t is a $J \times 1$ vector of the time series, β_i is an $J \times J$ autoregressive coefficient matrix, and lastly ϵ_t is an $J \times 1$ vector of error terms. It has a zero mean with a Σ covariance matrix. No intercept is included in Eq. (1), without loss of generality, I assume, that all the J time series are mean centered. The VAR(p) process is assumed to be stable and stationary, while the covariance matrix Σ is needed to be positive definite (Lütkepohl, 2013).

To estimate the DY framework's most important element, the DY spillover table, I need to estimate the coefficient matrices β_1, β_2, ... β_p and the error covariance matrix Σ efficiently. The β_i coefficient matrices reveal the temporal dependence between the time series and Σ reveals the contemporaneous linkages among them (Han et al., 2015; Davis et al., 2016).

The starting point for the DY framework (Diebold and Yilmaz, 2012) is to transform the time series of the VAR(p) in Eq. (1) into its vector moving average (VMA) representation using the Wold representation theorem (Diebold and Yilmaz, 2012; Gabauer et al., 2020) to get Eq. (2):

$$Y_t = \sum_{i=0}^{\infty} A_i \epsilon_{t-i} \tag{2}$$

where A_i is an $J \times J$ moving average coefficient matrix. Based on the Wold's theorem A_i is given by the following recursion $A_i = \sum_{j=1}^{p} \beta_j A_{i-j}$ where $A_j = 0$ for $j < 0$ and A_0 is an identity matrix.

As Diebold and Yilmaz (2012) emphasized, the calculated moving average coefficients and the estimated error covariance matrix (or its nonlinear transformations such as impulse response functions or forecast error variance) are the keys to understand the dynamics of the time series network.

Forecast error variance (FEV) allows me to calculate the fraction of the H step-ahead error variance in forecasting Y_i ($Y_i(H)$) that is due to shocks to other time series such as Y_j (Diebold and Yilmaz, 2012), to which I will hereafter refer as a spillover between Y_i and Y_j. Generally, in the DY framework, the FEVD of the VAR(p) model gives the measures of spillovers between the time series. Unfortunately, the calculation of the FEV requires orthogonal innovations. However, the VAR innovations are generally contemporaneously correlated (Diebold and Yilmaz, 2012; Diebold and Yılmaz, 2014; Basu and Michailidis, 2015).

There are two widely used approaches in the early DY framework-related papers for obtaining the variance decomposition. The first method uses the Cholesky factor orthogonalization of the covariance matrix Σ which generates orthogonalised innovations. The weakness of this decomposition is that its results order dependent FEVDs (Diebold and Yilmaz, 2012; Fengler and Gisler, 2015). The other approach uses the generalized VAR framework, which was introduced in the seminal papers by Koop et al. (1996) and Pesaran and Shin (1998). This framework allows correlated shocks. As a result, this second method produces an order-independent FEVD. Applying the second method is more widespread in empirical DY network studies.

The following equation shows the calculation of the generalized FEVD:

$$\theta_{ij}^{g}(H) = \frac{\sigma_{jj}^{-1} \sum_{h=0}^{H-1} (e_i' A_h \Sigma e_j)^2}{\sum_{h=0}^{H-1} (e_i' A_h \Sigma A_h' e_j)} \tag{3}$$

Where σ_{jj} is the j-th diagonal element of the error term's covariance matrix Σ, A_h is

the moving average coefficient matrix multiplying the h-lagged shock vector in the Wold's moving average representation (Eq. (2)) and e_i is a selection vector.

The numerator in Eq. (4) of $\tilde{\theta}^g_{ij}(H)$ represents the contribution of shocks in variable Y_j to the H-step FEVD of time series Y_i. The denominator is the forecast error variance of the time series Y_j. Unfortunately, the sum of the network time series' contribution to the forecast error variance is not necessarily one because, generally, FEVD of the shock terms are not orthogonalized (Diebold and Yilmaz, 2012). Normalization is therefore required, which I calculate in the following way:

$$\tilde{\theta}^g_{ij}(H) = \frac{\theta^g_{ij}(H)}{\Sigma^J_{k=1}\theta^g_{ik}(H)}. \tag{4}$$

The generalized FEVD is used to construct the several systemic/network connectedness measures of the DY framework (Diebold and Yilmaz, 2012; Diebold and Yılmaz, 2014). Firstly, the sum of directional spillovers to time series Y_i from all other time series (FROM spillover index $S^g_{i\leftarrow\bullet}$) is defined with the following equation:

$$S^g_{i\leftarrow\bullet}(H) = \frac{\sum^J_{k=1,k\neq i}\tilde{\theta}^g_{ik}(H)}{\sum^J_{k=1}\tilde{\theta}^g_{ik}(H)} * 100\%. = \sum_{k=1,k\neq i}^{J}\tilde{\theta}^g_{ik}(H) * 100\%. \tag{5}$$

Second, I am interested in the sum of the shocks transmitted by time series Y_i tO other time series (TO spillover index $S^g_{\bullet\leftarrow i}$):

$$S^g_{\bullet\leftarrow i}(H) = \sum_{k=1,k\neq i}^{J}\tilde{\theta}^g_{ki}(H) * 100\%. \tag{6}$$

The third relevant measure is the NET spillover index (S^g_i, Eq. (7)), which calculates the difference between the gross transmitted (TO) and received (FROM) shocks from all other time series:

$$S^g_i(H) = S^g_{\bullet\leftarrow i}(H) - S^g_{i\leftarrow\bullet}(H) \tag{7}$$

Finally, at the macro level of the network analysis, the system-wide spillover index (SUM spillover index S^g_{total}) offers information about the average influence one time series has on all other time series, regardless of the direction on the following way:

$$S^g_{total}(H) = \frac{1}{J}\sum_{i,k=1,i\neq k}^{J}\tilde{\theta}^g_{ik}(H) \tag{8}$$

In other words, the total spillover index is the sum of all the off-diagonal elements of the generalized FEVD matrix relative to the number of time series considered in the VAR(p) model (Diebold and Yilmaz, 2012; Gabauer et al., 2020).

2.2 Example for Diebold-Yilmaz framework

To illustrate the Diebold-Yilmaz framework, I calculate the spillover table and the FROM, TO, NET, and SUM spillover indices from a low-dimension system (for five time series). I consider that that data-generating process (DGP) is a VAR(1) model, and the coefficient matrix (B) is sparse, with one dominant time series:

$$B = \begin{bmatrix} 0.40 & 0 & \mathbf{0.08} & 0 & 0 \\ 0 & 0.40 & \mathbf{0.07} & 0 & 0 \\ 0 & 0 & 0.40 & 0 & 0 \\ 0 & 0 & \mathbf{0.08} & 0.40 & 0 \\ 0 & 0 & \mathbf{0.09} & 0 & 0.40 \end{bmatrix} \tag{9}$$

The diagonal elements in the coefficient matrix refer to the autoregressive dependence within the time series. Based on the third column of the matrix, the third time series leads to the other ones.

To illustrate the role of the covariance matrix (Σ) in the Diebold-Yilmaz framework, I consider two different structures. In the first case, both the covariance matrix is the identity matrix:

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{10}$$

In the second case, I consider a is a Toeplitz-type structure with a $\rho = 0.3$ parameter:

$$\Sigma = \begin{bmatrix} 1.00 & 0.20 & 0.04 & 0.01 & 0.00 \\ \mathbf{0.20} & 1.00 & 0.20 & 0.04 & 0.01 \\ \mathbf{0.04} & 0.20 & 1.00 & 0.20 & 0.04 \\ \mathbf{0.01} & 0.04 & 0.20 & 1.00 & 0.20 \\ \mathbf{0.00} & 0.01 & 0.04 & 0.20 & 1.00 \end{bmatrix} \tag{11}$$

In this case, the covariance matrix is dense, and the entries exponentially decrease in the distance from the diagonal.

Based on the coefficient and covariance matrix, I calculate the theoretical spillover table and the spillover indices by applying the Eqs. (1-8). The following Diebold-Yilmaz spillover table (Table 1) represents the first system where the B is sparse with one dominant time series, and the Σ is the identity matrix:

Table 1: Diebold-Yilmaz spillover table for sparse coefficient matrix B and sparse covariance matrix Σ

	TS1	TS2	TS3	TS4	TS5	FROM
TS1	98.96	0.00	**1.04**	0.00	0.00	1.04
TS2	0.0	99.20	**0.80**	0.00	0.00	0.80
TS3	**0.00**	**0.00**	**100.00**	**0.00**	**0.00**	**0.00**
TS4	0.00	0.00	**1.04**	98.96	0.00	1.04
TS5	0.00	0.00	**1.31**	0.00	98.69	1.31
TO	0.00	0.00	**4.20**	0.00	0.00	**4.20**
NET	**-1.04**	**-0.80**	**4.20**	**-1.04**	**-1.31**	0

Note: *The coefficient matrix (B) is sparse with one dominant time series, and the covariance matrix (Σ) is the identity matrix.*

The spillover table shows that the third time series transmits the shocks to the other parts of the system (TO = 4.2), and it doesn't receive any shock (FROM = 0). It plays a dominant role in a system as a shock transmitter. There are no connections between the other time series of the network. The total spillover index is very low (SUM = 4.2/5 = 0.8%) due to the sparse coefficient and covariance matrix. In the second case, when the coefficient matrix is the same, but the covariance matrix is dense, the Diebold-Yilmaz spillover table is the following:

Table 2: Diebold-Yilmaz spillover table for sparse coefficient matrix B and dense covariance matrix Σ

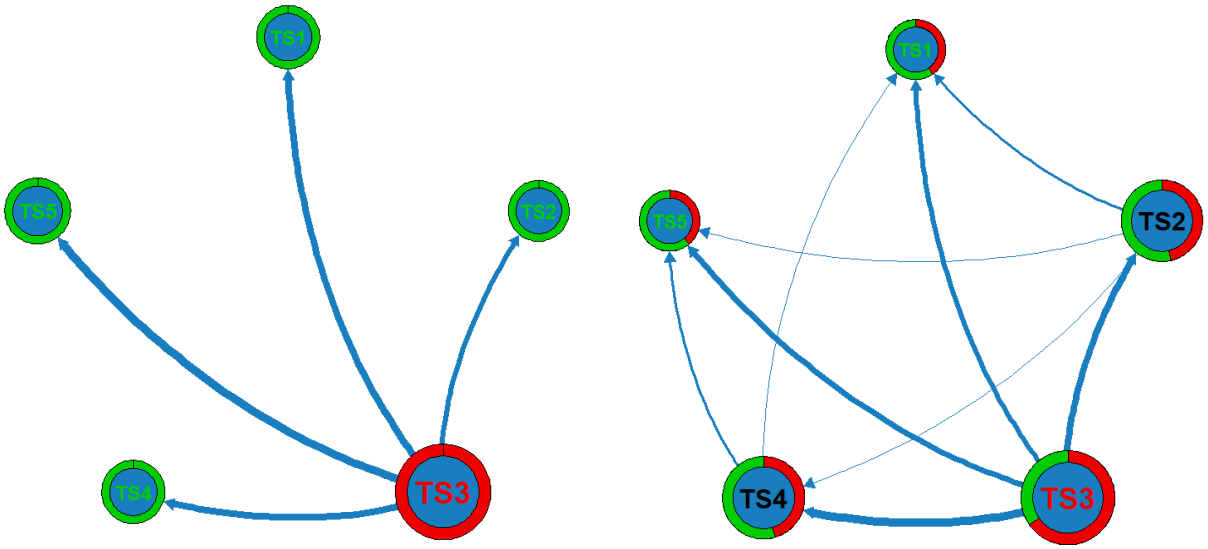
	TS1	TS2	TS3	TS4	TS5	FROM
TS1	94.42	4.09	**1.43**	0.06	0.00	5.58
TS2	3.63	90.65	**5.49**	0.22	0.01	9.35
TS3	**0.15**	**3.69**	**92.32**	**3.69**	**0.15**	**7.68**
TS4	0.01	0.23	**5.85**	90.30	3.61	9.70
TS5	0.00	0.08	**1.72**	4.12	94.08	5.92
TO	3.79	8.09	**14.48**	8.10	3.77	**38.24**
NET	**-1.80**	**-1.25**	**6.80**	**-1.61**	**-2.15**	0

Note: *The coefficient matrix (B) is sparse with one dominant time series, and the covariance matrix (Σ) is a Toeplitz-type matrix with $\rho = 0.3$.*

The third time series dominated the others in this case, too, but the FROM spillover value is not equal to zero (FROM = 7.68) due to the non-diagonal covariance matrix. Besides the connections between the other part of the network, all the time series has positive TO and FROM values. Regarding these network characteristics, the total spillover is higher than in the earlier case (SUM = 38.24/5 = 7.6%).

Most of the articles represent the network with graphs (besides the spillover table) to interpret the connections between the time series and the key participants of the network more easier. In the following two graphs (Fig. 1), I illustrate the two networks with sparse and dense covariance matrices:

Figure 1: Theoretical Diebold-Yilmaz networks for sparse and dense Σ



Notes: First (Second) DGP-based network is in the left (right) chart. The coefficient matrix (B) in both cases is sparse with one dominant time series, the covariance matrix (Σ) is the identity matrix in the first case, and a Toeplitz-type matrix with $\rho = 0.3$ in the second case. Nodes represent the time series. Node sizes show the SUM values. The green (red) slice of the pie charts represents the share of the FROM (TO) spillover indices of the time series. Edge directions mark the net spillover indices; their thickness represents the magnitudes. All of the edges are shown on the charts. Nodes with green letters represents time series where $FROM/SUM \geq 55\%$, reds where $TO/SUM \geq 55\%$.

To get a relevant conclusion from these figures, it is essential to clarify the interpretation of the graphs. These graphs encompass a wealth of information. The nodes represent the times series. The pie charts' colors around the nodes represent the proportion of shocks transmitted (TO spillover value marked with red color) and received (FROM spillover value marked with green color) by the time series. The last information related to the nodes is the SUM value, which is indicated by the size of the nodes. The edges represent the NET spillover indices of the time series, and their thickness represents their magnitudes. As in the 5-node network, there are just 10 NET pairwise connections; I present all of them. Nodes with green letters represent time series where $FROM/SUM \geq 55\%$, reds where $TO/SUM \geq 55\%$. These graphs aim to identify the key participants of the system (both big shock transmitters and receivers). The SUM indicator measures the centrality of a given time series in the system, regardless of the direction of the shock transmission. Nodes with red (green) letters represent the biggest shock transmitters (receivers).

In the first graph, it's clearly visible the dominance and the shock transmitting role the third time series. The lack of connections between the other time series shows the sparsity of the coefficient and the inverse covariance matrices. In contrast to this, the network on the second graph is dense due to the nonlinear relationship between the B and Σ matrices and the Diebold-Yilmaz spillover table, and the dense structure of the covariance matrix.

These examples with the spillover tables and the network graphs illustrate the nonlinear relationship between the B and Σ and Θ and the role of the coefficient and the covariance matrix in the Diebold-Yilmaz framework.

2.3 VAR estimation for high-dimension time series

Estimating high dimensional VAR models poses practical challenges to researchers: On one hand, a problem of dimensionality is caused by incorporating a huge number of time series and modeling higher-order autoregressive processes (Kock and Callot, 2015; Davis et al., 2016; Basu et al., 2019). Second, the temporal dependence structure in the sparse VAR model gives rise to some theoretical challenges (Basu and Michailidis, 2015). It is a challenging issue to determine which variables and (their) lags are relevant when the sample size is moderate (relative to dimensionality and lag number) (Nicholson et al., 2017; Hecq et al., 2023).

The heavy parameterization is a serious drawback of the traditional VAR estimation, and this limits its applicability for economics and financial high-dimension time series modeling (Hecq et al., 2023). The existing data lack adequate information for efficient parameter estimation using standard ordinary least squares (OLS) and maximum likelihood methods. This leads to noisy parameter estimates, potential instability in predictions, and challenges in providing clear interpretations of temporal dependence. Besides that, if $T < J^2p$, equation- by-equation least squares is not even feasible (Basu and Michailidis, 2015; Kock and Callot, 2015). Addressing the estimation of VAR model parameters is essential for effectively modeling high-dimensional time series data.

Recent advances in high-dimensional time series modeling have established that estimating a VAR model with relatively few samples is possible even when least squares estimation is not appropriate (Kock and Callot, 2015; Basu and Michailidis, 2015; Davis et al., 2016). It needs to impose a special structure on the coefficient matrix to estimate it consistently. It is generally believed that, for most economic and financial applications (Fan et al., 2011b; Giannone et al., 2021), the true model of the system of time series is sparse, only a small, unknown subset of the variables have significantly non-zero coefficients, and all the other variables have negligible (or even exactly zero) coefficients (Kock and Callot, 2015; Demirer et al., 2018; Hecq et al., 2023). Therefore, it is preferable to fit a sparse VAR model, with many AR parameters set to zero to more accurately approximate the sparse data generating process. An additional advantage of sparse models is that they allow for an ultra-high dimensionality of time series (Fan et al., 2011a, 2014).

The two most common methods to address the high-dimension problem are to apply pure shrinkage (L2-penalization) or to estimate with pure selection (L1-penalization).

The most prominent algorithm among penalized estimators is the LASSO (least absolute shrinkage and selection operator), blending the two core concepts; it shrinks and selects together. The LASSO approach was proposed by Tibshirani (1996). Theoretical aspects of this method have been thoroughly examined since the publication of the seminal paper; see the following references from machine learning literature Zou (2006), Meinshausen and Bühlmann (2006), Bickel et al. (2009).

The LASSO method incorporates a penalty based on the absolute value of the parameters into the objective function of least squares as the estimation problem in Eq. (12) shows:

$$\underset{\beta ij}{\operatorname{argmin}} \left[\sum_{t=p+1}^{T} (y_{jt} - \sum_{i=1}^{p} \sum_{j=1}^{J} \beta_{ij} y_{j,t-i})^2 + \lambda_j \sum_{i=1}^{p} \sum_{j=1}^{J} |\beta_{ij}| \right] \tag{12}$$

The L1 penalty serves a dual purpose. Initially, by incorporating this penalty into the objective function, it enables the estimation of β even when the number of coefficients in

the VAR model surpasses the time series length. Subsequently, it induces sparsity in the estimated autoregressive parameters B by enforcing some elements of β exactly equal to zero. The value of the penalty clearly determines the amount of this selection: the larger of the regularization parameter λ, the sparser the coefficient matrix (B). The concept behind the LASSO estimation is to shrink the OLS estimated parameters (elements of B matrix) towards zero to reduce variance (Kock and Callot, 2015; Demirer et al., 2018).

As Eq. (12) shows, LASSO penalizes all parameters equally. If it were feasible to assign a higher penalty to the truly zero parameters (β_{ij}, which are zeros in the data generating parameters) compared to the non-zero ones, one would expect a better estimation and forecast performance. Zou (2006) improved the LASSO regression by introducing an additional weight parameter. The adaptive LASSO method is a two-step algorithm that uses a first-step estimator result (usually the Ridge) to weight the lagged time series. If the first-step estimator classifies one of the estimated parameters as zero $\beta_{ij} = 0$, it is not included in the second step of the estimation. This concept results in a smaller size of the problem in the second phase. It also has sparse solutions and an even more efficient estimation algorithm than the original LASSO. Shortly, it generalizes the popular LASSO method.

The following equation shows the adaptive LASSO estimation problem:

$$\underset{\beta ij}{\operatorname{argmin}} \left[\sum_{t=p+1}^{T} (y_{jt} - \sum_{i=1}^{p} \sum_{j=1}^{J} \beta_{ij} y_{j,t-i})^2 + \lambda_j \sum_{i=1}^{p} \sum_{j=1}^{J} w_{ij} |\beta_{ij}| \right] \tag{13}$$

where ($w_{ij} = \frac{1}{|\beta_{ij}^{ridge}|}$) is the adaptive penalty term.

A nice feature of the adaptive LASSO is what Fan and Li (2001) call the oracle property, which means that the adaptive LASSO correctly identifies the zero components of the coefficient matrix with probability tending to 1 (setting all zero parameters in the data generating process exactly equal to zero through the estimation (Zou, 2006)). This property of the algorithm is true both in cross-section regression and high-dimension VAR(p) settings (Kock and Callot, 2015).

Compared to linear regression (cross-sectional regression), analysis of high-dimension VAR models requires critical consideration from a specific point. Since the response variable is multivariate in the VAR systems, choosing the loss function in the estimation step (sum of squared residuals or negative log-likelihood) is challenging (Basu and Michailidis, 2015). In a multivariate VAR setting, the applied loss function plays a critical role as it influences the regularized VAR model parameter estimation and accuracy, even if the multivariate time series have Gaussian errors[1]. This phenomenon becomes more serious when the multivariate error process has highly correlated components (Basu and Michailidis, 2015). The reason is that the negative log-likelihood function considers the inverse covariance matrix of the error term, but the OLS (based on the sum of squared residuals) estimation is not. This distinction will generally lead to different estimations of high-dimensional VAR models unless the unknown covariance matrix equals of the identity matrix[2] (Davis et al., 2016; Basu and Michailidis, 2015).

In estimating the LASSO-VAR method, a standard, and most common approach is to apply the uni-variate OLS regression estimation separately on each response variable

[1] which is a common assumption in econometrics

[2] or a scalar multiple of identity matrix

(Kock and Callot, 2015)). This essentially involves restructuring the VAR model into a linear regression model, where the current values of the time series are regarded as the response variable, and lagged values are considered as the explanatory variables. After that, researchers apply penalized estimation equation-by-equation separately, often with different tuning parameters for different equations (Lee and Liu, 2012; Basu and Michailidis, 2015; Kock and Callot, 2015; Davis et al., 2016; Deshpande et al., 2019). Hereafter I refer this estimation as a row-wise LASSO (RW LASSO), and the adaptive version of this estimation as an adaptive row-wise LASSO (ARW LASSO)[3].

Although simple and popular, this estimation strategy ignores the joint information among the response variables. It does not consider the dependence between the time series (the contemporaneous correlation), which may lead to poor predictive performance under certain circumstances. A research of Lee and Liu (2012) points out that the correlation between error terms of the VAR model has a significant impact on the estimated parameters in a penalized regression. Disregarding the serial correlation in the regularization step can be dangerous because the theoretical risk bounds of the estimation depend on the degree of contemporaneous cross-correlation between the time series (Song and Bickel, 2011; Basu and Michailidis, 2015).

2.4 Adaptive Joint LASSO

In order to achieve more efficient estimation, I propose an extended penalized maximum likelihood method for VAR estimation. To develop a better estimator for B and Ω, I consider incorporating the inverse covariance matrix of the error term (Ω) in the estimation process, and regularize both the coefficient and the inverse covariance matrix, following Rothman et al. (2010), Lee and Liu (2012) and Barbaglia et al. (2020). Unlike previous studies, I estimate the model with adaptive-LASSO, minimizing the following weighted negative log-likelihood function:

$$(\hat{B}, \hat{\Omega}) = \underset{B,\Omega}{\operatorname{argmin}} \left[\frac{1}{2J} trace\left[((Y - XB)\Omega(Y - XB))' \right] - \frac{1}{2} \log |\Omega| + \lambda \sum_{s=1}^{p} \sum_{i,k=1}^{J} w_{s,ik} |\beta_{s,ik}| + \gamma \sum_{i=k}^{J} w_{ik} |\omega_{ik}| \right] \tag{14}$$

where λ and γ are the two hyperparameters of the method.

I apply the adaptive version of LASSO for two reasons. Firstly, as Kock and Callot (2015) stands, only the row-wise adaptive LASSO has an oracle property for the high-dimension VAR models. Second, as Lee and Liu (2012) pointed out, the adaptive version of the multivariate regression with covariance estimation (MRCE) approach has better parameter estimation accuracy for cross-sectional data.

I take the weights of the adaptive LASSO as a reciprocate of the L2 estimator, where the ridge estimator is based on the following minimization:

[3]Eq. (12) is related to row-wise Lass, and Eq. (14) to the adaptive row-wise LASSO estimation

$$(\hat{B}, \hat{\Omega}) = \underset{B,\Omega}{\operatorname{argmin}} \left[\frac{1}{2J} trace \left[(Y - XB)\Omega(Y - XB)' \right] - \frac{1}{2} \log |\Omega| + \lambda_2 \sum_{s=1}^{p} \sum_{i,k=1}^{J} \beta_{s,ik}^2 + \gamma_2 \sum_{i=k}^{J} \omega_{ik}^2 \right] \tag{15}$$

I refer to my new algorithm as the adaptive joint LASSO (AJ LASSO) method. The essence of the model is the maximum likelihood estimation with joint penalization on both B and Ω matrices. The adaptive joint LASSO extends the MRCE method (Rothman et al., 2010) and penalized maximum likelihood LASSO method (Barbaglia et al., 2020) with the adaptive estimation step. It has two substantial advantages compared to the commonly used adaptive row-wise LASSO method. Firstly, it incorporates the inverse covariance matrix (Ω) into the coefficient matrix (B) estimation. Based on the earlier results of the MRCE literature (Rothman et al., 2010; Lee and Liu, 2012) this feature of the estimation can be important if the inverse covariance matrix has a dense or band structure ($\Omega \neq I_J$). Second, besides the coefficient matrix (B) regularization, this approach penalizes the elements of the inverse covariance matrix (Ω) too[4]. It can be effective if the inverse covariance matrix is sparse ($\Omega = I_J$). Besides these advantages, I highlight that the Diebold-Yilmaz spillover table depends on both B and Ω, and the strength of my method is that it jointly estimates these matrices. Mention the disadvantages of my model as well, it uses only just one-one hyperparameter for the coefficient and inverse covariance matrix estimation, while the row-wise LASSO optimizes the selection equation-by-equation. If some time series dominate the system the row-wise LASSO can be a better choice for the estimation of B.

2.5 Moving Block Bootstrap method

By constructing bootstrap confidence intervals for spillover measures on macro (total spillover index) and micro levels (net pairwise spillover indices) of the estimated network, I can distinguish significant spillover changes from others. As a result, we can evaluate the magnitude and statistical significance of the shocks. Following Buse et al. (2022) and Greenwood-Nimmo and Tarassow (2022), a bootstrap-based analysis is taken to explore the uncertainty of the estimated spillover measures. The choice of the bootstrap method needs to reflect the dependency structure inherent in the analyzed time series.

Various bootstrapping techniques have been designed in the last two decades to preserve the asymptotically valid inference for VAR estimation (Lahiri and Lahiri, 2003). The most common methods in the VAR context are recursive-design wild bootstrap, fixed-design wild bootstrap, and pairwise bootstrap (Brüggemann et al. (2016)). In my analysis, I apply the residual-based moving-block bootstrap (MBB) method developed by (Brüggemann et al., 2016). Both Buse et al. (2022) and Greenwood-Nimmo and Tarassow (2022) applied this method in the DY framework due to its desirable features. First, it doesn't require knowledge about the distribution of the DY spillover indices, it is suitable for small samples, and third, it remains valid under conditional heteroscedasticity (Greenwood-Nimmo and Tarassow, 2022). For the validity of the residual moving-block bootstrap algorithm, I refer to Brüggemann et al. (2016).

[4]row-wise LASSO and adaptive row-wise LASSO only penalize B

The logic behind the MBB approach is to simulate $s = 1, 2, \ldots, S$ bootstrap samples of the residuals from the estimated VAR model and re-estimate the model on the bootstrap sample too. After re-estimating the VAR model for every bootstrap sample (MBB-VAR), I calculate the DY spillover indices, and the resulting bootstrap spillover metrics (SUM, FROM, TO, and NET indices) can be used for inference. For the VAR estimation, I apply the adaptive joint LASSO algorithm.

The MBB-VAR method efficiency depends on the sample size and bootstrap sample. We can estimate the confidence intervals more accurately if the sample size and the bootstrap sample are higher. Through my empirical analysis, I conducted 1000 bootstrap samples. I use 68% confidence interval for the network snapshots, and 95% when I analyze the structural changes in the total spillover index.

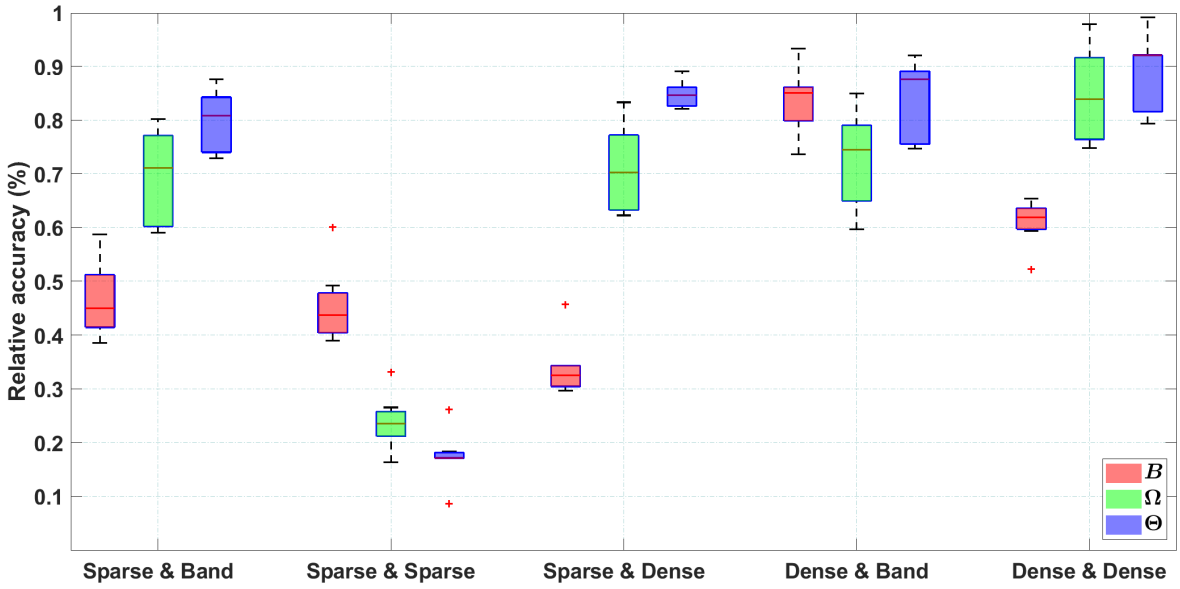
3 Chapter 3. Thesis contibution

3.1 Research method innovations

Regularisation in the estimation step of the VAR model may be useful for high-dimension time series modeling. In Chapter 5 of my thesis, I extend the commonly used row-wise LASSO method. I propose a new regularization method that is useful in the DY framework, as it is able to shrink and select the essential coefficients of the VAR and simultaneously account for possible sparsity in the distribution of the VAR model errors. In Chapter 5.2, I perform an extensive Monte Carlo simulation study I to present my proposed estimator's statistical properties. I compare my new approach with the most widely used VAR model estimation methods (OLS, RW LASSO, ARW LASSO).

To illustrate the results of the Monte Carlo simulation study, I illustrate the results on Figure 2. The figure represents the relative MAEE (mean absolute estimation error) accuracy for my method in B, Ω, and Θ estimation for the selected DGPs.

Figure 2: Summary of relative estimation accuracy of the AJ LASSO algorithm compared to the better RW LASSO for different coefficient (B) and inverse covariance (Ω) matrices



Notes: Red, green, and blue boxplots represent the relative estimation accuracy of the AJ LASSO algorithm compared to the better RW LASSO for the coefficient (B), inverse covariance (Ω), and DY spillover (Θ) matrices. The estimation error is measured by MAEE. The cases on the x-axes are the following: sparse B and band Ω; sparse B and sparse Ω; sparse B and dense Ω; dense B and band Ω; dense B and dense Ω. T, J, and ρ are varies through the cases.

The main findings are the following observations:

1. Based on Figure 2, in all the five cases (sparse and dense B; band, sparse and dense Ω), the AJ LASSO estimator outperforms RW LASSO estimators significantly in both B, Ω and Θ estimation.

2. For the basic setup (sparse coefficient B and band inverse covariance Ω matrices), simulation results indicate that in highly correlated systems, the relative perfor-

mance of the AJ LASSO method performs remarkably well, especially for the coefficient matrix (B) estimation.

3. The accuracy gain is sensitive to the covariance structure, in case of sparse Ω – the common assumption in theoretical econometrics – setting it is the highest (between 75-90% for Θ estimation). The results demonstrate that the new method efficiently accounts for the characteristics of the sparse Ω in the estimation via the jointly penalized (regularization on both B and Ω) objective function.

4. The relative accuracy is high in the case of the dense Ω settings, too (compared to the basic band Ω set), especially for the estimation of coefficient matrix (B). Incorporating dense Ω into the B estimation process is the critical point of the new method efficiency. The superior accuracy for the DY spillover matrix estimation (Θ) is still present with dense Ω.

5. In both sparse and dense Ω cases, the gain is also significant in the low-dimension systems ($J = 10$). That means the new method can be useful for DY network analysis in low-dimension systems too.

6. The accuracy gain of the AJ LASSO method is smaller for dense B but still significant. The smallest gains of the new estimator relative to the others are obtained for the dense B and band Ω case.

7. Even when AJ LASSO is only marginally superior (dense B), there are still benefits because the number of extreme outliers where the error is more than 20% in the estimation is significantly lower than for the other methods.

8. In general, the advantage in the estimation accuracy becomes much greater in shorter time series. Relative estimation accuracy is sensitive to the dimension of the system, the highest in high-dimension settings.

Based on these results, AJ LASSO can be useful for high-dimension VAR system (and DY network) modeling. In Chapter 6 of my thesis, I use my proposed method to estimate a VAR model for the DY volatility and illiquidity network.

Combining the DY framework for network modeling with event study methodology may be useful at turbulent times. In Chapter 6.1.3 of my thesis, I extend the original Diebold-Yilmaz framework with an event study tool to provide more insights into the contagion channels appearing in the networks during structural changes.

I introduce a formal test to compare the distribution of the FROM, TO, NET, and SUM (Eqs. (5-8)) spillover indices over time. I integrate the bootstrap algorithm into the DY framework's estimation step to determine whether the DY spillover indices' empirical distribution on different periods is equal. The gain of this method is that we can use it not just for analyzing the significant structural changes in the network on a macro level (total spillover index) but on a micro level (pairwise spillover indices) too. Combining the residual-based MBB Brüggemann et al. (2016) method with the DY framework (MBB-based DY framework), I can investigate how the observed shocks transform the financial networks. By interpreting the system as a network, I can monitor the network on a daily basis and analyze structural changes with the combined method.

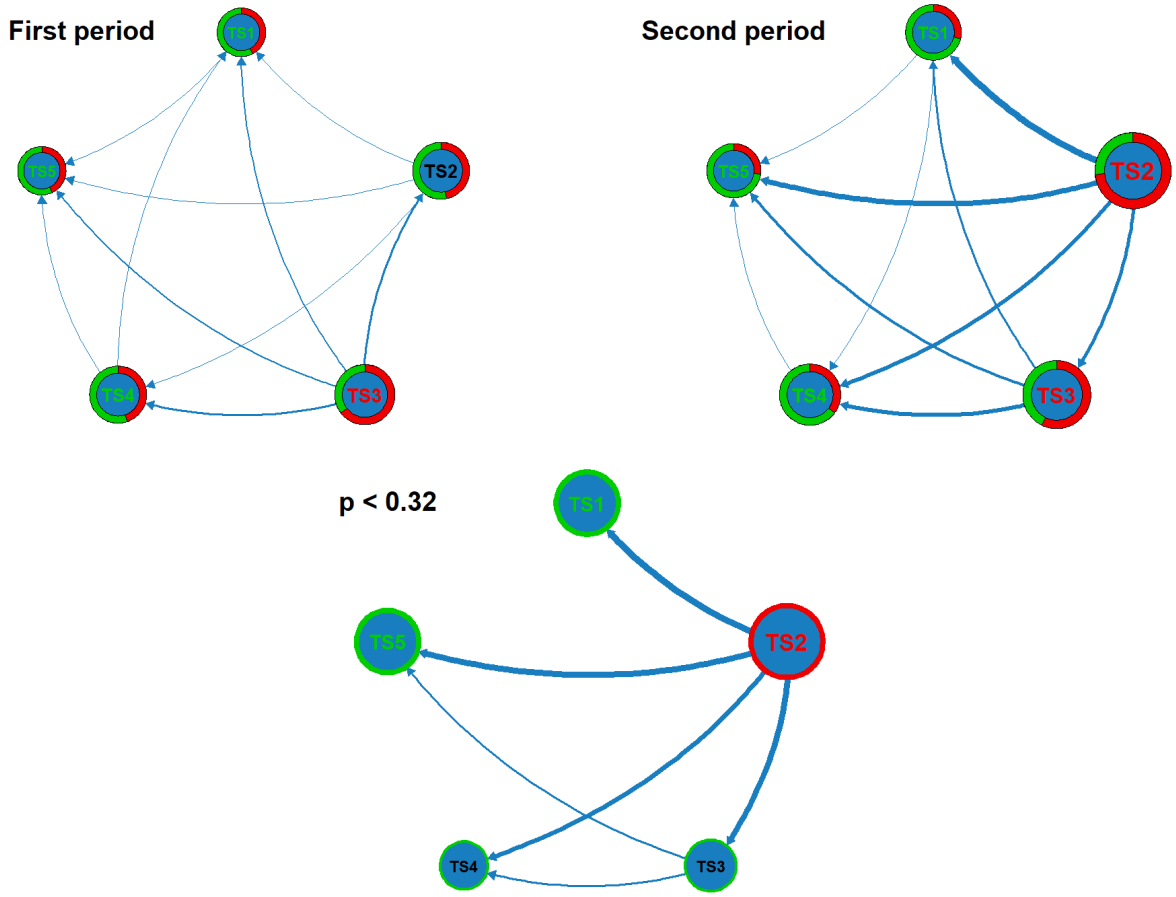
To illustrate my bootstrap method's usability in structural shock analysis, I estimate the Diebold-Yilmaz framework on two networks and estimate the significant differences with the Moving Block Bootstrap Method. Let's consider a low-dimension system (J=5) where the data-generating process is the same as the second case in Chapter 1. The coefficient matrix B is sparse, with one dominant time series (Eq. (9)), and the covariance matrix (Σ) is a Toeplitz-type matrix with $\rho = 0.3$ (Eq. (11)). To analyze the structural change in a network, I add a shock to the system. The following coefficient matrix (B) represents the data-generating process after the shock[5]:

$$B = \begin{bmatrix} 0.40 & \mathbf{0.12} & 0.08 & 0 & 0 \\ 0 & 0.40 & 0.07 & 0 & 0 \\ 0 & \mathbf{0.12} & 0.40 & 0 & 0 \\ 0 & \mathbf{0.11} & 0.08 & 0.40 & 0 \\ 0 & \mathbf{0.13} & 0.09 & 0 & 0.40 \end{bmatrix} \tag{16}$$

The NET values changed for every time series in the following way: -5.31, 15.35 -0.79, -4.11, - 5.25. These numbers illustrate the change in the network. During the structural change, the shock-transmitting role of the second time series increased. Besides that, the shock-receiving role strengthened for the first, fourth, and fifth time series. To illustrate a structural change more clearly, I show this change in a Fig. 3, which is similar to Fig. 1, but more complex.

[5] the covariance matrix is the same after the shock

Figure 3: Structural change in the Diebold-Yilmaz network



Notes: The network before the shock (after the shock) is in the upper left (right) chart. In the first case, the coefficient matrix (B) is sparse with one dominant time series (TS1). In the second case, it's also sparse with two dominant time series (TS1, TS2). The covariance matrix (Σ) is a Toeplitz-type matrix with $\rho = 0.3$ in both cases. The bottom chart shows their difference at 32% significance levels. Nodes represent the time series. ***Top panels:*** Node sizes shows the SUM values. The green (red) slice of the pie charts represents the share of the FROM (TO) spillover indices of the time series. Edge directions mark the net spillover indices; their thickness represents the magnitudes. All of the edges are shown on the charts. ***Bottom panel:*** Green (red borders represent a significant decrease (increase) of the node's net spillover index. Smaller (Larger) sizes than the average size represent a significant decrease (increase) in SUM values. ***Nodes:*** Nodes in the upper charts with green letters represent time series where $FROM/SUM \geq 55\%$, reds where $TO/SUM \geq 55\%$. In the bottom chart, green letters represent time series where the FROM/SUM ratio increased more than 10% and reds where TO/SUM increased more than 10%. ***MBB-VAR:*** Significance levels are estimated with moving-block bootstrap method (block size = 15, simulation run = 1000). The information of the Diebold-Yilmaz network is calculated from a $T = 1000$, adaptive joint LASSO-VAR(2) estimation.

To get a relevant conclusion from these snapshots, it is important to clarify the interpretation of the network graphs. These graphs encompass a wealth of information. At this case, each snapshot figure combines three different charts. The top two are obtained from the original time series-based VAR estimations, while the bottom one displays the Moving Block Bootstrap-based significant differences between the first and second day

at 32% levels. Nodes represent the time series. The pie charts' colors around the nodes represent the proportion of shocks transmitted (TO spillover value marked with red color) and received (FROM spillover value marked with green color) by the time series. The last information related to the nodes is the SUM value, which is indicated by the size of the nodes. On both charts in the top panel, the edges represent the NET spillover indices of the time series, and their thickness represents their magnitudes. As in the 5-node network, there are 10 NET pairwise connections, and I present all of them. Nodes with green letters represent time series where $FROM/SUM \geq 55\%$, reds where $TO/SUM \geq 55\%$. The node sizes and the edge thickness are rescaled based on the first and second periods; the biggest node size in the bottom charts represents the maximum SUM value in the two periods, and the smaller one represents the minimum SUM value. The thickest edge in the bottom charts represents the maximum NET value, and the least thicker represents the minimum NET value.

The bottom graphs indicate the significant differences between the SUM, FROM, TO, and NET measures shown by the charts of the two periods at the top. In this graph, only the significantly changed edges are displayed. Their direction and thickness indicate the direction and magnitude of the impact that the structural change generated during the periods. Here, the red (green) color of the node borders represents if there was a significant change based on the MBB method in the transmitted (received) shocks. That means the border color is red (green) if the net spillover index increased (decreased) between the selected days. If the node's size is above (below) the average size, a significant increase (decrease) happens in the SUM spillover value. Nodes with green letters represent time series where the FROM/SUM ratio increased more than 10%, and reds where TO/SUM increased more than 10%. The SUM indicator roughly translates to the centrality of a given time series in the system.

These snapshots aim to identify the key participants of the system (both big shock transmitters and receivers) and analyze how these roles change during specified days. Besides that, I can separate the significant and insignificant changes in the network with the Moving Block Bootstrap method. The SUM indicator measures the centrality of a given time series in the system, regardless of the direction of the shock transmission. Nodes with red (green) letters represent the biggest shock transmitters (receivers). Increased size in the top panel marks the elevated role of the time series in the system, while the smaller size means a less important participant in the network. I interpret the significant changes in the transmitter role with significant red pie slice growth on the top panel (red border on the bottom). Similarly, the receiver role changes significantly with significant green pie slice growth on the top panel (green border on the bottom panel).

Figure 3 illustrates the structural change in the network. The upper charts show the increased total spillover in the system: the edges are thicker, especially that are related to the second time series. Besides that, the node sizes are higher for most of the time series. The node of the second time series shows the changed role of this network's participant. However, the two upper charts only show the increased total spillover and whose role changed in the system. Based on these graphs, we can't decide which change was significant and which was not. The bottom part of the figure solves this problem. This chart illustrates that the shock-transmitting role significantly increased for the second time series, and for the other four participants in the system became more shock-transmitting due to the structural change.

3.2 Conceptual innovations

To address the gap in the empirical literature, in Chapter 6.2, I present a characterization of the illiquidity connectedness of a US financial network with my proposed AJ LASSO method in the Diebold-Yilmaz framework. I investigate the significant similarities and differences between the volatility and illiquidity network on the macro and micro levels with the combined MBB-based DY framework.

The empirical results suggest that illiquidity total spillover indices are also relevant in analyzing systemic risk as they behave differently than total volatility connectedness. On the macro level, both indices react to shocks in the financial system, but even in tranquil periods, the dynamics might vary. Furthermore, if I analyze the pairwise connections between FIs periodically, dominant illiquidity shock transmitters and receivers appear in the system. I show that the volatility and illiquidity network differs not just on the macro level (total spillover index) but on the micro (pairwise net spillover indices) level too. I conclud that micro and macro-level illiquidity spillover indices display important information for the financial network.

In Chapter 6.3 of my thesis, I consider the critical financial events of the GFC, as Diebold and Yılmaz (2014), with the analytical purpose of testing the performance of my new event study framework. Applying the MBB-based DY framework, I analyze the volatility and liquidity shocks, focusing on four key events of the GFC. Besides that, I provide daily snapshots of the financial network to illustrate that illiquidity spillovers act as contagion channels during turmoil time.

In my empirical analysis, I show that event study analysis tools like visualization, dynamic measures, and the new formal statistical test help to identify key events, FIs, and contagion channels in the financial system. Furthermore, I find that on both macro and micro level analysis, the illiquidity networks of FIs during the GFC display relevant information. I show that while volatility networks do not always react to financial turmoil, troubled FIs become the main shock transmitters in illiquidity networks. Illiquidity connections act as financial linkages described by Acemoglu et al. (2015); thus, they indicate the spread of contagion in the system during turmoil. Severe financial shocks temporarily altered the illiquidity network, and dominant shock transmitters appeared in the system. My empirical results are in line with Gai and Kapadia (2010) findings, too. I find that the impact of a shock depends on which node of the network it hits. I conclude that the DY framework with the MBB approach is a powerful tool for regulators to identify potential failure cascades and SIFIs. Through event study analysis, they can observe shocks that lead to significant changes in the financial network.

The analysis in Chapter 6 of my thesis contributes to the empirical network analysis literature at three related points. First, as far as I know, I am the first to compare the volatility and illiquidity network with the MBB-based DY framework. Second, I am the first who apply the MBB-based DY framework to analyze the effect of key events in the network. An important limitation of the original DY framework is the absence of a formal statistic to test whether the changes in the spillover matrix are significant. Lastly, I show that DY illiquidity spillover better tracks the dominant shock transmitters in the system and signals during financial turmoil than volatility spillover both in the macro and micro

level of the network. My empirical results strengthen the findings of Gai and Kapadia (2010) and Acemoglu et al. (2015), because I found that the impact of a shock depends on which node of the network it hits, and the network transformation's persistence depends on the shock size.

4 The author's Publications

Badics, M. Cs., Huszár, Zs. R., and Kotró, B. B. (2023). The impact of crisis periods and monetary decisions of the Fed and the ECB on the sovereign yield curve network. *Journal of International Financial Markets, Institutions and Money*, 88, 101837

Badics, M. Cs. (2016). Cross-correlation analysis of international foreign exchange markets: An EEMD-based approach. in *10th International Conference on Computational and Financial Econometrics*, ISBN 978-9963-2227-1-1

Badics, M. Cs., and Kutasi, D. (2016). Valuation methods for the housing market: evidence from Budapest. *Acta Oeconomica*, 66(3), 529-548.

Badics, M. (2015). Integrating independent component analysis-based SSA with neural network for stock price prediction. in *International work-conference On Time Series*, 635-636, ISBN 9788416292202

Badics, M., Berlinger, E., Illés, F., and Tuza, Á. (2015). Technical Analysis, neural networks and logoptimal portfolios. in *Mastering R for Quantitative Finance*, Chapeter 10, 227-256, ISBN 139781783552078

Badics, M. Cs. (2014). Tőzsdei idősorok előrejelzése adatbányászati módszerekkel. *Hitelintézeti szemle*, 13(4), 207-227.

References

Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. *American Economic Review*, 105(2):564–608.

Barbaglia, L., Croux, C., and Wilms, I. (2020). Volatility spillovers in commodity markets: A large t-vector autoregressive approach. *Energy Economics*, 85:104555.

Basu, S., Das, S., Michailidis, G., and Purnanandam, A. (2019). A system-wide approach to measure connectivity in the financial sector. *Available at SSRN 2816137*.

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of lasso and dantzig selector.

Brüggemann, R., Jentsch, C., and Trenkler, C. (2016). Inference in vars with conditional heteroskedasticity of unknown form. *Journal of econometrics*, 191(1):69–85.

Buse, R., Schienle, M., and Urban, J. (2022). Assessing the impact of policy and regulation interventions in european sovereign credit risk networks: What worked best? *Journal of International Economics*, 139:103673.

Davis, R. A., Zang, P., and Zheng, T. (2016). Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics*, 25(4):1077–1096.

Demirer, M., Diebold, F. X., Liu, L., and Yilmaz, K. (2018). Estimating global bank network connectedness. *Journal of Applied Econometrics*, 33(1):1–15.

Deshpande, S. K., Ročková, V., and George, E. I. (2019). Simultaneous variable and covariance selection with the multivariate spike-and-slab lasso. *Journal of Computational and Graphical Statistics*, 28(4):921–931.

Diebold, F. X. and Yilmaz, K. (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal*, 119(534):158–171.

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.

Diebold, F. X. and Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134.

Elliott, M., Golub, B., and Jackson, M. O. (2014). Financial networks and contagion. *American Economic Review*, 104(10):3115–53.

Fan, J., Feng, Y., and Song, R. (2011a). Nonparametric independence screening in sparse ultra-high-dimensional additive models. *Journal of the American Statistical Association*, 106(494):544–557.

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.

Fan, J., Lv, J., and Qi, L. (2011b). Sparse high-dimensional models in economics. *Annu. Rev. Econ.*, 3(1):291–317.

Fan, J., Ma, Y., and Dai, W. (2014). Nonparametric independence screening in sparse ultra-high-dimensional varying coefficient models. *Journal of the American Statistical Association*, 109(507):1270–1284.

Fengler, M. R. and Gisler, K. I. (2015). A variance spillover analysis without covariances: What do we miss? *Journal of International Money and Finance*, 51:174–195.

Gabauer, D., Gupta, R., Marfatia, H., and Miller, S. M. (2020). Estimating us housing price network connectedness: Evidence from dynamic elastic net, lasso, and ridge vector autoregressive models. *Lasso, and Ridge Vector Autoregressive Models (July 26, 2020)*.

Gai, P. and Kapadia, S. (2010). Contagion in financial networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466(2120):2401–2423.

Giannone, D., Lenza, M., and Primiceri, G. E. (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica*, 89(5):2409–2437.

Greenwood-Nimmo, M. and Tarassow, A. (2022). Bootstrap-based probabilistic analysis of spillover scenarios in economic and financial networks. *Journal of Financial Markets*, 59:100661.

Han, F., Lu, H., and Liu, H. (2015). A direct estimation of high dimensional stationary vector autoregressions. *Journal of Machine Learning Research*.

Hecq, A., Margaritella, L., and Smeekes, S. (2023). Granger causality testing in high-dimensional vars: a post-double-selection procedure. *Journal of Financial Econometrics*, 21(3):915–958.

Kock, A. B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics*, 186(2):325–344.

Koop, G., Pesaran, M. H., and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of econometrics*, 74(1):119–147.

Lahiri, S. and Lahiri, S. (2003). *Resampling methods for dependent data*. Springer Science & Business Media.

Lee, W. and Liu, Y. (2012). Simultaneous multiple response regression and inverse covariance matrix estimation via penalized gaussian maximum likelihood. *Journal of multivariate analysis*, 111:241–255.

Lütkepohl, H. (2013). Vector autoregressive models. *Handbook of research methods and applications in empirical macroeconomics*, 30.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso.

Nicholson, W. B., Matteson, D. S., and Bien, J. (2017). Varx-l: Structured regularization for large vector autoregressions with exogenous variables. *International Journal of Forecasting*, 33(3):627–651.

Pesaran, H. H. and Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics Letters*, 58(1):17–29.

Rothman, A. J., Levina, E., and Zhu, J. (2010). Sparse multivariate regression with covariance estimation. *Journal of Computational and Graphical Statistics*, 19(4):947–962.

Song, S. and Bickel, P. J. (2011). Large vector auto regressions. *arXiv preprint arXiv:1106.3915*.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.