

ESSAYS IN FINANCIAL ECONOMETRICS AND SPILLOVER ESTIMATION

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

vorgelegt von

Ruben Hipp

im FSS 2019

Abteilungssprecher: Prof. Dr. Jochen Streb
Referent: Prof. Dr. Carsten Trenkler
Korreferent: Prof. Dr. Carsten Jentsch

Tag der Verteidigung: 03.07.2019

Acknowledgements

First and foremost, I would like to thank my supervisor Carsten Trenkler for his support and guidance throughout the last years. I am truly indebted for the excellent research environment, and support he provided from the beginning of my bachelor thesis. I would also like to thank Carsten Jentsch for the research opportunities, encouragements, and his fantastic co-supervision. I am also grateful to Christian Brownlees, who has been a superb host at my research stay at the Universidad Pompeu Fabra. His welcoming and relaxed nature have been an inspiration for the pursuit of my academic goals.

During my job market phase, I experienced a fair and kind behavior of my letter writers Carsten Trenkler, Carsten Jentsch and Christian Brownlees. I particularly want to thank all of them for their work at this time, but also Christoph Rothe for his guidance as a job market coordinator.

In Mannheim, I had the pleasure to experience the courteous help of Matthias Meier, who showed interest in my work and offered amazing advice. My neighbor and colleague, André Stenzel, always provided strategic advice, and valuable help when it was needed. Their unconditional help and advice were very much appreciated.

I also want to thank my Ph.D. friends Florian Böser, Felix Brunner, Esteban Cattaneo, Tobias Etzel, Jasmin Fliegner, Niklas Garnadt, Karl Schulz, and many others who helped me with the completion of my thesis.

For making my Ph.D. life so much easier, I am thankful to Anja Dostert, Regina Mannsperger, and Sylvia Rosenkranz from the University of Mannheim, as well as Sandro Holzheimer, Marion Lehnert, Dagmar Röttsches, and Golareh Khalilpour from the CDSE.

Eventually, I want to thank my parents Ingrid and Gerhard, my brother David, my sister Marlene, my girlfriend Luisa, and all whom I forgot to mention.

Contents

| | |
|--|-----------|
| Acknowledgements | iv |
| 1 On Causal Networks of Financial Firms | 4 |
| 1.1 Introduction | 4 |
| 1.2 The Model | 8 |
| 1.2.1 Terminology and General Setup | 8 |
| 1.2.2 Structural Identification in a Nutshell | 10 |
| 1.2.3 Identification via Non-parametric Heteroskedasticity Modeling | 12 |
| 1.2.4 Extremum Estimator | 16 |
| 1.3 Causal Financial Connectedness | 18 |
| 1.3.1 A causal network model | 18 |
| 1.3.2 Forecast Error Variance Decompositions for SVARs | 20 |
| 1.3.3 Measures of Interconnectedness | 21 |
| 1.3.4 Data | 23 |
| 1.3.5 Estimation Strategy | 26 |
| 1.3.6 Empirical Results | 28 |
| 1.4 Conclusion | 39 |
| 2 On Time-Variation of Financial Connectedness and its Statistical Significance | 41 |
| 2.1 Introduction | 41 |
| 2.2 Preliminaries | 44 |
| 2.2.1 Notation | 44 |
| 2.2.2 Local stationarity | 47 |

| | | |
|----------|---|------------|
| 2.3 | Methodology | 49 |
| 2.3.1 | Local-Linear Estimator for Time-Varying VAR Coefficients . . | 49 |
| 2.3.2 | Local-Linear Estimator for a Time-Varying Innovation Covari- ance Matrix | 51 |
| 2.3.3 | Bandwidth Selection | 52 |
| 2.3.4 | Time-Varying Forecast Error Variance Decompositions | 55 |
| 2.3.5 | Inference with Bootstrap Confidence Intervals | 58 |
| 2.4 | Simulation Study | 59 |
| 2.5 | Time-Varying Financial Connectedness | 65 |
| 2.5.1 | Data | 65 |
| 2.5.2 | Empirical Results | 68 |
| 2.6 | Concluding Remarks | 72 |
| 3 | Estimating Large Dimensional Connectedness Tables | 73 |
| 3.1 | Introduction | 73 |
| 3.2 | Methodology | 76 |
| 3.2.1 | Generalized Forecast Error Variance Decompositions in a Nut- shell | 76 |
| 3.2.2 | Estimating Large Vector Autoregressions | 78 |
| 3.2.3 | Data-Driven Choice of λ and δ | 85 |
| 3.3 | Simulation Study | 88 |
| 3.3.1 | Data Generating Processes | 88 |
| 3.3.2 | Bias, Accuracy and ROC | 89 |
| 3.3.3 | Comparison of Different Regularization Methods | 92 |
| 3.4 | Empirical Application | 95 |
| 3.4.1 | Data | 96 |
| 3.4.2 | Results | 96 |
| 3.5 | Conclusions | 104 |
| A | Appendix to Chapter 1 | 105 |
| A.1 | Propositions and proofs | 105 |
| A.2 | Complementary Figures | 106 |

| | | |
|----------|---|------------|
| B | Appendix to Chapter 2 | 107 |
| B.1 | Proofs | 107 |
| B.2 | Derivation of local linear least square tvVAR | 109 |
| B.3 | Residual-based estimation of $\Sigma(\tau)$ | 111 |
| B.4 | Simulation: Data Generation | 113 |
| C | Appendix to Chapter 3 | 117 |
| C.1 | Complementary Graphs | 117 |

List of Figures

| | | |
|-----|--|----|
| 1.1 | Mean Causal Network for the 20 % strongest connections. Node colors indicate the strength of the average To-Connectedness of the respective stock. The analysis includes American Int. Group (AIG), American Express (AXP), Bank of America (BAC), Bank of NY Mellon (BK), Citigroup (C), Goldman Sachs (GS), JP Morgan Chase (JPM), Morgan Stanley (MS), Wells Fargo (WFC). | 30 |
| 1.2 | Comparison of Forecast Error Variance Decompositions with the generalized version as in Diebold and Yilmaz (2014) on the left and the structural version on the right. Observation at the 09/14/2008 for the 20 % strongest connections. Node colors indicate the strength of the average effect of a respective stock to others. | 32 |
| 1.3 | Total Average Connectedness for the FEVDs with structural decomposition (solid blue) and the generalized version of Diebold and Yilmaz (2009) (dotted green) | 33 |
| 1.4 | Total Average Connectedness for the causal network matrix G_t (solid blue), the impact matrix $A_{0,t}^{-1}B_t$ (dotted green) | 34 |
| 1.5 | From-Connectedness of G_t | 36 |
| 1.6 | To-Connectedness of G_t | 37 |
| 1.7 | Systemic Relevance for particular firms | 39 |
| 2.1 | Monte Carlo Simulation of A_t with 500 repetitions. Mean comparison of the local-linear estimator (2.12) (red) and the local-constant estimator <i>aka</i> kernel-weighted LS-estimator (dashed blue). The black line depicts the true DGP parameters. The grey area shows the Monte-Carlo bands. | 61 |

| | | |
|-----|--|----|
| 2.2 | Monte Carlo Simulation of Σ_t with 500 repetitions. Mean comparison of the local-linear estimator (2.12) (red) and the local-constant estimator <i>aka</i> kernel-weighted LS-estimator (dashed blue). The black line depicts the true innovation covariance entries. The grey area shows the Monte-Carlo bands. | 62 |
| 2.3 | Monte Carlo Simulation of D_t^{gH} with 500 repetitions. Mean comparison of the local-linear estimator (2.12) (red) and the local-constant estimator <i>aka</i> kernel-weighted LS-estimator (dashed blue). The black line depicts the true innovation covariance entries. The grey area shows the Monte-Carlo bands. | 63 |
| 2.4 | Distributional results for the distance between the oracle bandwidth and the cross validation selected bandwidth. The oracle bandwidth is the bandwidth which minimizes the norm to the true parameter curves. The CV bandwidths are chosen with the cross-validation criteria in (2.17) and (2.21). Distributions are kernel smoothed with a gaussian kernel. | 64 |
| 2.5 | Network graph of the mean network over time. The strongest 25% dependencies are shown. Colors indicate the out-degree of the respective institution. | 69 |
| 2.6 | Network graph of the network at the 30/06/2018. The strongest 25% dependencies are shown. Colors indicate the out-degree of the respective institution. | 70 |
| 2.7 | Average Connectedness (blue) over time. The grey areas depict the 90% and 99% bootstrap confidence interval. The red lines depict the selected bandwidth for the coefficient (solid) and the covariance (dashed) estimate. | 71 |
| 3.1 | Simulation results for 500 Monte-Carlo repetitions of <i>DGP 5</i> and $N = 100$. The first two panels show the mean difference and the norm of $\hat{D}^{gH} - D^{gH}$ respectively. The third panel shows the Frobenius norm of the estimate to the one matrix, i.e., $\ \hat{D}^{gH} - \mathbf{1}_{N \times N}\ $. The sample size T is on the x-axis. | 90 |

| | | |
|-----|---|-----|
| 3.2 | Simulation results for 500 Monte-Carlo repetitions of <i>DGP 5</i> and $N = 100$. The first two panels show the true positive rate (TPR) and false positive rate (FPR) respectively. The sample size T is on the x-axis. The right panel shows the receiver operator characteristic (ROC) for $T = 150$, which is FPR on the x-axis and TPR on the y-axis. The diagonal thin black line is the equivalent of a random estimate. | 91 |
| 3.3 | Growth rates of monthly industrial production with annual rates. . . | 97 |
| 3.4 | 10-fold cross-validation results for the tuning parameter α in the adaptive elastic-net. The blue curve shows for different α 's the minimal MSE with \hat{A}_{OLS} as the initial estimate in the weights of the individual penalties for different α s. The red curve shows the same for \hat{A}_{ENET} , as suggested by Zou and Zhang (2009) . Both curves show the results of the first 10-fold CV over values of λ | 99 |
| 3.5 | 10-fold cross-validation results for the tuning parameter of different covariance regularization methods. Values of α are on the x-axis and the minimal mean squared error on the y-axis. The two thresholding estimators tune δ in equation (3.14), the GLASSO tunes λ in equation (3.16), and the manual Ledoit-Wolf the shrinkage term. | 100 |
| 3.6 | Connectedness network for the respective periods. The size of the node relates to the respective From-Connectedness of the sector. The colors depict the To-Connectedness. The sector with the highest To-Connectedness is labeled. | 101 |
| 3.7 | Histogram of the From-Connectedness of the 88 three-digit level sectors. The 50% strongest sectors by weight are highlighted in red. . . . | 102 |
| 3.8 | Histogram of the To-Connectedness of the 88 three-digit level sectors. The 10% largest sectors by To-Connectedness in the pre-Great Moderation period are highlighted in red. | 103 |
| A.1 | Lead-lag correlation: $\text{corr}(C(G_t), \overline{\text{vola}_{t-k}})$. The average volatility at time t is taken over the firms. | 106 |
| B.1 | Monte Carlo Simulation with 500 repetitions of A_t . Cumulative distribution function of the norm of the local-linear estimator (2.12) (red) and the local-constant estimator minus the true parameter. . . . | 115 |

| | | |
|-----|--|-----|
| B.2 | Monte Carlo Simulation with 500 repetitions of Σ_t . Cumulative distribution function of the norm of the local-linear estimator (2.12) (red) and the local-constant estimator minus the true parameter. . . . | 116 |
| C.1 | Estimated CDF of the To-Connectedness of the 88 three-digit level sectors. | 117 |

List of Tables

| | | |
|-----|---|-----|
| 1.1 | US. financial institutions key figures in bn. US\$ | 29 |
| 1.2 | Important events concerning the overall connectedness of financial institutions | 31 |
| 1.3 | Important events on firm-level | 38 |
| 2.1 | Monte-Carlo simulation results for 200 repetitions and a 1000 bootstrap repetitions. Coverage rates are shown for the basic Hall bootstrap interval and the percentile interval. $\hat{\theta}$ denotes the estimation target \hat{A}_t and quantiles are determined by α | 64 |
| 2.2 | Key figures of financial institutions in bn. US\$ (source: Compustat, *interpolated missing values, **CS reported for 2000Q4 and 2017Q2***MUFG reported for 2001Q2 and 2018Q1,) | 67 |
| 3.1 | Simulation results for the regularization of A paired with the sample covariance. $ \hat{D}_{reg}^{gH} - D^{gH} / \hat{D}_{OLS}^{gH} - D^{gH} $ are shown for different N and T with 500 Monte-Carlo repetitions. <i>DGP 4</i> and <i>DGP 5</i> have 25 different random realizations of A and Σ_u | 93 |
| 3.2 | Simulation results for the regularization of Σ_u paired with the best performing adaptive elastic-net estimator. $ \hat{D}_{reg}^{gH} - D^{gH} / \hat{D}_{OLS}^{gH} - D^{gH} $ are shown for different N and T with 500 Monte-Carlo repetitions. <i>DGP 4</i> and <i>DGP 5</i> have 25 different random realizations of A and Σ_u | 94 |
| 3.3 | Summary of the estimation results for different levels of sectoral disaggregation. The columns labeled $C(D^{gH})$ show the estimated Average Connectedness. The columns labeled df show the sparsity level or percentage degree of freedom of the autoregression coefficient A | 101 |

General Introduction

This dissertation consists of three self-contained chapters. The common theme is the estimation of spillovers and their interpretation as networks. In financial econometrics, spillovers are mostly regarded as forecast error variance decompositions, i.e., a matrix constructed by a combination of vector autoregressive coefficients and an innovation covariance matrix. For a pristine interpretation of these spillovers, the estimation of the model's parameters is essential. The three chapters deal with different peculiarities of the estimation. All chapters contain an empirical application with separate insights. In Chapter 1, I deal with the estimation of structural matrices in order to obtain a proper representation of contemporaneous spillovers. In the empirical application, I look at the US financial system and volatility spillovers. This chapter has also been my job market paper and, thus, it is the central part of this dissertation. The second chapter is joint work with Carsten Jentsch. We investigate the estimation of time-varying spillovers in the setup of local stationarity. Empirically, we investigate financial spillovers between the biggest banks in the US, Europe, and Japan. The last chapter is joint work with Matteo Barigozzi and Christian Brownlees. We analyze the effect of high-dimensions on the estimation of spillover tables. An application on the industrial production index in the US aims to answer the question of whether the Great Moderation has changed spillovers between sectors. A compound of the respective abstracts follows to give a summary.

Chapter 1

On Causal Networks of Financial Firms: Structural Identification via Non-Parametric Heteroskedasticity Modeling.

We investigate the dependency structure of the US financial system. To account for contemporaneous relations, we introduce a novel non-parametric approach to identify structural shocks, which exploits the fact that most financial models empirically exhibit heteroskedasticity. The identification works locally and, thus, allows structural matrices to vary smoothly with time. A network application demonstrates the functionality of our approach by analyzing volatility spillovers of financial firms. The estimation of a causal network uncovers their contemporaneous directional dependencies. In this setup, we derive a new measure of systemic relevance, which highlights the most central institutions over the last two decades. Finally, we detect a change in the network architecture beginning in March 2017. This change is due to an increase in the systemic relevance of JPMorgan Chase, which is the most central institution as of June 2018.

Chapter 2

On Time Variation of Financial Connectedness and its Statistical Significance

[Diebold and Yilmaz \(2014\)](#) introduced a new way of estimating a unified network and successfully established a new standard in monitoring systemically important risk figures. Using a rolling window approach which sweeps through the sample, they implicitly assume that networks evolve smoothly over time. Although rolling windows are heuristically easy to interpret, their theory lacks asymptotics. In this paper, we aim to fill the gap of statistical inference and generalize the idea of financial connectedness within the framework of local stationarity. For this purpose, we propose a local linear kernel estimator for VAR coefficients curves. As the limiting distributions are too complex for practical applications, we propose a new bootstrap scheme for inference. In an extensive simulation study, we show the performance and

accuracy of this method. An application on financial volatility spillovers provides new insights into the dynamics of financial connectedness. We also advise on how to handle bandwidth selection.

Chapter 3

Estimation of Large Dimensional Connectedness Tables

Forecast error variance decompositions are a popular way to describe spillover networks in a unified fashion. For this unified network estimation, it is essential to include all variables. Obtaining such tables in a high-dimensional setup is challenging as they result from estimations of vector autoregressive coefficients and the covariance matrices. Naturally, we resort to regularized estimators to get consistent and accurate estimates. In this study, we carry out a comprehensive analysis of different regularization methods and introduce a novel way to regularize network tables. We compare these methods in an extensive simulation to shed light into their estimation uncertainty. We find that when the number of nodes in the network is large, ordinary least squares induces a bias for the entries of interconnectedness tables towards unity. Also, we show that regularization of the innovation covariance matrix is key to optimal performance. An application on sectoral spillovers of industrial production in the US from 1972 to 2007 gives insights into the amendments happening at the Great Moderation. With the assistance of regularization methods, we obtain a meaningful distribution of in- and outgoing connectedness. We find that a handful of sectors decreased their outgoing links to an extent which could have caused the Great Moderation.

Chapter 1

On Causal Networks of Financial Firms

1.1 Introduction

In the financial crisis of 2007, the term “systemic risk” has been popularized by the media after stock market reactions symptomized the entanglement of the financial system. American International Group (AIG) asked for liquidity support from the Federal Reserve, which decided for a bailout due to reactions on preceding defaults. Although the individual systemic risk was the reason for the bailout, the term “interconnectedness” better describes the origin of the problem: the tight entanglement of firms.

In particular, [Diebold and Yilmaz \(2015\)](#) point out the importance of this concept and link it to various aspects of risk. For example, portfolio risk and default connectedness are not only the sum of their increments but rather inherit their attributes by a combination of idiosyncratic risks and interdependencies. Similarly, gridlock (network) risk (see [Brunnermeier, 2009](#)) and systemic risk are results of the connectedness of firms. Alongside linear correlation measures, it is essential to uncover the interconnectedness of firms to allow for more sophisticated risk management.

Therefore, this study investigates the dependency structure of financial institutions in the US by estimating a network graph, which represents interconnectedness. In contrast to previous studies, we include a contemporaneous directed network by employing a structural vector autoregression (SVAR). We work under the assumption that parameters in our model are time-varying but still smooth enough for

estimation. In this framework, we find conditions under which structural parameters are identified.

To build intuition, we want to highlight the importance of information about the full set of contemporaneous dependencies. It is essential to know about direct dependencies caused by contractual obligations such as liability structures, deposits, and payments through the interbank clearing system. However, from a network perspective, the stability of a system is not only affected by the average intensity of pairwise associations, but it also depends on its network architecture (see [Acemoglu et al., 2015](#)). It is conceivable that different network architectures with the same average connectedness are exposed differently to individual shocks. Take, for example, the case of a cyclical dependency structure of banks. A significant shock propagates step by step through all banks such that a default of one firm starts a domino effect, which causes the whole system to default. In the case of a complete network, i.e., all banks are connected, the diversification of the links absorbs the shock. In a nutshell, the estimation of a unified network of directional dependencies is closely related to the question of a system's financial stability and, thus, it is of vital importance for policymakers in order to react adequately.

In principle, multivariate time series analysis could estimate the full set of directional dependencies by observing responses over an extended period. However, we require high-frequency data such that fast responses can be captured. Unfortunately, the only high-frequency data available is prone to the bid-ask spread and other problematic phenomena. Less-frequent data almost always misses out on direct responses and, thus, standard statistical tools can only visualize them as undirected co-movements or correlations. In other words, uncovering contemporaneous directional dependencies with standard time series analysis is usually a fruitless challenge. One way to overcome this issue is to ignore potential short-term dependencies by approximating them with an undirected covariance matrix of the error terms. Consequently, long-term dependencies, e.g., Forecast Error Variance Decompositions, are not fully understood.

A methodological contribution of this paper, then, is to introduce an approach which identifies contemporaneous dependencies. This approach exploits the fact that most financial models exhibit heteroskedasticity. That is, the covariance matrix of the error terms varies over time. We follow an approach free of any functional form and parametrize its local time-trend using a Taylor expansion. To decompose the

covariance matrix into its structural components, we propose two separate assumptions on the time-variation of the connectedness parameter and the idiosyncratic volatility of structural shocks. More precisely, volatility has to alternate faster than connectedness. Intuitively, this assumption ensures that we can attribute all local time-variation to changes in idiosyncratic volatility. The additional covariance structure from the Taylor expansion doubles the number of equations such that it matches the number of unknowns. This way, we can identify structural parameters when the connectedness matrix alternates reasonably slower than the idiosyncratic volatility of the shocks. Unlike previous work, this local identification improves on the assumption of static connectedness by allowing for dynamic parameters. Finally, in the network application, we introduce a new intuitive centrality measure for financial firms.

The estimation of return volatility spillovers offers new insights into the average connectedness of US financial firms. The results suggest that the most significant peaks in spillovers occurred in the financial crisis. However, we find that spillovers also peak at times where average return volatility does not. Moreover, we see a higher spillover from stocks of financial firms to the stock market index than vice versa. This finding is in line with the observation that financial crises are almost always more severe than general economic crises. A snapshot analysis reveals that all firms but Goldman Sachs receive spillovers from AIG around the time of the bailout. Finally, we highlight the systemic relevance of institutions over the last two decades. In particular, we find that JPMorgan Chase shows a significant increase in relevance since the official Brexit. As of June 2018, it is the most systemic relevant financial institution. This result suggests that policymakers and regulators are well advised to monitor JPMorgan Chase’s financial health carefully.

Our study is related to three strands of the literature: the empirical studies of connectedness, the identification of structural shocks and the estimation of time-varying coefficients.

While the earlier literature measures dependencies as pairwise associations, e.g., [Adrian and Brunnermeier \(2011\)](#) and [Brownlees and Engle \(2012\)](#), the whole concept of connectedness as a network has first been addressed by [Diebold and Yilmaz \(2014\)](#). Their main claim is the interpretation of Forecast Error Variance Decompositions as networks. However, this framework only considers contemporaneous relations as undirected correlations. Then, if most significant reactions occur con-

temporaneously, this method is prone to misspecifications since it only estimates slower reactions. Thus, dynamic propagation is less credible for most applications. For example, in a volatility context, investors are more alerted in turbulent periods and, thus, it is conceivable that a substantial part of the responses occurs within the same day, i.e., contemporaneously.

Another approach to estimating directed dependencies is Granger-Causality testing. For example, [Billio et al. \(2012\)](#) provide a binary network with entries based on positive Granger-Causality tests. However, these particular tests are carried out for multiple lags and, hence, also ignore contemporaneous directionality. In contrast, [Barigozzi and Brownlees \(2013\)](#) tackle directional dependencies with equation-wise LASSO-type techniques to estimate a sparse causal contemporaneous network. Yet, a precise non-sparse contemporaneous causal network remains imperfectly estimated at best. [De Santis and Zimic \(2017\)](#) directly estimate a network by employing a structural VAR, which is, however, only set-identified. In a nutshell, empirical applications suffer under restrictive assumptions on the contemporaneous relations.

Note that the question of contemporaneous relations is equivalent to the identification of structural shocks. Such shocks are generally interpreted as unexpected uncorrelated exogenous innovations with economic interpretation. As aforementioned, structural shocks always require identification restrictions. The earlier literature mostly tackled this issue with the exclusion of effects, e.g., via triangularization as in [Sims \(1980\)](#) or via long-run restrictions as in [Blanchard and Quah \(1988\)](#). Equality constraints, however, are too restrictive for many economic applications. Therefore, set-identification of structural shocks by inequality restrictions as in [Uhlig \(2005\)](#) are popular in applications.¹

A smaller strand of the literature finds that changes in the variances of shocks can identify parameters of the contemporaneous connectedness matrix (see [Rigobon and Sack, 2003](#)). This relatively mild assumption generated attention as the presence of heteroskedasticity in the form of time-varying volatility is uncontroversial in most applications. In particular, they assume discrete volatility regimes which need to be determined. Prominent extensions as Markov-switching processes by [Lanne et al. \(2010\)](#) and GARCH by [Milunovich and Yang \(2013\)](#) model heteroskedasticity parametrically absent of external information. However, for most applications, the

¹[Fry and Pagan \(2011\)](#) point out the advantages and disadvantages of this method and highlighted explicitly that the estimations are not interpretable with probabilistic language.

implications of Markov-switching and GARCH models are too restrictive due to the functional form assumptions. Thus, [Lewis \(2017\)](#) uses non-parametric heteroskedasticity to identify a finite set of possible solutions. In contrast, the non-parametric approach in our paper point-identifies structural parameters and, moreover, allows for time-variation in the response matrix.

In the time-variation literature, a critical challenge is the derivation of restrictions to ensure positive definiteness. For example, [Primiceri \(2005\)](#) use a triangular decomposition and impose a prior on the evolutionary process. However, such priors are prone to misspecification for two reasons. First, the identification assumptions for the decompositions are just a loose approximation of the truth and hence lead to estimation uncertainty (see [Bognanni, 2018](#)). Second, distributional assumptions on the dynamics dissent with the econometrician’s direct interest of unveiling the evolution of the coefficients itself. To address this issue in the context of auto-regressive models, [Dahlhaus et al. \(1997\)](#) provide a prior-free estimation procedure motivated by the concept of infill asymptotics. In line with this idea, our prior-free estimation of the structural components may be adapted to Bayesian estimation since it is almost always concordant with less restrictive sampling methods.

The remainder of the paper is organized as follows. Section 1.2 first sets up the mathematical framework and briefly summarizes the structural identification problem. Within this framework, the same sections states identifying assumptions in 1.2.3 and offers a likelihood-based extremum estimator in 1.2.4. In Section 1.3, we apply this approach to provide further insights into contemporaneous causal networks of financial firms in the US. Section 1.4 concludes.

1.2 The Model

1.2.1 Terminology and General Setup

In the application, we consider a structural VAR model, but more broadly, we first introduce an N dimensional vector process u_t with contemporaneous interdependence. u_t can either be directly observed or obtained from a reduced form model. For example, for a time series y_t we have $A(L)y_t = u_t$, where $A(L)$ is a matrix function of the lag operator L . Contemporaneous interdependence is described by

$$A_{0,t}u_t = B_t\epsilon_t, \quad \forall t = 1, \dots, T, \quad (1.1)$$

where the structural matrix $A_{0,t}$ is a real valued parameter matrix of size $(N \times N)$ with full rank and unit diagonals. B_t is diagonal matrix with real valued positive entries. The structural shocks ϵ_t have a multivariate distribution with mean zero and unit variance. The unit diagonal of $A_{0,t}$ and the diagonal structure of B_t ensure that all connections of the N variables are in the off-diagonal of $A_{0,t}$. Heteroskedasticity is now included by $B_t \epsilon_t$ and is without further restrictions unconditional. Moreover, since $A_{0,t}$ is also assumed to be unconditional, we can use multiple estimation techniques (e.g. General Methods of Moments, likelihood methods, or General Least Squares).

Note that the basic structural equation reads $u_t = S_t \epsilon_t$ with S_t as the structural matrix. Both problems can be regarded as equivalent since diagonality of B_t and the unit diagonal of $A_{0,t}$ ensure that there exists a unique decomposition of $S_t = A_{0,t}^{-1} B_t$. The decomposition into $A_{0,t}$ and B_t allows to impose different assumptions and is essential for heteroskedasticity identification. The restrictions we consider focus on the time evolution of matrix entries and are generally in line with the most recent literature on time-varying parameters.²

Time-Variation Assumptions. For all $t \in (1, \dots, T)$, Parameters $\theta_t = (A_{0,t}, B_t)$ are bounded random and/or deterministic processes independent of ϵ_t . They satisfy

(A0) smoothness: for $1 \leq k \leq t$ and $k \rightarrow \infty$: $\sup_{d: |d| \leq k} \|\theta_t - \theta_{t+d}\| = O_p(k/t)$

Further,

(A1) local-linear volatility: B_t has a time-derivative or time-trend different from zero

(A2) local-constant connectedness: $A_{0,t}$ has a slower alteration rate, i.e., for some $c > 0$: $\sup_{d: |d| \leq k} \|A_{0,t} - A_{0,t+d}\| = O_p(k/t^{1+c}) = o_p(k/t)$

The smoothness condition (A0) implies that parameters drift slowly with time. In particular, this condition enables consistent estimation and allows for local approximations. Moreover, (A1) and (A2) ensure identification of parameters since we assume different time-variation behavior. In fact, the difference in the convergence (alteration) rate is the identification assumption. By (A0), the heteroskedastic volatility parameter B_t is assumed to have an asymptotic derivative, which (A1) ensures to be different from zero. In contrast, (A2) states that $A_{0,t}$ changes slower

²see e.g. Giraitis et al. (2016).

such that its derivative or time-trend is negligible. In a nutshell, we expect volatility depicted by $B_t B_t'$ to evolve faster than connectedness in $A_{0,t}$.

Assumption (A0) is a generalization of the standard assumptions for locally stationary processes. Namely, in the work of [Dahlhaus et al. \(2006\)](#) the parameter $\theta_t = \theta(t/T)$ is assumed to be smooth, deterministic and piecewise differentiable. In contrast, (A0) also allows for stochastic parameter processes but ensures the necessary degree of persistence in the entries. Assumptions (A1) and (A2) are further specifications under the smoothness condition. While B_t is linear with a non-zero gradient for sufficiently small segments, $A_{0,t}$ is constant on the same segment. Intuitively, variations in B_t dominate variations in $A_{0,t}$ and thus we can neglect the latter. In sections 1.2.3 and 1.2.4, we make clear how this dominance comes into play.

Different from previous studies about heteroskedasticity identification, these conditions are more flexible since they allow $A_{0,t}$ to be time-varying. Whereas this approach also functions under static $A_{0,t}$, we are later forced to adopt time-variation due to the estimation procedure. Relaxation of the time-invariant assumption, however, tackles the most prominent critique of identification via heteroskedasticity, which states that time-variation in B_t is expected to accompany with time-variation in $A_{0,t}$.

1.2.2 Structural Identification in a Nutshell

To get further insights in the identification and estimation, we take the respective reduced form of (1.1),

$$u_t = A_{0,t}^{-1} B_t \epsilon_t, \quad E[u_t u_t'] = \Sigma_t = A_{0,t}^{-1} B_t B_t' A_{0,t}^{-1'}. \quad (1.2)$$

Consequently, we get u_t as a forecast error with unconditional covariance matrix Σ_t . Although, we can estimate Σ_t , we are not able to deduce the structural components from it. In fact, we need at least $(N - 1)N/2$ further relations to uniquely identify $A_{0,t}$ and B_t . This requirement immediately follows from the $(N^2 + N)/2$ equations provided by Σ_t and the N^2 unknowns in $(A_{0,t}, B_t)$.

In order to understand the challenges of structural identification, we follow [Rubio-Ramirez et al. \(2010\)](#), and introduce the concept of observational equivalent matrices. In short, two structural parameter points are observational equivalent if and

only if they yield the same reduced form distribution of u_t . In our setting, this holds true if they yield the same Σ_t . To visualize this phenomenon, we redefine the structural parameters such that

$$\begin{aligned} \Sigma_t &= S_t S'_t, \\ \text{with } S_t &= A_{0,t}^{-1} B_t. \end{aligned} \tag{1.3}$$

Consequently, any parameter set $\{\tilde{S}_t = S_t Q | Q Q' = I_N\}$ satisfies (1.3) and thus produces the same distribution. In the context of SVARs, we end up with the same observations. The equivalence follows directly from the orthogonality of Q and can be observed by plugging it into (1.3).

Note that whenever there exists more than one observational equivalent structural parameter, the structural model is not identified since they all yield the same distribution. Thus, the set of orthogonal matrices defines all possible solutions in an estimation. However, the process in (1.1) suggests that only one S_t is correct. The fact that we can only observe and directly estimate (1.3) limits identification and estimation of the real structural parameters in (1.1) to the set of observational equivalent matrices. To obtain local identification of the structural parameters (A_0, B_t) , we need to rule out all Q 's but $Q = I_N$ such that only the true $\tilde{S}_t = S_t$ remains.

In the literature overview, various identification schemes have already been pointed out. To sum up, conditions on the parameter space, such as exclusion and long-run restrictions, are sufficient for local identification. However, economically motivating these restrictions is usually a fruitless challenge. Therefore, we desire weaker conditions to fit more applications. Sign restrictions, for example, are easy to motivate, but only reduce the number of observational equivalent parameters to more than one \tilde{S}_t . In this case, we call the parameters to be partially or set identified.

Although sign restrictions are frequently applied, they fall short when it comes to the interpretation of the estimation results. [Fry and Pagan \(2011\)](#) point out that solutions to the estimation cannot be interpreted with probabilistic language anymore, due to the fact that all but one \tilde{S}_t are untrue.³ Moreover, in the context of time-variation, set identification prohibits to depict point dynamics and only allows

³In a first attempt of this paper, time-variation was used to evaluate the goodness of fit of a single \tilde{S}_t in the set of partially identified parameters, but it proved to be infeasible obtaining the complete set of observationally equivalent solutions.

for interval dynamics, which are, as pointed out, non-probabilistic. Therefore, this study focuses on point-identification by exploiting the mild assumptions (A1) and (A2) for time-varying parameters.

1.2.3 Identification via Non-parametric Heteroskedasticity Modeling

Since this approach intends to work under parsimonious conditions, identification and estimation of parameters have two conceptional tasks. First, identification of structural parameters should proceed absent of any equality restrictions and parametric assumptions, and second, time-varying parameter estimation should be prior-free. To overcome these challenges, we double use the local-constant and local-linear assumptions for both tasks. That is, identification works under the assumptions (A1) and (A2), and estimation requires (A0) and (A2). In so doing, we take the idea of [Rigobon and Sack \(2003\)](#), who point out that structural parameters are identifiable when models exhibit heteroskedasticity. More precisely, since B_t must have at least two distinct values in (1.1), it is time-varying.

In the past, time-varying B_t have been modeled similarly with different approaches. Two examples are the previously mentioned [Lanne et al. \(2010\)](#), who use a two-state Markov-Switching model, and [Milunovich and Yang \(2013\)](#), who estimate B_t with a parametric GARCH model. Both papers establish an identification scheme by parametrizing heteroskedasticity in their models. However, it comes with the cost of functional form assumptions, which makes estimations sensitive to their compliance. For example, we find structural GARCH estimation numerically unstable due to the possible dissents between GARCH estimation and structural identification. Therefore, this paper targets to avoid the parametrization of heteroskedasticity.

Identification

Non-parametric heteroskedasticity modeling challenges identification due to the missing parametric gains. To still gain additional parametric information, we exploit the assumption of asymptotic differentiability and, thereby, we can stay in a non-parametric environment. Precisely, this assumption allows us to apply a derivative process to model heteroskedasticity and fit a locally weighted kernel estimator. The

local optimization of an objective function with respect to the functional value and its derivative gives us additional knowledge about the drift of the parameters. Such drifts can be estimated and, therefore, be parametrized.

To benefit from infill asymptotics, we approximate continuity of time by rescaling the time domain $[1, \dots, T]$ to the unit interval. We replace $A_{0,t}$, B_t and Σ_t by $A_{0,t/T}$, $B(t/T)$ and $\Sigma(t/T)$. Note, that while we have a functional notation for $B(t/T)$ and $\Sigma(t/T)$, we leave time as subscript character for $A_{0,t/T}$ since the asymptotic derivative $\partial A_{0,t}/\partial t = 0$. In particular, we use a Taylor-type expansion around τ for a piecewise differentiable function $f(\cdot)$,

$$f(t/T) = f_\tau + (t/T - \tau)\dot{f}_\tau + 1/2(t/T - \tau)^2\ddot{f}_\tau + \dots$$

For clarity, we left out the functional notation for realisations. Precisely, we denote the functional value at τ with the subscript τ . The number of dots above f_τ denotes the respective derivative at τ , e.g. $\ddot{f}_\tau = (\partial^2 f / \partial t^2)(\tau)$. In order to keep assumptions as parsimonious as possible, we use a Taylor series of degree one, which is sufficient for exact identification. Fitting any other degree larger than one makes the assumptions over-identifying and, hence, testable.⁴ A higher degree is easily achievable by following the same steps as in this paper.

Before we start, it is worth spending a thought on the parametric target of the Taylor expansion. The main issue is its linear nature, which approximates time-variation as a sum of the functional value and its derivative. While summands can easily be torn apart for estimation methods such as General Method of Moments, this form makes it hard for objective functions which include the inverted argument. More precisely, likelihood-based estimations suffer since the covariance matrix appears in an inverted fashion. To understand this problem, take the innovation covariance matrix $\Sigma(t)$ and Taylor expand it around τ . It reads $\Sigma_\tau(t) \approx \Sigma_\tau + (\frac{t}{T} - \tau)\dot{\Sigma}_\tau$. Plugging this matrix function in the log-likelihood results in an inconvenient representation due to the inverse of $\Sigma_\tau(t)$ (see section 1.2.4).

Taylor expanding the inverse covariance matrix (also called the concentration or precision matrix) obtains a more elegant representation. [Barigozzi and Brownlees \(2013\)](#) point out the advantages of parametrizing the concentration matrix instead of the covariance matrix itself. In fact, the entries of the concentration matrix re-

⁴Note, that the test can detect if the assumptions are too restrictive but can not indicate which assumptions violate the observations.

late to the contemporaneous correlations between two variables conditional on others. Namely, partial correlations ρ_{ij} depend on entries of the concentration matrix, $\Sigma^{-1}(t)$, in the following way,

$$\rho_{ij} = -\frac{\Sigma^{-1}(t)_{ij}}{\sqrt{\Sigma^{-1}(t)_{ii}\Sigma^{-1}(t)_{jj}}}.$$

This matrix is still symmetric, but it already comes close to the matrix of causal dependencies. The entries in the covariance matrix, in contrast, are affected by many conditional correlations and thereby are a mix of many dependencies. Although, the dependency structure of this matrix still contains undirected connections,⁵ we see the Taylor expansion of the concentration matrix, $\Sigma^{-1}(t)$, superior. Moreover, targeting the concentration matrix allows for a more appealing objective function but does not affect estimation and identification.

The Taylor expansion for $\Sigma^{-1}(t/T) = A'_{0,t/T}B(t/T)^{-2}A_{0,t/T}$ around τ reads

$$\Sigma_{\tau}^{-1}(t/T) \approx \Sigma_{\tau}^{-1} + \left(\frac{t}{T} - \tau\right)\dot{\Sigma}_{\tau}^{-1}, \quad (1.4)$$

$$\Sigma_{\tau}^{-1} = A'_{0,\tau}B_{\tau}^{-2}A_{0,\tau} \quad (1.5)$$

$$\begin{aligned} \dot{\Sigma}_{\tau}^{-1} &= \frac{\partial(A'_{0,\tau}B(\tau)^{-2}A_{0,\tau})}{\partial\tau}(\tau) \\ &= \dot{A}'_{0,\tau}B_{\tau}^{-2}A_{0,\tau} - 2A'_{0,\tau}B_{\tau}^{-3}\dot{B}_{\tau}A_{0,\tau} + A'_{0,\tau}B_{\tau}^{-2}\dot{A}_{0,\tau}, \end{aligned} \quad (1.6)$$

where the last equation follows from the chain rule and the diagonality of B_{τ} . To build intuition for the derivative of a function under the assumptions (A0) and (A2), we illustrate the (infill) asymptotics of the time-varying process. For two arbitrary points τ and $\tau + t/T$, the functional difference in the limit becomes the derivative,

$$\lim_{T \rightarrow \infty} \frac{\theta(\tau + t/T) - \theta(\tau)}{t/T} = \frac{\partial\theta}{\partial\tau}(\tau).$$

By assumption (A0), this limit exists and therefore the derivative is asymptotically defined. Additionally, (A2) ensures that the derivative of $A_{0,\tau}$ goes to zero with convergence rate T . For the sake of clarity, derivatives of $A_{0,t}$ will be dropped since

⁵Take, for example, the case with B_t as the identity. The direct dependencies in $A_{0,t}$ stay hidden due to the symmetry of $\Sigma^{-1}(t) = A'_{0,t}A_{0,t}$

assumption (A2) ensures asymptotical negligibility. (1.6) becomes

$$\dot{\Sigma}_\tau^{-1} = -2A'_{0,\tau}B_\tau^{-3}\dot{B}_\tau A_{0,\tau}. \quad (1.7)$$

Intuitively, the assumptions allow to attribute all variation in Σ_τ to variations in B_τ , and, hence, provides additional information about $A_{0,t}$. The mapping from the structural parameters to the reduced form ones is consequently given by (1.5) and (1.7), where there are $N(N+1)$ reduced form parameters in $(\Sigma_\tau^{-1}, \dot{\Sigma}_\tau^{-1})$ and $N(N+1)$ structural parameters in $(A_{0,\tau}, B_\tau, \dot{B}_\tau)$.

In the Appendix, Proposition 1 shows the conditions for identification of the mappings. However, as B_τ^{-2} is squared, B_τ 's and \dot{B}_τ 's identifications are subject to sign changes. We can easily solve this issue by restricting the elements to be positive, which in turn imposes sign restrictions on \dot{B}_τ . Fortunately, from construction, we never sought to identify the derivative. It just serves as a tool to find $A_{0,\tau}$ and B_t and does not hold any interpretational value. In summary, identification results from the mappings in (1.5) and (1.7) and, therefore, estimation of the structural parameters is possible if we can find estimates for Σ_τ^{-1} and $\dot{\Sigma}_\tau^{-1}$. Note that, in contrast to previous approaches, identification works for one observation τ and thus is independent of others. Not only this peculiarity allows for time-variation in both parameters, but also, it leads to a local estimation function, which is well tractable.

The condition $\dot{b}_i/b_i \neq \dot{b}_j/b_j$ for all $i \neq j$ assumes that the relative derivatives of the structural variances are never the same for two variables. Considering that it is highly unlikely that relative marginal changes of structural variances match the same value, we come to the conclusion that this assumption holds for heteroskedasticity applications with (A1) and (A2) fulfilled. Nevertheless, we are aware of the potentially increasing identification uncertainty in case two relative derivatives are close in magnitude. In applications where a common factor drives the structural variances, violation of this condition represents a more serious problem. For example, the market's mood mostly drives heteroskedasticity for stock volatility, and it is conceivable that relative changes in orthogonal variances are subject to changes in mood. In this case, since we expect a common factor in the mood, the model should equate for it.

1.2.4 Extremum Estimator

In this section, we propose an estimator for the local-constant local-linear assumptions. Recall that, in contrast to previous heteroskedasticity literature, we only require parameters at one observation for identification. This idiosyncrasy allows us to estimate the structural parameters point-wise. In so doing, we tackle a weakness of previous papers, which need to assume that $A_{0,t}$ is constant over multiple periods. In our setup, this assumption leads to over-identification. We could test over-identified models, but it is highly likely that we reject the null hypothesis. Therefore, we focus on the estimation of time-varying $A_{0,t}$.

Let $l(u_t|\theta_\tau)$ be the likelihood of the reduced form vector u_t to occur under the parameters of θ_τ . With slight abuse of notation, θ_τ now consists of the structural parameters $(A_{0,\tau}, B_\tau)$ and the derivative \dot{B}_τ .

By plugging in the structural parameters in the log-likelihood for normally distributed errors we get

$$\begin{aligned} l(u_t|\theta_\tau) &= -\frac{1}{2}\ln|2\pi| - \frac{1}{2}\ln|\Sigma_\tau(\frac{t}{T})| - \frac{1}{2}u_t'\Sigma_\tau^{-1}(\frac{t}{T})u_t, \\ &= -\frac{1}{2}\ln|2\pi| + \ln|A_{0,\tau}| + \frac{1}{2}\ln|B_\tau^{-2}(I_N - 2(\frac{t}{T} - \tau)\dot{B}_\tau B_\tau^{-1})| \\ &\quad - \frac{1}{2}u_t' A_{0,\tau}' B_\tau^{-2} A_{0,\tau} u_t - \frac{1}{2}u_t' A_{0,\tau}' (-2)(\frac{t}{T} - \tau)\dot{B}_\tau B_\tau^{-3} A_{0,\tau} u_t, \end{aligned}$$

with $|\cdot|$ denoting the matrix determinant. Note that, we take the log-likelihood of any time point t with parameters at τ . This step is necessary for local estimation techniques and requires us to have approximations for $t \neq \tau$. In particular, we use (1.4) to get an idea of other time points. Although, the setup implies unconditional covariance matrices, we plug in the conditional moments in the likelihood. Implications of this necessity are pointed out at the end of this section.

Due to the nature of Taylor approximations and local smoothing estimators, we choose an extremum estimator as in [Giraitis et al. \(2016\)](#). This estimator is similar to [Fan et al. \(1995\)](#)'s quasi maximum likelihood (QML) estimator. It is well known, that under correct specifications and identification of parameters the QML with normal distribution is consistent. In the time series context, QML is consistent and asymptotically normal under regularity conditions (see [Bollerslev and Wooldridge \(1992\)](#)). In contrast to QML, the extremum estimator estimates locally due to different weights for log-likelihood realizations. Precisely, it weights residual

log-likelihoods with a pre-specified kernel $K_h(x) = K(x/h)/h$ and bandwidth h . The kernel is a symmetric continuous bounded function with compact support and normalized to one. Moreover, the bandwidth h satisfies $h \rightarrow \infty$ and $h = o(T^{1/2})$. The log-likelihood at t has weight $K_h(t/T - \tau)$ for the estimate at τ such that the extremum estimator reads

$$L_\tau(\theta_\tau) = \sum_{t=1}^T K_h\left(\frac{t}{T} - \tau\right) l(u_t | \theta_\tau). \quad (1.8)$$

We reformulate (1.8) with the properties of the Kronecker product, determinant and inner product, and obtain

$$\begin{aligned} L_\tau(\theta_\tau) = & c + \ln|A_{0,\tau}| - \ln|B_\tau| + \sum_{t=1}^T K_h\left(\frac{t}{T} - \tau\right) \frac{1}{2} \ln|I_N - 2\left(\frac{t}{T} - \tau\right) \dot{B}_\tau B_\tau^{-1}| \\ & - \frac{1}{2} \text{trace}(\tilde{\Sigma}_\tau A'_{0,\tau} B_\tau^{-2} A_{0,\tau}) + \text{trace}(\tilde{\Sigma}_\tau A'_{0,\tau} \dot{B}_\tau B_\tau^{-3} A_{0,\tau}), \end{aligned} \quad (1.9)$$

with

$$\begin{aligned} \tilde{\Sigma}_\tau &= U W U', \quad W = \text{diag}(K_h(\frac{1}{T} - \tau), \dots, K_h(\frac{T}{T} - \tau)), \\ \dot{\tilde{\Sigma}}_\tau &= U D W U', \quad D = \text{diag}((\frac{1}{T} - \tau), \dots, (\frac{T}{T} - \tau)). \end{aligned}$$

where c is a constant term encompassing $-\frac{1}{2} \ln|2\pi|$ and $U = [u_1, \dots, u_T]$ is the matrix of realizations of u_t . Note that $\tilde{\Sigma}_\tau$ and $\dot{\tilde{\Sigma}}_\tau$ represent the local (least square) estimates with a kernel weighting. Finally, optimizing (1.9) with respect to $(A_{0,\tau}, B_\tau, \dot{B}_\tau)$ obtains the estimates for the structural parameters.

To build intuition for the functionality of the estimation, we inspect its two approximation errors. First, the Taylor expansion of $\Sigma_\tau(t/T)$ around τ yields an error at observations different from τ . In the local polynomial estimation, this error is asymptotically negligible by the smoothness assumption (A0). Second, the mapping of $\theta_\tau = (A_{0,\tau}, B_\tau, \dot{B}_\tau)$ to $\dot{\Sigma}_\tau$ ignores the trend/derivative of $A_{0,\tau}$. Clearly, accounting for this term makes computations numerically expensive and requires priors. In the time-varying setup, assumption (A2) depicts a mild prior and ensures that this approximation is asymptotically negligible as well.⁶ Intuitively, it is sufficient that

⁶In fact, the term $\sum_{t=1}^T K_h(\frac{t}{T} - \tau) (l(u_t | \theta_t) - l(u_t | \theta_\tau))$ is asymptotically negligible due to similar arguments as in [Giraitis et al. \(2016\)](#). Clearly, the kernel automatically gives less weight to τ^* distant from τ and comprises the approximation errors for Σ_{τ^*} . Note that by construction Σ_t also fulfills

$A_{0,\tau}$ is locally constant over intervals where B_τ is not such that all variations in Σ_τ are evoked by variations in B_τ . This condition is ensured by $A_{0,\tau}$'s slower alteration rate.

1.3 Causal Financial Connectedness

1.3.1 A causal network model

In empirical applications, structural models are prominent in macroeconomics and financial econometrics. While macro data such as unemployment, GDP, and inflation are mostly available at low frequencies, financial data such as return and volatility are rich in, both, frequency and variables. This abundance allows for a variety of applications but also challenges us to deal with its peculiarities. In particular, return, and volatility parameters are mostly considered to be time-varying due to different market sentiments over time.

We exploit time-variation of the coefficients and innovation covariance matrix by aligning our model within the setup of [Diebold and Yilmaz \(2014\)](#)'s daily stock return volatilities. In their application, they estimate rolling window FEVDs and interpret them as financial networks. As aforementioned, generalized FEVDs fall short when it comes to accounting for contemporaneous dependencies. In particular, in the setup of daily volatilities, it sums up all within-day dependencies and models them as undirected correlations. However, it is highly conceivable that most of the significant reactions on strong adverse shocks occur on the same day. The generalized FEVD ignores the direction of these reactions. Therefore, we extend [Diebold and Yilmaz \(2014\)](#)'s approach by a contemporaneous directed network to provide new insights into the contagion process.

We model a system of N financial institutions and assume it to contain the most important firms, such that there are no other significant effects. Then, business between the firms in form of contractual obligations creates financial dependencies. The set of financial dependencies spans a structure, which we want to estimate to deduce systemic issues. More precisely, the dependencies in this structure can present various key figures. For example, we can see how gains, losses, and risk of firms are dependent on the financial success of others. For a given horizon, we

(A0)

interpret this structure as a network and analyze it with tools from the literature.

Various response times help us to understand different kinds of dependencies and, therefore, we define networks for various time horizons. In particular, networks appear in three forms in our analysis. First, the causal network G_t depicts immediate reactions/direct dependencies. For daily observations, it is loose but perhaps helpful to see this matrix as a within-seconds reaction type. Second, $A_{0,t}^{-1}$ is the impact network and also appears contemporaneously, but quantifies the contagion result at the one-step forecast errors u_t . Loosely speaking, this matrix shows the end-of-the-day outcome of chain reactions in case of daily observations. Third, a spillover network shows the dynamic propagation of shocks after H periods. More precisely, this matrix is employed as a Forecast Variance Decomposition, which we introduce in Section 1.3.2. To account for propagation, we include dependency over periods via lags in the regression.

Analogously to Diebold and Yilmaz (2014), we employ a VAR(3) model with daily observation of volatility. The structural model reads

$$y_t = \alpha_t + G_t y_t + A_{1,t} y_{t-1} + A_{2,t} y_{t-2} + A_{3,t} y_{t-3} + B_t \epsilon_t, \quad \forall t = 4, \dots, T \quad (1.10)$$

where y_t is an N dimensional time series vector of observables and G_t is an $(N \times N)$ adjacency matrix with zero on the diagonal. The network G_t contains nodes and links representing firms and dependencies respectively. We assume it to be directed and weighted, i.e. G_t is non-symmetric and has non-negative values.⁷ $A_{1:3,t}$ are real valued time-varying autoregressive matrices within the boundaries of stationarity. B_t is a $(N \times N)$ diagonal matrix representing the idiosyncratic risks and ϵ_t is the normalized structural shock vector. α_t is a vector of intercepts. Note that the infill asymptotics require a rescaled time domain. In the application however, we stick to the notation of discrete time for a more convenient representation.

To build intuition how non-negative entries in G_t ensure a network decomposition, we inspect $A_{0,t}$ in (1.1) more carefully and link it to (1.10). Since $A_{0,t}$ has ones on the diagonal by assumption, we decompose the structural equation in an additive

⁷Note that the assumption of non-negative values for G_t is not necessary for identification, but helps us to ensure a network decomposition. In the context of volatility spillovers, however, this assumption is plausible and helps to make the numerical optimization computationally more stable. Moreover, test for over-identification restrictions did not reject the null.

structural equation

$$\begin{aligned} A_{0,t}u_t &= B_t\epsilon_t, \\ u_t &= G_t u_t + B_t\epsilon_t, \\ G_t &= (I_N - A_{0,t}). \end{aligned}$$

Clearly, the existence of Σ_t implies that $A_{0,t}$ is invertable. Hence, the geometric series $(I_N - G_t)^{-1} = \sum_{k=0}^{\infty} G_t^k$ exists and is finite. In return, the maximum magnitude of G_t 's eigenvalues, i.e. the spectral radius $\rho(G_t)$, has to be smaller than 1. With all entries being non-negative, G_t has only values in $[0, 1)$. Then, G_t is, in fact, an adjacency matrix. In the literature of financial connectedness, this condition is known as a magnitude restriction. For example, [De Santis and Zimic \(2017\)](#) condition the effect to other shocks being smaller than the idiosyncratic effect.

1.3.2 Forecast Error Variance Decompositions for SVARs

For insights into the propagation of shocks, we analyze Variance Decompositions for a given time horizon. Forecast Error Variance Decompositions (FEVD) are essential for connectedness analyses since they predict how much of the forecast's variance is explained by other variables. They, therefore, show the estimated contributions to other variables. The resulting matrix shows the dependency structure for the forecast horizon and depicts a network matrix in our analysis.

Unfortunately, the estimation of the first impulse responses still poses a crucial problem for higher horizons. The respective FEVD, therefore, suffers under the decomposition of the covariance matrix. To address this issue, Generalized Forecast Error Variance Decompositions (GVD) from [Pesaran and Shin \(1998\)](#) use a correlation-based approach to get an idea of the first forecast error. The main advantage over the (orthogonal) triangular Cholesky decomposition is that it is invariant to reordering. However, it misses out on the directions of contemporaneous dependencies due to the pairwise nature of correlations.

In contrast, the structural VAR implies a decomposition of the covariance matrix. More precisely, the structural representation allows to label structural shocks, and thereby, we can see structural shocks as they are emanating from the respective source.

We start with the $MA(\infty)$ representation for forecast errors u_t

$$y_t = \sum_{k=0}^{\infty} \Phi_{k,t} u_{t-k}, \quad \Phi_{0,t} = I_N, \quad \forall t = -p+1, \dots, T, \quad (1.11)$$

where we can use $u_t = A_{0,t}^{-1} B_t \epsilon_t$ such that (1.11) becomes

$$y_t = \sum_{k=0}^{\infty} \Theta_{k,t} \epsilon_{t-k}, \quad \Theta_{k,t} = \Phi_{k,t} A_{0,t-k}^{-1} B_{t-k}. \quad (1.12)$$

Then matrix $\Theta_{k,t}$ contains response functions for horizon k . We observe impulse responses of a unit shock on variable j in the respective column of matrix

$$[\theta_{ij,k,t}] = \Phi_{k,t} A_{0,t-k}^{-1} B_{t-k}.$$

Since ϵ_t has mean zero and unit variance, squaring the elements of $\Theta_{k,t}$ gives us the error variance for the forecast at horizon k . Summing these error variances up from 0 to $H-1$, gives us the H -step forecast error variances:

$$\Psi_t^2(H) = \sum_{k=0}^{H-1} ((\Phi_{k,t} A_{0,t-k}^{-1} B_{t-k})^2). \quad (1.13)$$

where $(\cdot)^2$ denotes the element-wise squared matrix. The FEVD-table $D_t^H = [d_{ij,t}^H]$ reads

$$d_{ij,t}^H = \frac{\psi_{ij,t}^2(H)}{\sum_g \psi_{ig,t}^2(H)}, \quad (1.14)$$

where $\psi_{ij,t}^2(H)$ denotes the ij -th entry of $\Psi_t^2(H)$. The ratio explains j -th percentage contribution on the total forecast variance of all variables on i . In a financial setting, we can directly link this ratio to the expected capital shortfall of i conditional on j 's shortfall. Note that, in contrast to the generalized version of [Pesaran and Shin \(1998\)](#), the rows of D_t^H sum up to one.

1.3.3 Measures of Interconnectedness

Visualization of the results proves to be difficult for this analysis since we end up with almost T network matrices. In order to provide useful insights into systemic issues of the network, we define measures of interconnectedness in this section. The measures will apply to all three network matrices: the *causal network* matrix G_t ,

the contemporaneous *impact* matrix $A_{0,t}^{-1}B_t$ and the H -step *Forecast Error Variance Decomposition* D_t^H outlined in 1.3.2.

In analogy to [Diebold and Yilmaz \(2014\)](#), we add "From", "To" and "Average" effects to see which institutions receive or spread risk and how connectedness evolves. For connection matrix $M = [m_{ij}]$ we define

$$\begin{aligned} C_{i \leftarrow \cdot} (M) &= \sum_{j \neq i} m_{ij}, & (\text{From-Connectedness to } i) \\ C_{\cdot \leftarrow j} (M) &= \sum_{i \neq j} m_{ij}, & (\text{To-Connectedness from } j) \\ C (M) &= \frac{100}{N} \sum_i \sum_{j \neq i} m_{ij}. & (\text{Average Connectedness}) \end{aligned}$$

Note that the first two measures are on firm-basis and produce N numbers each, while the last distills the spillovers into a single number. From-Connectedness and To-Connectedness show how much effect a firm "takes from" or "gives to" others. These measures relate to the in-degree and out-degree respectively. The average over firms multiplied by 100 represents the Average Connectedness and provides information about the overall spillover potential. For a row-normalized connection matrix, a value of 100 depicts that an idiosyncratic shock only affects others. Similarly, a value of 50 shows that 50% of the shock's effect is on other firms. Finally, institutions are *Receivers*, *Distributors*, and *Diffusers* if they have a relatively high $C_{i \leftarrow \cdot}$, $C_{\cdot \leftarrow j}$, and both, respectively.

Although these measures provide insights into the importance of institutions, they neglect the underlying network architecture entirely. For instance, a Distributor is less relevant if all connected receiving institutions are not forwarding shocks. Thus, we should be concerned with measuring the centrality of firms within the network. For example, we can use the eigenvector centrality, which is arguably the most popular measure. It uses the attributes of eigenvalues and -vectors of the adjacency matrix and takes the eigenvector with the highest absolute eigenvalue as a centrality measure. However, this measure does not allow for an economic interpretation as such.

In contrast, we add a new model-derived centrality measure in the context of financial risk. First, we note that the impact matrix is a result of contagion through the causal network G_t . In one period, the reactions from the causal network G_t

mathematically occur infinitely often. Precisely, the contemporaneous reactions of y_t on a shock are depicted by the inverse structural matrix $A_{0,t}^{-1} = (I_N - G_t)^{-1}$. As aforementioned, the inverse is, in fact, the geometric sum $\sum_{k=0}^{\infty} G_t^k$. This breakdown shows that a minor change in G_t can change the contemporaneous contagion result $A_{0,t}^{-1}$ completely.

For risk mitigation, regulators are primarily interested in the contagion result. While a market intervention would come equal to a change in the causal network G_t , the impact on the system is the most essential result in the end. Therefore, we construct an artificial control experiment by treating one variable i as the placebo, i.e., all outgoing causal connections of i are set to zero. In algebra, the placebo variable has a zero column in the causal network G_t . The new causal and impact networks are denoted by $G_{t,-i}^*$ and $A_{0,t,-i}^{-1} B_t = (I - G_{t,-i}^*)^{-1} B_t$, respectively. The difference in the impact is defined as the Systemic Relevance of i ,

$$\Delta C_{i,t}^s = C((I_N - G_t)^{-1} B_t) - C((I_N - G_{t,-i}^*)^{-1} B_t). \quad (\text{Systemic Relevance of } i)$$

Note that this artificial control experiment is equivalent to a regulator making concessions about the safety of an institution's obligations. In particular, as the obligations from this institution are secured, other institutions do not take any more risk from it. Then, $\Delta C_{i,t}^s$ measures the reduction in the average risk exposure if i was a non-emanating variable. Consequently, the spillovers in the impact matrix decrease more if institution i is systemically more relevant, and hence higher values of this measure relate to a higher Relevance. The extension to longer time horizons works analogously.

1.3.4 Data

For an analysis of financial dependencies and spillovers, high-frequency balance sheets and other frequently updated obligations are ideal. However, the lack of such data forces us to estimate dependencies with publicly observable data. We use market-based data since it reflects the expectations of many strategically acting investors who might even have access to private firm information. As in most studies about spillovers, we study volatility to track investors fear. The connectedness of volatility represents the investor-anticipated dependencies of risk, gains, and losses of firms. However, we have to estimate volatility since it is latent. The so-called

realized volatility introduced by Andersen et al. (2003) has proven as a standard in estimating past volatility. Diebold and Yilmaz (2014) use daily log-realized volatility to track investors fear by averaging 5-minute return variances of high-frequency data. They argue that the result is a measure of risk and apply it successfully to spillovers.

Following Barigozzi and Brownlees (2013), we measure volatility with the extreme-value estimation by Parkinson (1980),

$$\tilde{\sigma}_{i,t}^2 = 0.361(p_{i,t}^{high} - p_{i,t}^{low})^2.$$

Where $p_{i,t}^{high/low}$ denotes the daily maximum and daily minimum of intra-day log-prices respectively. When it comes to applications, both volatility estimations show remarkably similar values. In fact, we found the same magnitude of both estimators and a correlation of more than 0.9 for the sample. The extreme value estimation is particularly appealing due to its simplicity and we believe it to be more robust to recording errors.⁸

More frequent observations help us to see reaction timings, but as for the estimation of volatility, daily data is the best we can do. To build intuition, we review investor reaction timings in the setup of daily observations. The main difference in reaction timings is between institutional investors and private investors.⁹ While private investors tend to be less informed and act slower, institutional investors have more information and react swiftly. Even though they are often rigid because of the size of their investments, they are still incentivized to respond on price signals as forcefully and quickly as possible. Therefore we expect the strongest and most sophisticated reaction type from institutional investors. Their swift reactions are most notable on the day of the occurrence of a shock. We, therefore, expect to see a relevant part of the market-based dependencies in the contemporaneous network.

To identify the contemporaneous network, assumptions (A1) and (A2) must be fulfilled.¹⁰ First, (A1) implies that the exogenous shock vector $B(t)\epsilon_t \sim (0, B(t)B(t)')$ has covariance matrix $B(t)B(t)'$ with a trend component. In particular, we expect shocks on realized volatility to have time-varying variance in the form of stable

⁸Since the extreme value estimation only requires daily high and low, we do not require high-frequency data anymore.

⁹Market movements caused by algorithmic tradings are ascribed to the institutional investor type reaction.

¹⁰(A0) follows from (A1) and (A2).

processes. It is conceivable that this cyclical component comes from an investor sentiment cycle and, thus, we see this assumption satisfied. Second, (A2) implies that entries in G_t evolve smoothly and have slower alterations than B_t . As argued above, we assume that this network is a result of dependencies caused by contractual obligations between financial firms. Stacks of contractual obligations constantly grow and shrink due to new contracts getting signed and old ones maturing. Since one contract is just a small increment of a stack, such stacks evolve at a slow rate and can be assumed to be constant for sufficiently small periods. Thus, we see market sentiments to have a faster variation rate than the whole stack of contracts.

The choice of institutions is crucial for the interpretation of the results. For example, consider the case of neglecting a significant source of shocks. Then, the statistical identification of orthogonal shocks does not provide an economically meaningful interpretation. Clearly, we want to include all institutions who have a potential effect, but numerical optimizations restrict us computationally. Thus, we focus only on the US financial market. For this market, we want to capture all main five categories: primary and non-primary dealer banks, non-bank financial institutions, non-financial institutions, and, to measure common effects, the rest.

We separate US banks into primary dealers and non-primary dealers since their status differs substantially in the auction of US government securities and the open-market operations of the Federal Reserve. In the US, a primary dealer is a financial institution which is permitted to trade with the Federal Reserve directly. Other financial institutions which are not classified as such, have to trade with primary dealers to fund themselves. Therefore, non-primary dealers only get funding via the interbank market. To account for all significant exogenous variations, we include as primary dealers the six largest banks in the US: JPMorgan Chase, Citigroup, Bank of America, Wells Fargo, Goldman Sachs, and Morgan Stanley.¹¹

On the counterpart, the non-primary dealers are assumed to take less than average risk spillovers since they lend less money to other banks. To see these connections, we add a frequently traded non-primary dealer. Bank of New York Mellon represents smaller banks and non-primary dealers. However, we expect these banks to have significant effects on other banks since they refinance solely via the interbank market.

¹¹Bank of America is not listed as a primary dealer. However, Merrill Lynch under the auspices of Bank of America is listed and, hence, we treat Bank of America as a primary dealer.

We also include two non-bank financial institutions, American Express, as a financial service institute and American International Group (AIG), which played a crucial role in the 2007 financial crisis. Moreover, we include Apple in the analysis as the representative for big non-financial institutions. It had the most substantial returns in our sample period and is the most valuable stock in the market as of today.¹²

As already argued, the statistical method does orthogonalize ϵ_t . However, we also want to underpin the shock to variables with some economic meaning. For that reason, we account for market-specific shocks by including an exchange-traded fund (ETF) that tracks a broad market index. An obvious choice is the S&P 500 Ex-Financials. Unfortunately, ETFs for this index do not date back far enough to complete our sample, and thus we are not able to overcome this issue.¹³ Thus, we include the classical SPDR S&P 500 trust (SPY). Common shocks, such as shocks to macroeconomic factors, are embedded in this variable. In doing so, we obtain orthogonality of shocks in a more meaningful way. If we assume that common shocks exclusively occur to the whole market, we can interpret all other shocks as economically meaningful. Nevertheless, there is, most likely, no helpful interpretation of shocks to the market since it includes various macroeconomic factors. Eventually, we can rule out price effects since the construction of volatility already comprises price shocks.

Our sample period starts at the 01/03/2000 and ends at the 06/30/2018 and, hence, includes two major crisis, which is the early 2000's crisis with the 9/11 crash in 2001 and the financial crisis with the Lehman Brothers default and AIG bailout in September 2008. The data source of all aforementioned is CRSP.

1.3.5 Estimation Strategy

The estimation includes two steps. First, we estimate the reduced form, and, second, we take the resulting residual series for the structural estimation. For the empirical application, the companion form

$$Y_t = A_t^* Y_{t-1} + U_t, \quad (1.15)$$

¹²Apple also serves as a sanity check of our approach since we can easily find economically motivated arguments for the direction of its dependencies.

¹³Since we need daily high and low prices, it becomes complicated to create indexes for daily high and low prices.

demonstrates usefulness. Where

$$Y_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}, A_t^* = \begin{bmatrix} A_{0,t}^{-1}\alpha & A_{0,t}^{-1}A_{1,t} & A_{0,t}^{-1}A_{2,t} & \cdots & A_{0,t}^{-1}A_{p,t} \\ 0 & I_N & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & I_N & 0 \end{bmatrix}, U_t = \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

In the first step, we perform a time-variation estimation of (1.15). Note that we explicitly need assumptions on the time-variation of $A_{i,t}$. We interpret them as increments contributing to the long-run dependency. In return, we impose the same assumptions on them as on $A_{0,t}$. The smoothness condition (A0) and the local-constant connectedness (A2) restrict us to use local-constant estimation techniques on the reduced form slope coefficients. However, for the same reasons as in the local extremum estimator, the approximation errors for the slope parameters are asymptotically negligible. We use kernel weighted least square (LS) estimation, i.e., a special form of the generalized least squares,

$$\hat{A}_t^* = XW_tZ'(ZW_tZ')^{-1} \quad (1.16)$$

where W_t is defined as in (1.9) and $X = [Y_1, \dots, Y_T]$, $Z = [Y_0, \dots, Y_{T-1}]$. The corresponding reduced form residuals are the first N values in $\hat{U}_t = Y_t - \hat{A}_t^*Y_{t-1}$. We use this series in (1.9) to obtain the structural parameters. Similar to the rolling window estimation, this estimation approach has the advantages of great simplicity and allows us to stay in a prior-free environment. Hence, in contrast to prior-based approaches, this estimation unveils a more meaningful evolution of connectedness.

A crucial step in the estimation concerns the selection of bandwidths for both steps: h_1 and h_2 . In the first estimation, we explored robustness for the choice of the bandwidth. The smoothness of the estimated coefficients was the only visible difference between the bandwidths. Due to this robustness and the fact that misspecification potentially increases estimation uncertainty in the second step, we apply a fully data-driven Cross-Validation approach for this bandwidth. In order to minimize estimation errors in the first step, we use the likelihood of the leave-one-out interpolation errors,

$$\hat{u}_{t,h_1}^\circ = y_t - \hat{y}_{t,h_1}^\circ, \quad (1.17)$$

where $\widehat{y}_{t,h_1}^\circ$ denotes the estimate of y_t with bandwidth h_1 and weight 0 on observation t . This step is equal to setting the t -th diagonal entry of W in (1.16) to 0. The bandwidth we consider is chosen with the following log-likelihood type leave-one-out criterion¹⁴

$$h_1^{CV} = \arg \max_h - \frac{1}{2} \sum_t \left(\widehat{u}_{t,h}^\circ{}' \widehat{\Sigma}_{\widehat{u}_{t,h}^\circ}^{-1} \widehat{u}_{t,h}^\circ + \log(|\widehat{\Sigma}_{\widehat{u}_{t,h}^\circ}|) \right), \quad (1.18)$$

where $\widehat{\Sigma}_{\widehat{u}_{t,h}^\circ}$ is the full-sample covariance estimate for the interpolation errors. Then, h_1^{CV} is the bandwidth which maximizes the explanatory power of the estimation given a training dataset. The training set is simply the bandwidth weighted sample without the value it predicts.

In the second step, we eliminate autocorrelation from the time series by using the residual series $\widehat{u}_t = \widehat{u}_{t,h_1^{CV}}$ with the previously selected bandwidth. Optimally, we also apply a data-driven bandwidth selection for the extremum estimation, but, unfortunately, we did not find any convincing criterion to choose the respective bandwidth. To the best of our knowledge, bandwidth selection for structural matrix estimates is still little studied. That is, an optimal bandwidth for the covariance matrix does not necessarily imply that it is also optimal for the structural parameters.

Thus, in this paper, we focus on a bandwidth which is meaningful for the needs of our story. Namely, since we want to account for the average maturity of contractual obligations of firms, we choose a bandwidth which assigns 95 % of the kernel's weight to one calendar year. For our data, this weight corresponds to 250 trading days. In particular,

$$h_2 = \frac{250}{2q_{0.95}T},$$

where T is the sample length and $q_{0.95}$ denotes the 95% quantile of the respective kernel distribution. We use a Gaussian kernel due to its good asymptotic behavior.

1.3.6 Empirical Results

In this section, we present the empirical results. Due to the issue of visualizing a time series of networks, we include two different types of analyses. First, we visualize

¹⁴The sum of squared interpolation errors is potentially inaccurate for a selection criterion because we expect the covariance matrix to be non-diagonal.

the whole network as a snapshot at a specific event, and the sample mean. Second, we make use of the measures introduced in section 1.3.3 to visualize the evolution of connectedness.

In the forthcoming analysis, we do not show results for Apple since we are not able to label shocks for this firm. In particular, Apple only serves as a representative of non-financial institutions, and we use it as a sanity check for the results. Table 1.1 shows descriptive statistics of the nine financial institutions. Note that JPMorgan Chase and Bank of America increased their market capitalization and their number of total assets by a large margin. In 2018, JPMorgan Chase had about 50 % more assets than the runner-up Bank of America. In general, all institutions increased their total assets, but note that assets generally increased in value since 2000Q1.

| Institution | Ticker | Market Cap. | | Total Assets | |
|---------------------|--------|-------------|--------|--------------|--------|
| | | 2000Q1 | 2018Q2 | 2000Q1 | 2018Q2 |
| American Int. Group | AIG | 168.8 | 47.2 | 279.3 | 496.8 |
| American Express | AXP | 66.2 | 84.4 | 150.7 | 184.9 |
| Bank of America | BAC | 86.9 | 282.3 | 656.1 | 2291.7 |
| Bank of NY Mellon | BK | 30.6 | 53.9 | 76.0 | 352.9 |
| Citigroup | C | 201.8 | 168.4 | 738.2 | 1912.3 |
| Goldman Sachs | GS | 44.8 | 82.9 | 276.9 | 968.6 |
| JP Morgan Chase | JPM | 71.9 | 350.2 | 391.5 | 3360.9 |
| Morgan Stanley | MS | 79.9 | 82.9 | 408.1 | 875.9 |
| Wells Fargo | WFC | 66.4 | 268.3 | 222.3 | 1879.7 |

Table 1.1: US. financial institutions key figures in bn. US\$

In Figure 1.1, we show the mean causal network over the whole sample with the 20 % strongest directional connections. The circles (nodes) represent the firms and are labeled with the respective ticker. The thickness of the arrows depicts the strongest 5%, 10%, 15%, and 20% percentile dependencies (in decreasing order of thickness). First, we note that the regular banks, J.P. Morgan, Wells Fargo, Citigroup, and Bank of America are well connected and seem to be completely intertwined. For the investment banks, Morgan Stanley and Goldman Sachs, we see a mutual connection, but just small associations to other institutions or the market. Further, the smaller non-primary dealer bank, Bank of New York Mellon, shows mainly outgoing connections. This finding is in line with the story of non-primary dealers being in primary dealers' debt. Bank of New York Mellon needs to finance itself via other banks and, thus, its financial health is essential for the lenders.

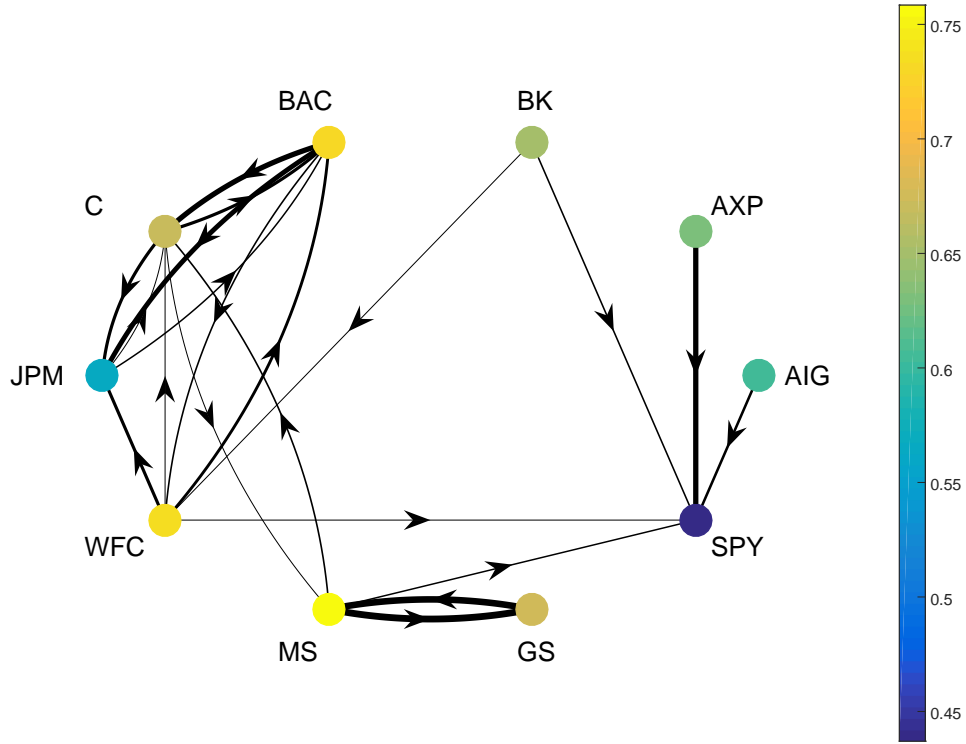


Figure 1.1: Mean Causal Network for the 20 % strongest connections. Node colors indicate the strength of the average To-Connectedness of the respective stock. The analysis includes American Int. Group (AIG), American Express (AXP), Bank of America (BAC), Bank of NY Mellon (BK), Citigroup (C), Goldman Sachs (GS), JP Morgan Chase (JPM), Morgan Stanley (MS), Wells Fargo (WFC).

Moreover, we see that American Express and AIG only connect to the market variable SPY and appear more isolated than other institutions. In particular, both affect the market but do not have a single incoming connection (in the highest 20%). While we ascribe SPY's dependency on American Express to the fact that we did not include other non-bank financial institutions, AIG clearly plays a central role as one of the biggest insurance companies in the US. Finally, we direct our attention to the role of the market within this constellation. The node color of SPY indicates that the outgoing effects are the weakest for all nodes, but we also observe that it has many ingoing effects. This perhaps surprising finding is in line with the claim that financial crises are more severe than general economic crises. We see in the graph that for this sample, there was more spillover from financial firms to the market than vice versa.

In an analysis of the evolution of dependencies, we are mainly interested in ex-

treme points, such as peaks and troughs, and a general trend. Therefore, to frame the evolution over time, we include a list of events, which are potentially important for the whole system. Table 1.2 gives an overview of these events.

| # | Date | Event Description |
|---|------------------------|--|
| 1 | 11 September '01 | Stock market crash due to the 09/11 terror attacks |
| 2 | 07 November '07 | AIG's Q3 report reveals losses with credit default swaps |
| 3 | March-May '08 | Bear Stearns bailout by the FED and take over by JPM |
| 4 | September-October '08 | Lehman Brother Default, AIG bailout, and Q3 reports |
| 5 | 06 May '10 | Flash Crash |
| 6 | September-November '11 | Down-Rating WFC, Q3 reports, Greece bailout |
| 7 | October '15 | Q3 reports |
| 8 | 29 March '17 | Great Britain invoked Article 50 starting the Brexit |
| 9 | 01 February '18 | China bans trading of Bitcoins |

Table 1.2: Important events concerning the overall connectedness of financial institutions

In the liquidity crisis of 2008, AIG, in particular, stood in focus when it became public that major banks and trading partners, such as Goldman Sachs, Morgan Stanley and Bank of America, depended on AIG's liquidity. Credit default swaps (CDSs) and collateralized debt obligations (CDOs) were mainly responsible for the tight entanglement of the firms. On the 14th of September, AIG sought for \$ 40 Billion in FED aid to survive the uprising liquidity crisis. It followed the default of Lehman Brothers on the 15th and the bailout of AIG on the 16th. For this sample, the 14th of September 2008 marks a key date since it moved investor attention towards AIG. Figure 1.2 shows the 10-day Forecast Error Variance Decompositions for this very date. In the network graph, the 20% strongest connections are plotted, and node colors represent outgoing effects of the respective stock. We compare the generalized version as in [Diebold and Yilmaz \(2009\)](#) on the left with the structural version as introduced in 1.3.2 on the right.¹⁵

For the generalized version, we see a higher connectedness of financial institutions in the node colors and mainly see connections for the traditional banks. This

¹⁵The generalized version has an implicit undirected contemporaneous connectedness assumption. Namely, for a forecast horizon of 1 day, we would not be able to see directions in the graph.

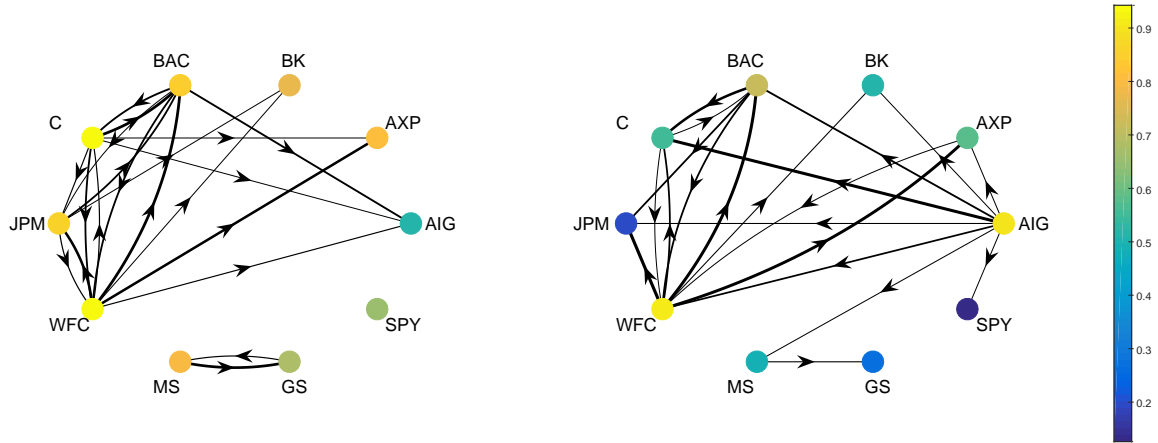


Figure 1.2: Comparison of Forecast Error Variance Decompositions with the generalized version as in Diebold and Yilmaz (2014) on the left and the structural version on the right. Observation at the 09/14/2008 for the 20 % strongest connections. Node colors indicate the strength of the average effect of a respective stock to others.

side-by-side comparison highlights the issue of using undirected correlations for the contemporaneous effect. In crises, investors are more alerted and react faster on shocks. Thus, most significant reactions occur within the same day. Since the generalized FEVD approximates contemporaneous connections with undirected correlations, it cannot uncover the full directed FEVD. We can also see that it connects all receiving institutions as they appear to co-move contemporaneously. For the correlation based decomposition, a firm-specific shock, as it happened to AIG, is not labeled and thus firms, which have high incoming and outgoing connections are over-estimated.

In contrast, AIG has mainly outgoing connections in this period, and thus its effect gets highly under-estimated in the generalized FEVD. For the structural FEVD, in comparison, AIG has outgoing connections to all other institutions but Goldman Sachs. This finding is in line with the investors' awareness of a potential default. Connections did not increase necessarily, but investors valued connections more due to the uncertainty emanating from AIG. Moreover, anticipating AIG's default, Goldman Sachs at that time protected itself against such an event and stood to collect 1.7 \$ bn from credit default swap had AIG defaulted.¹⁶ While investors anticipated

¹⁶see <https://www.bloomberg.com/news/articles/2010-07-24/citigroup-credit-suisse-top-list-of-goldman-sachs-counterparties-on-aig>

the shrinking dependency of Goldman Sachs to AIG, all other effects of AIG grew.

For further comparison, Figure 1.3 shows the dynamic of both, the generalized and the structural version. The Average Connectedness for the generalized version is about 30 index points higher, but movements are visually the same. That is, the generalized version overestimates spillovers. Recall, that the Generalized Forecast Error Variance Decomposition uses the undirected correlation for the one period forecast error. Due to the undirectedness, we witness two entries in the off-diagonal, even if there is only one in reality. This approximation generally results in an overestimation of connectedness.

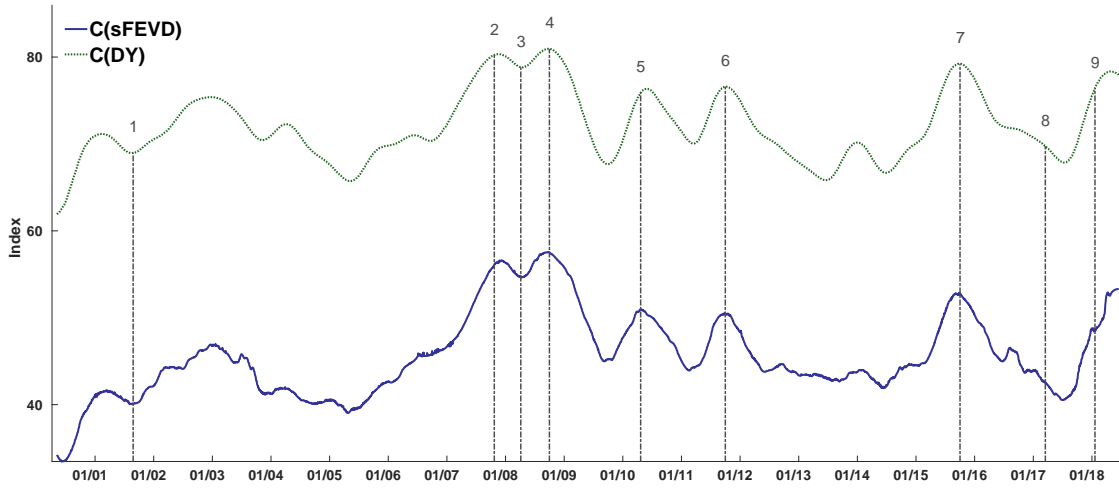


Figure 1.3: Total Average Connectedness for the FEVDs with structural decomposition (solid blue) and the generalized version of Diebold and Yilmaz (2009)(dotted green)

Forecast Error Variance Decompositions indicate general connectedness dynamics. In stressed investment periods, however, we expect to see most reactions contemporaneously since investors are alerted. Thus, we investigate contemporaneous relations further in an analysis of average connectedness for the causal matrix G_t and impact matrix $A_{0,t}^{-1}B_t$. Figure 1.4 shows both curves plotted on different scales. Namely, $C(G_t)$ ranges circa from 60 to 75, while $C(A_{0,t}^{-1}B_t)$ ranges from ca 80 to 145.

The impact's maximum value of 145 interprets as follows: A unity shock to i affects others on average with factor 1.45 in the same period. For the causal network, the maximum value of 75 states that a one standard deviation shock is immediately carried on with an average weight of 0.75. At first glance, the impact connectedness

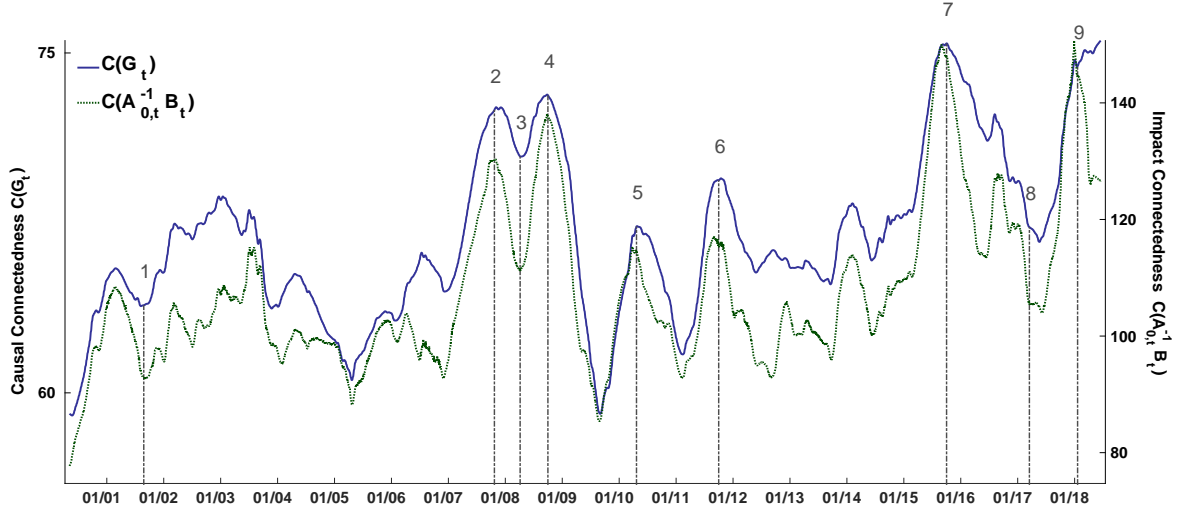


Figure 1.4: Total Average Connectedness for the causal network matrix G_t (solid blue), the impact matrix $A_{0,t}^{-1} B_t$ (dotted green)

looks like an enhanced version of the causal connectedness. Clearly, this visual similarity results from the relation $A_{0,t}^{-1} = \sum_{k=0}^{\infty} G_t^k$. However, for some periods, e.g., August 2006 to December 2006 and the first half of 2018, the two curves differ not just by magnitude but also by evolution. We ascribe these differences to changes in the network architecture. At the end of this section, we discuss the reasons for these differences with an analysis of the Systemic Relevances. In particular, this analysis points out the changes in the centralities of institutions for these periods.

Moreover, we analyze the peaks of connectedness. In particular, the most prominent peaks are seen during the financial crisis in events 2 and 4. These events are the third quarter financial report in 2007 of AIG and the chain reactions on the market caused by the Lehman Default, respectively. Whereas we would expect the Lehman Default to be the most severe event in the sample, the consecutive month was more stressful for the stock market. The third quarter report season in October 2008 revealed most of the losses and thereby symptomized the real connections of financial firms. More broadly, third quarter financial reports appear to have a significant effect on the connectedness perceived by the investors. For example, quarter three financial reports at event 6 and event 7 also date peaks of connectedness. Event 6 in 2011 was in the midst of the European sovereign debt crisis. The peak suggests that investors were surprised by the connectedness of banks in the US to European banks. Similarly, the quarter three reports in 2015 updated investors beliefs about the connectedness.

In contrast, a local trough is observable at event 3 between two peaks at the financial crisis. The successful take over of Bear Stearns cooled down some frayed nerves as it showed that an intervention of the FED buffers up the highest risks. In March 2017, when the UK government invoked Article 50 of the Treaty on European Union, we see another hard change in causal connectedness (event 8). Because the analysis does not include European Banks, the Brexit affects connections of US banks via the European ones. In particular, this event seemingly changed the architecture of the network but did not affect the Average Connectedness of the causal network.

Finally, we see a sharp increase in 2018 until the end of the sample. This increase visualizes the growing fear of adverse shocks to other institutions and awareness towards real connections. In particular, connectedness for the causal network matrix G_t is at an all-time high. In contrast, after event 9 the connectedness of the impact matrix $A_{0,t}^{-1}B_t$ shrank again. Note that, for risk management, the impact matrix plays a more important role than the causal network matrix. However, this difference indicates a change in the centrality of firms. Thus, for policymakers, it is key to identify the cause of it, namely, to find changes in institutions' relevance.

Excursion The 'May 6, 2010, Flash Crash' also known as the 'Crash of 2:45' (event 5) was a stock market crash which lasted for approximately 36 minutes. On this day, the stock market indices nearly lost 10% within minutes only to rebound half an hour later. Volatility measured exceptionally high and, perhaps surprising, connectedness as well. Since we can not identify any other events around this peak, the question arises whether the volatility level causes the level of Average Connectedness. In particular, more uncertainty in the stock market goes hand in hand with a higher average correlation between the stock returns. However, it is unclear whether a higher volatility level causes more connectedness or vice versa. Through the story, we expect that connections in the network form exogenously. That is, a higher degree of connectedness enhances the spillover of idiosyncratic shocks. By investigating the lead-lag correlation (see Appendix) of Average Connectedness and average volatility, we find support for exogenous network formation. This finding suggests that Average Connectedness is a predictor of risk since its variations lead variations in average volatility. Thus, this finding opens up new research questions.

To get more insights into single institutions, we study From-Connectedness, To-Connectedness, and Systemic Relevance. Figures 1.5 and 1.6 depict the From-

and To-Connectedness measures of the nine financial institutions and the market.

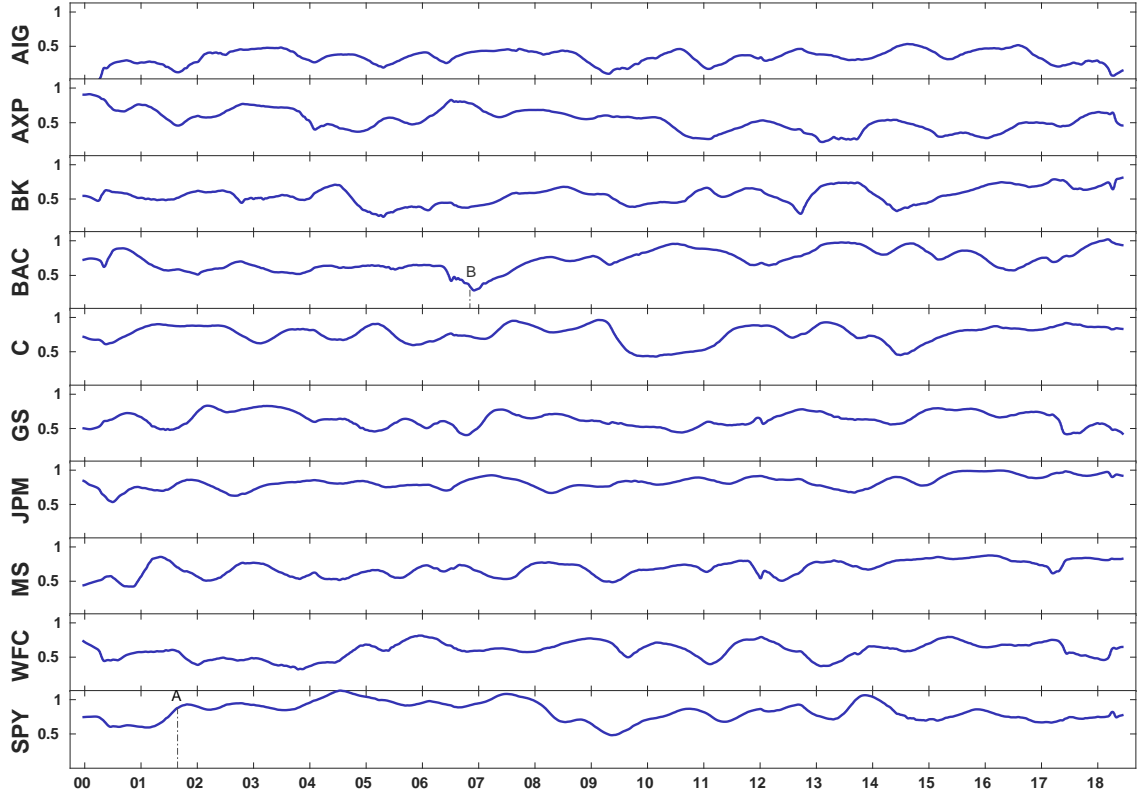


Figure 1.5: From-Connectedness of G_t

The most notable detail is the low To- and the high From-Connectedness of the market variable SPY. Since the market is substantially larger than the nine institutions combined, it seems surprising that the market has more incoming connections than outgoing ones. Nevertheless, it is common knowledge that financial institutions play a significant role in the expected mean and variance of general stock returns. Thus, we find that there is more spillover from the financial institutions to the market than vice versa. This finding is in line with [Jordà et al. \(2011\)](#), who point out that financial crises are more severe than 'normal' recessions.

Table 1.3 provides an overview of events which happened at specific peaks and troughs. At the 09/11 terror attacks (event A), American Express and the market variable (SPY) were the only ones reacting in connectedness. While both are dependent on private consumers' sentiments, other firms' business climate did not change significantly. These two kinks highlight that a shock to the market does not

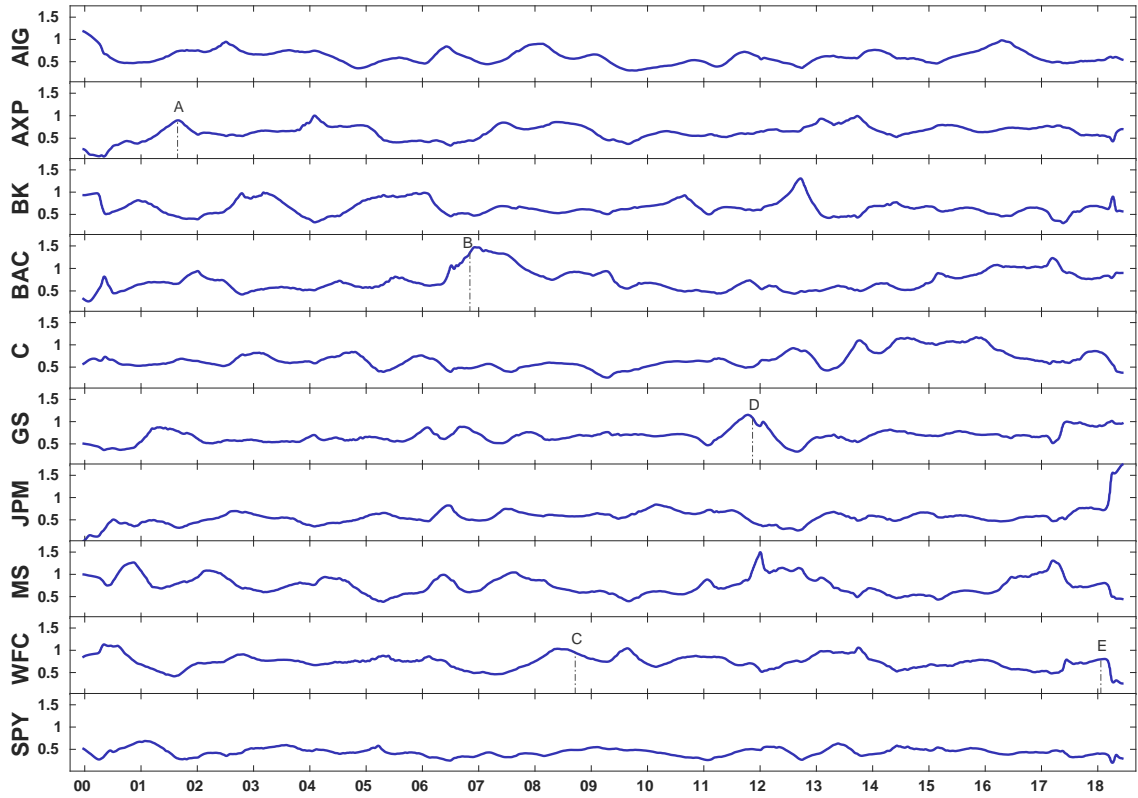


Figure 1.6: To-Connectedness of G_t

affect causal connectedness of financial institutions. We see this as a support for the structural decomposition of this paper.

Event B stated a trough in the From-Connectedness when Bank of America was about to acquire The United States Trust Company, which was famous for its private banking business. Bank of America seemingly was less dependent on other's financial status but also increased its outgoing connectedness (see To-Connectedness). Moreover, event D shows that the Greece bailout in 2011 affected Goldman Sachs. The To-Connectedness of Goldman Sachs rose from 0.5 to more than one as they stood to collect profits from a default of the crippled state. Finally, event E shows the FED prohibiting Wells Fargo from growing their asset base. This event led to a drop in the To-Connectedness.

As a final analysis, we show the Systemic Relevance of three institutions. This measure, which we introduced in 1.3.3, demonstrates the centrality of institutions. Recall that the measure is equivalent to a placebo experiment and intends to reveal

| | Date | Event Description |
|---|------------------|---|
| A | 11 September '01 | Stock market crash due to the 09/11 terror attacks |
| B | 20 November '06 | Bank of America announced to purchase of The United States Trust Company |
| C | 3 October '08 | Wachovia agrees to be bought by Wells Fargo |
| D | 27 November '11 | Writedown of 50% of Greek bonds |
| E | 2 February '18 | FED barred Wells Fargo from growing asset base |

Table 1.3: Important events on firm-level

the importance of the respective financial institution. That is, it measures the difference in the average connectedness of the impact matrix when institution i 's outgoing causal connections are mitigated. A higher value of this measure relates to a higher relevance for the system. The result is a combination of outgoing, ingoing connections and the general architecture of the system's network. We analyze the most noteworthy movements: Goldman Sachs, JPMorgan Chase, and Wells Fargo. Whereas they all increased market capitalization and total assets over the sample period, it remains unclear how their Systemic Relevance changed.

Figure 1.7 depicts the Systemic Relevance of Goldman Sachs, JPMorgan Chase, and Wells Fargo. While most of the pattern appears to be somewhat random, we can still see a slight trend for Goldman Sachs and a stronger trend for JPMorgan Chase. In particular, since Great Britain filed for the Brexit on the 29th of March 2017, JPMorgan Chase's Relevance experienced an almost strict upwards trend. More precisely, the average spillover of the whole system increases by 0.7 only due to JPMorgan Chase forwarding shocks. This development makes JPMorgan Chase the most important financial institution as of June 2018. Remarkably, this is the highest value of a single institution for the whole sample. As aforementioned, the formal process of Great Britain invoking Article 50 forms a milestone in the network architecture. Since then, European and British banks stay behind the US banks in financial profits. The sharp increase of JP Morgan Chase's Relevance hints that it benefits from a weaker competition overseas.

Moreover, Wells Fargo experienced a significant increase in importance in the financial crisis. On October 28th 2008, the bank took 25 \$ bn of Emergency Economic Stabilization Act funds. Wells Fargo's Systemic Relevance peaked at this date and decreased after the financial aid. More recent developments at the beginning of 2018 showed a sharp decrease in the importance of Wells Fargo. This decrease is

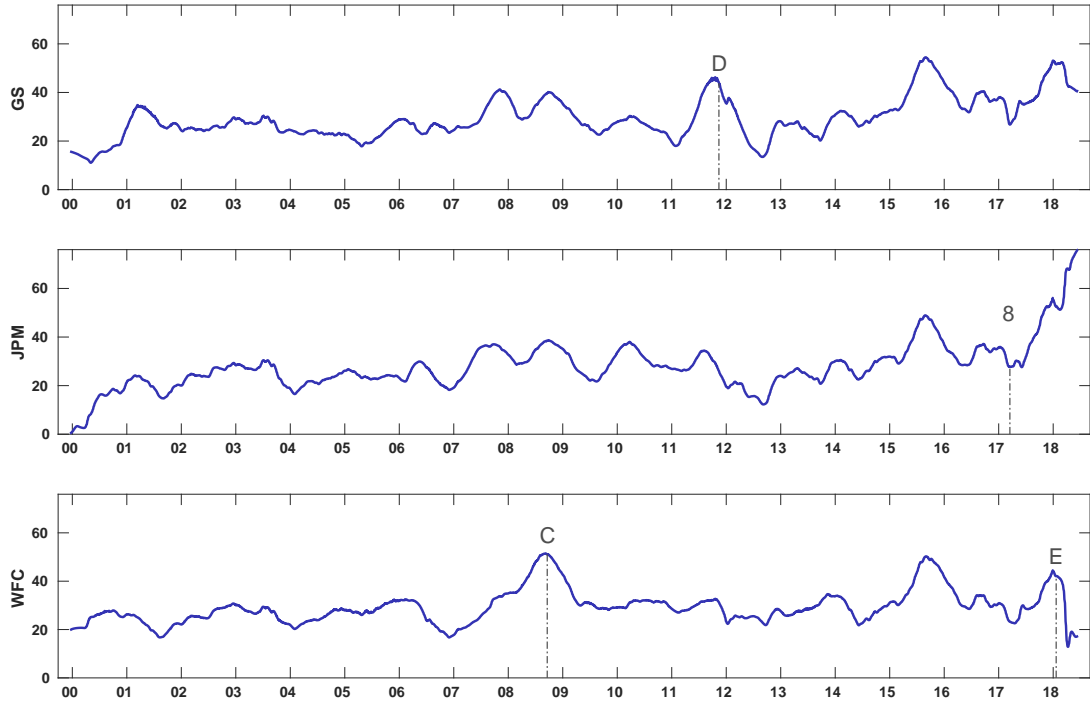


Figure 1.7: Systemic Relevance for particular firms

dated on February 2, 2018, when the FED barred Wells Fargo from growing its asset base. Up to the end of this sample, Wells Fargo was not able to solve their issues. Consequently, the Relevance of Wells Fargo is relatively low, as seen by investors. Based on the analysis of Systemic Relevance, regulators are well-advised to keep an eye on the financial health of JPMorgan Chase.

1.4 Conclusion

In this paper, we tackled the question of how to analyze directed contemporaneous connectedness. In the methodological section, we developed time-varying assumptions under which contemporaneous connections can be identified. More precisely, we modeled time-variation non-parametrically allowing to estimate a time-trend, which ensures identification. The identification works under parsimonious conditions and comes with a well tractable estimation function. Furthermore, we seized the problem of static response matrices by allowing all parameters to be time-varying.

In the application, we studied volatility spillovers for US financial firms in the framework of [Diebold and Yilmaz \(2014\)](#). Our approach highlighted the issue of approximating contemporaneous relations by correlations and, thus, it successfully addressed a major drawback of Generalized Forecast Error Variance Decompositions. The application included a contemporaneous causal dependency matrix, which provided new insights into the connectedness of financial firms. With an explicit contemporaneous network in hand, we introduced a new centrality measure that is economically meaningful. This centrality measure highlights the relevance of financial firms and serves as a critical tool for policymakers to monitor risk.

Empirically, we found that investors mostly react within a day in uncertain periods. This finding stressed the importance of using directed contemporaneous dependencies. Namely, we exemplified the directed connectedness of financial firms by comparing Generalized and structural Forecast Error Variance Decompositions. Moreover, we observed that the Average Connectedness measure is robust to stock return shocks. In line with the story, it appeared that the connectedness analysis based on market data peaked mostly for the financial reports seasons. Namely, financial reports in some periods might have differed from the expectations of the investors. In other words, public reports contained valuable information for investors since they hinted how strong the connections of financial firms are by revealing gains and losses.

Finally, we investigated the Systemic Relevance of financial firms. In this analysis, we focused on three banks, Goldman Sachs, JPMorgan Chase, and Wells Fargo. In particular, JPMorgan Chase stood out at the end of the sample. In June 2018, JPMorgan Chase was not only the biggest bank, but it also asserted its dominance by being the most relevant institution in our sample. With this finding, policymakers and regulators are well advised to monitor JPMorgan Chase’s financial health carefully.

Chapter 2

On Time-Variation of Financial Connectedness and its Statistical Significance

This paper is joint work with Carsten Jentsch.

2.1 Introduction

In 2008, the financial crisis peaked when Lehman Brothers filed for chapter 11 bankruptcy protection on the 15th of September. A day later, on the 16th of September, the FED took over 80% of the shares of American International Group (AIG). These events removed any remaining doubts about the systemic scale of the financial crisis. In order to understand why this crisis was systemically so spread, it is useful to see the underlying network structure. In particular, we want to know the full unified network. This unified network structure provides insights into a battery of essential risk figures, such as market risk, gridlock risk, and systemic risk. Because of insightful network graphs and network measures, the estimation of a unified dependency structure enjoys popularity in time series econometrics.

If nothing else is known, we usually estimate an unknown dependency structure with time series regressions. For example, a regression on the lags reveals an explicit, directed dependency. However, when it comes to network analysis, it is essential that the links have a meaningful economic interpretation. For financial firms, it is plausible to interpret dependencies as risk spillover resulting from stacks

of contractual obligations, such as asset and liability structures, derivatives, and mutual portfolio holdings. Since these stacks change over time, this interpretation implies that links, and consequently networks, are non-static. Moreover, to detect potential grievances, policy institutions are not only interested in the level of dependency but also its evolution. In a nutshell, it is advisable to use a time-varying parameter approach to capture financial networks accurately.

In general, many different strands of the literature incorporate time-variation to study lagged correlation structures, because the assumption of a static sample is often too strict. If not explicitly, authors assume implicitly that the sample's dependency structure is dynamic. However, time-variation in the parameters contradicts the assumption of ergodicity of a standard vector autoregression (VAR) and least squares estimation for rolling windows is not theoretically justified. Although the application's results are heuristically interpretable, they lack rigorous inference tools. An exhaustive analysis of financial dependencies requires confidence sets of the estimators. Without such, it is unclear whether a change in the estimators is rooted in estimation uncertainty or the dynamics of the actual parameter. Thus, we aim to fill this gap in the literature of time-varying parameters with an improved local-linear estimator with theoretical justification.

We face three main challenges. First, local-constant regressions, such as rolling windows, suffer from bias, and boundary issues. Local-linear estimations are known to eradicate this bias but have to be extended to the multivariate case. Second, inference for connectedness is one of the most obvious extensions of this fundamental analysis. However, naive estimations, such as local-constant kernel regressions, do not allow for meaningful statistical inference. Thus, to apply bootstrap techniques, we need to find asymptotical results for the estimators first. Third, we want to provide a data-driven approach for bandwidth selection in time-varying VARs. Hence, we face the challenge of finding the level of time-variation in the parameters. In summary, this paper aims for a fully-fledged time-varying VAR estimation in the setup of financial connectedness.

We start our analysis with a plain time-varying VAR model. That is, we allow for a dynamic process in the VAR coefficients and the covariance matrix. To derive their estimators, we impose assumptions on the process locally similar to the global assumptions in the static setup. For the dynamic estimation of coefficients and covariances, we provide a data-driven cross-validation approach to select band-

widths for the estimators. In particular, this step sheds light in the dynamics of connectedness and helps to understand the network formation. A simulation analysis underpins the functionality and efficiency of our approach. Finally, a residual-based bootstrap procedure gives confidence intervals for all quantities of interest and neatly avoids the problem of complex distributional results of the coefficients and covariance estimates.

We contribute in four ways to the literature. First, we translate univariate local-linear estimators to the multivariate setting of time-varying VARs with the respective stability conditions. The local stationarity setup of [Dahlhaus \(1996\)](#) gives limiting behavior on the observational level. Second, with the help of theoretical limit results, we extend the literature of time-varying parameters with the possibility to conduct inference on the estimates. In particular, bootstrap techniques enhance the comprehension of the time-varying results. Third, in the application of financial connectedness, the novel estimators give vital insights into the level and evolution of the average connectedness of firms. In particular, the confidence intervals quantify the goodness of the estimates. Last, the dynamic cross-validation is, to the best of our knowledge, the first of its kind and gives valuable information about the change in the dynamics of financial connectedness. This particular finding is essential for policymakers as it shows the accuracy of forecasts.

Our study relates to two strands of the literature: the theoretical considerations of estimating locally stationary processes, and the general analysis of financial spillovers. While we connect to the literature on local stationarity in Section 2.2, we embed this study into the financial spillover context here in the introduction. In particular, since the outburst of the financial crisis in 2007, many authors devoted their attention to the measurement of financial connectedness. Systemic risk papers, e.g. [Adrian and Brunnermeier \(2011\)](#), [Brownlees and Engle \(2012\)](#) and [Acharya et al. \(2017\)](#), focused on pairwise versions of financial dependency. While they estimate associations between firm- and market-movements or vice versa, another strand of the literature, e.g., [Billio et al. \(2012\)](#) and [Diebold and Yilmaz \(2014\)](#), studied the estimation of a network. In particular, the latter gained popularity as its estimation target was a unified network. [Korobilis and Yilmaz \(2018\)](#) extended the rolling sample estimation by a large Bayesian time-varying VAR estimation. In so doing, they observe increased connectedness with the occurrence of extremely adverse shocks and hence implying endogenous network formation. Yet, confidence sets of

spillovers remain uncared for.

The remainder of this study aims to extend the literature in this regard in the following way. Section 2.2 explains the technical concept of local stationarity. Section 2.3 introduces the methodological approach for the statistical audience. Section 2.4 shows the effectiveness and efficiency of the estimation strategy. In section 2.5, we apply the methodology to the concept of financial connectedness.

2.2 Preliminaries

2.2.1 Notation

To study connectedness between financial institutions, we propose a time-varying VAR. For this purpose, we observe data y_1, \dots, y_T from the following N -dimensional time-varying VAR(p) model of the form

$$y_t = \nu(t) + A_1(t)y_{t-1} + \dots + A_p(t)y_{t-p} + u_t, \quad (2.1)$$

where $u_t \sim (0, \Sigma(t))$ is an N -dimensional independent white noise sequence with time-varying variance-covariance matrix $\Sigma(t)$ such that $E(u_t) = 0$, $E(u_t u'_s) = 0$ if $t \neq s$ and $E(u_t u'_t) = \Sigma(t)$. The deterministic parameters of interest in (2.1) are an N -dimensional intercept term $\nu(t)$, the $N \times N$ VAR coefficient matrices $A_i(t)$, $i = 1, \dots, p$ and the innovations' covariance matrix $\Sigma(t)$ that are all allowed to depend on time t . Alternatively, (2.1) can be represented as

$$y_t - \mu(t) = A_1(t)(y_{t-1} - \mu(t-1)) + \dots + A_p(t)(y_{t-p} - \mu(t-p)) + u_t, \quad (2.2)$$

with the relation $\nu(t) = \mu(t) - A_1(t)\mu(t-1) - \dots - A_p(t)\mu(t-p)$ between (2.1) and (2.2).¹ In particular, this representation is useful when we want to derive $MA(\infty)$ representations. In applications, the companion form

$$Y_t = A(t)X_{t-1} + U_t, \quad (2.3)$$

¹Note that there is a difference between the mean μ of a process and an intercept ν in a regression. In the literature on local stationarity it is more common to include a time-varying mean, whereas in the stable VAR literature we rather prefer to have a time-varying intercept.

proves to be useful, where

$$Y_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \quad X_{t-1} = \begin{bmatrix} 1 \\ Y_{t-1} \end{bmatrix}, \quad U_t = \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$\text{and } A(t) = \begin{bmatrix} \nu(t) & A_1(t) & A_2(t) & \cdots & A_p(t) \\ 0 & I_N & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & I_N & 0 \end{bmatrix}.$$

Technically, a time-varying VAR model is a non-linear model which has to fulfill a stability condition of some sort. Clearly, without any further restrictions, the model class (2.1) is too rich and not feasible. It appears to be natural to impose a stability condition on the VAR coefficients that is uniform in t . However, even under a uniform stability condition, the resulting processes will be generally non-stationary due to the dependence on t . Furthermore, in contrast to stationary VAR models, increasing the sample size $T \rightarrow \infty$ in model (2.1), does not provide any further information about the process of the past. Consequently, a standard asymptotic treatment $T \rightarrow \infty$ is a fruitless challenge. To resolve this issue, one can impose (global) parametric assumptions, such as a functional form on the dynamics, but, following the simplicity of rolling-windows, we want to avoid any assumptions of that kind.

To allow for rigorous asymptotic treatments in the time-varying VAR as in (2.1), we rescale the parameter curves $A(t)$ and $\Sigma(t)$ from $[1, \dots, T]$ to the unit interval $(0, 1]$. To be precise, we replace them by $A(t/T)$ and $\Sigma(t/T)$ respectively. Formally, this results in replacing Y_t in (2.3) by a triangular array of observations $\{Y_{t,T} : t = 1, \dots, T, T \in \mathbb{N}\}$ following the recursion

$$Y_{t,T} = A\left(\frac{t}{T}\right)X_{t-1,T} + U_{t,T}, \quad (2.4)$$

where $(U_{t,T})$ is as above with $u_{t,T} \sim (0, \Sigma(\frac{t}{T}))$ and parameter curves $A(\cdot)$ and $\Sigma(\cdot)$ are defined on $(0, 1]$ with suitable domains. For a more convenient treatment of the above recursions (2.4), we extend all curves from the unit interval to the real line, by assuming $A(\tau) = A(0)$ and $\Sigma(\tau) = \Sigma(0)$ for $\tau \leq 0$ and $A(\tau) = A(1)$ and

$\Sigma(\tau) = \Sigma(1)$ for $\tau > 1$.

The main idea of local stationarity introduced to time-varying VAR models (2.1) is to obtain a rescaled model class (2.4) that is globally non-stationary but can be approximated sufficiently well by a stationary VAR process *locally*. This rescaling does not only allow to model changing dynamics of a VAR process but also enables us to apply established techniques for stationary VAR processes. More precisely, it provides a framework to establish a meaningful and rigorous asymptotic theory. In particular, it becomes feasible since, with $T \rightarrow \infty$, the mesh size between observations $y_{t,T}$ becomes finer such that for a given $\tau \in (0, 1]$ more information about the structure of the process becomes available locally. Surely, without any further regularity conditions on the parameter curves, the model class in (2.4) does not allow a meaningful asymptotic treatment. That is, we have to control the model class (2.4) in two aspects. First, we assume that the VAR parameters to fulfill a specific uniform stability condition. Second, we impose sufficient smoothness conditions on the parameter curves that allow for locally stationary approximations to the globally non-stationary process.

In detail, we make the following assumptions.

(A0) The N -dimensional noise process $\{u_{t,T}, t \in \mathbb{Z}\}$ is independent with $\sup_t E(u_{t,T}^{(a)} u_{t,T}^{(b)} u_{t,T}^{(c)} u_{t,T}^{(d)}) < \infty$ for all $a, b, c, d = 1, \dots, D$.

(A1) There exists a $\delta > 0$ such that for all $\tau \in (0, 1]$,² it holds

$$\det(I_K - A_1(\tau)z - \dots - A_p(\tau)z^p) \neq 0 \quad \forall z : |z| \leq 1 + \delta. \quad (2.5)$$

(A2) The parameter curves $\nu(\cdot)$, $A_i(\cdot)$, $i = 1, \dots, p$ and $\Sigma(\cdot)$ are (entry-wise) continuous on \mathbb{R} and twice continuously differentiable on $(0, 1)$.

Intuitively, (A1) guarantees that for each fixed $\tau \in (0, 1]$ the VAR process with parameters $A_1(\tau), \dots, A_p(\tau)$ is stable (stationary and causal). By imposing the smoothness conditions (A2), we achieve that the rescaled tvVAR model can be locally approximated by stationary processes uniformly on the unit interval and, hence, (A2) enables estimation.

²This is equivalent to “for all $\tau \in \mathbb{R}$ ” thanks to the extension of $A_1(\cdot), \dots, A_p(\cdot)$ to the real line.

2.2.2 Local stationarity

Dahlhaus (1996) provides the original definition of local stationarity for univariate time series processes. Even if not needed, he requires all moments of the time series to be finite. Similarly, he also gives the first definition of local stationary suitable for multivariate Gaussian time series in Dahlhaus (2000). Since finite moments or Gaussianity are too restrictive for most applications, we adopt the multivariate definition of Hirukawa (2004, 2012) (we drop their notation about presample values here) that explicitly does not need all moments to be finite or any normality assumption. However, we include an N -dimensional time-varying mean vector as in the definition of Dahlhaus (2000), which leads to the following definition.

Definition 2.2.1 (Local stationarity).

A sequence of multivariate (N -dimensional) processes $y_{t,T} = (y_{t,T}^{(1)}, \dots, y_{t,T}^{(N)})$ ($t = 1, \dots, T$) is called locally stationary with transfer function A^0 and mean μ if there exists a representation

$$y_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp(i\lambda t) A_{t,T}^0(\lambda) d\xi(\lambda) \quad (2.6)$$

where

(i) $\xi(\lambda) = (\xi^{(1)}(\lambda), \dots, \xi^{(D)}(\lambda))$ is a complex-valued stochastic vector process on $[-\pi, \pi]$ with $\overline{\xi^{(a)}(\lambda)} = \xi^{(a)}(-\lambda)$, $a = 1, \dots, D$, and

$$\text{cum}\{d\xi^{(a_1)}(\lambda_1), \dots, d\xi^{(a_k)}(\lambda_k)\} = \eta\left(\sum_{j=1}^k \lambda_j\right) \frac{\kappa_{a_1, \dots, a_k}}{(2\pi)^{k-1}} d\lambda_1 \cdots d\lambda_{k-1}$$

for some $k \geq 2$, $a_1, \dots, a_k = 1, \dots, D$, where $\text{cum}\{\dots\}$ denotes the cumulant of k th order and $\eta(\lambda) = \sum_{j=-\infty}^{\infty} \delta(\lambda + 2\pi j)$ is the period 2π extension of the Dirac delta function.

(ii) There exists a constant B and a 2π -periodic matrix valued function $\mathbf{A} : [0, 1] \times \mathbb{R} \rightarrow \mathbb{C}^{N \times D}$ with $\mathbf{A}(u, -\lambda) = \overline{\mathbf{A}(u, \lambda)}$ and

$$\sup_{t, \lambda} \left| \mathbf{A}_{t,T}^0(\lambda)_{ab} - \mathbf{A}\left(\frac{t}{T}, \lambda\right)_{ab} \right| \leq \frac{B}{T}$$

for all $a, b = 1, \dots, D$ and $T \in \mathbb{N}$. $\mathbf{A}(\tau, \lambda) = (\mathbf{A}(\tau, \lambda)_{ab})$ and $\mu(\tau) = (\mu_1(\tau), \dots, \mu_N(\tau))'$ are assumed to be continuous in τ .

Note that there are various definitions for local stationarity. [Dahlhaus \(2012\)](#) discusses most results under the setup of a time-varying $\text{MA}(\infty)$ representation, which is essentially equivalent. [Puchstein and Preuß \(2016\)](#) make use of an $\text{MA}(\infty)$ representation with Gaussian errors. In contrast, [Vogt \(2012\)](#) and [Vogt and Dette \(2015\)](#) use an alternative definition of local stationarity that relies on a direct approximation without using frequency domain representation. General classes of univariate locally stationary time series processes based on physical dependence conditions have been considered by [Zhou \(2013\)](#), who constructs confidence bands for time-varying AR coefficients. A definition based on time-varying $\text{MA}(\infty)$ representations would suffice for our purpose, but sticking to the original definitions allows to adopt the proofs from the corresponding literature and to extend them to the multivariate case.

Lemma 2.2.1 (tvVMA(∞) representation). *Suppose the parameter curves $A_i(\cdot)$, $i = 1, \dots, p$ and $\Sigma(\cdot)$ fulfill the uniform stability condition (A1) and the smoothness condition (A2). Then, the tvVAR(p) recursions (2.2) have a solution of the form*

$$y_{t,T} - \mu\left(\frac{t}{T}\right) = \sum_{j=0}^{\infty} \Psi_{t,T,j} u_{t-j}, \quad (2.7)$$

with

$$\sup_{t,T} |\Psi_{t,T,j}|_1 \leq C \rho^j$$

for constants $\rho \in (0, 1)$ and $C < \infty$ such that $\sum_{j=0}^{\infty} |\Psi_{t,T,j}|_1 < \infty$ holds uniformly in t and T .

Proof. All proofs can be found in Appendix B.1. □

Lemma 2.2.2. *Suppose the parameter curves $A_i(\cdot)$, $i = 1, \dots, p$ and $\Sigma(\cdot)$ fulfill the uniform stability condition (A1) and the smoothness condition (A2). Then, the tvVAR(p) recursions (2.2) have a solution of the form (2.7) by Lemma 2.2.1 which is locally stationary in the sense of Definition 2.2.1 with*

$$\mathbf{A}(\tau, \lambda) = \frac{1}{\sqrt{2\pi}} \left(I_N - \sum_{j=1}^p A_j(\tau) \exp(-ij\lambda) \right)^{-1} \Sigma^{1/2}(\tau) \quad (2.8)$$

and time-varying spectral density $\mathbf{f}(\tau, \lambda) = \mathbf{A}(\tau, \lambda) \overline{\mathbf{A}(\tau, \lambda)'}'$. Furthermore, there exists a sequence of tvVMA coefficient curves $\{\Psi_j(\cdot) : (-\infty, 1] \rightarrow \mathbb{R}^{N \times D}, j \in \mathbb{N}_0\}$ that are (entry-wise) differentiable on $(0, 1)$ and a process $\{\tilde{y}_{t,T}, t \in \mathbb{Z}, t \leq T, T \in \mathbb{N}\}$ with

$$\tilde{y}_{t,T} - \mu\left(\frac{t}{T}\right) = \sum_{j=0}^{\infty} \Psi_j\left(\frac{t}{T}\right) u_{t-j}, \quad (2.9)$$

such that

$$(i) \quad \sup_{t \in \mathbb{Z}, t \leq T} |y_{t,T} - \tilde{y}_{t,T}|_1 = O_P\left(\frac{1}{T}\right), \quad (2.10)$$

$$(ii) \quad \sup_{\tau \in [0,1]} \sum_{j=0}^{\infty} |\Psi_j(\tau)|_1 < \infty. \quad (2.11)$$

Precisely, it holds that $\sup_{\tau} |\Psi_j(\tau)|_1 \leq C\rho^j$.

Proof. Again, all proofs can be found in Appendix B.1. \square

It is worth noting that a tvVAR model as in (2.4) leads to a tvMA(∞) representation (2.7), which we only approximate by (2.9). This fact is different to the case where the process is already introduced as a tvMA(∞) process and $\mathbf{A}_{t,T}(\lambda) = A\left(\frac{t}{T}, \lambda\right)$ holds. This difference comes by the recursive derivation of the tvVMA representation, which includes VAR coefficient curves that reach into the past. In particular, we find this approach appealing as the tvMA parameter curves $\Psi_j(\cdot)$ are also useful in the calculation of connectedness.

2.3 Methodology

2.3.1 Local-Linear Estimator for Time-Varying VAR Coefficients

In this section, we introduce an easy-to-apply estimator in matrix notation. Appendix B.2 derives a local-linear estimator for local stationary processes. For the sake of notation, we start with the companion form (2.3). That is, the parameters are estimated as a VAR(1). Note, however, that it is possible to avoid the companion form to save computation time.

For any $t \in (1, T)$ and $\tau \in (0, 1]$, we want to estimate $\mu(\tau)$ and $A_i(\tau)$ for $i = 1, \dots, p$. Define matrices as follows,

$$\begin{aligned} Y &= (Y_{1,T}, Y_{2,T}, \dots, Y_{T,T}), & X &= (X_{0,T}, X_{1,T}, \dots, X_{T-1,T}), \\ D_\tau &= \text{diag} \left(\frac{1}{T} - \tau, \frac{2}{T} - \tau, \dots, \frac{T}{T} - \tau \right), & Z_\tau &= \begin{pmatrix} X \\ XD_\tau \end{pmatrix}, \\ W_\tau &= \text{diag} \left(K_{h_1} \left(\frac{1}{T} - \tau \right), \dots, K_{h_1} \left(\frac{T}{T} - \tau \right) \right), & E_1 &= \begin{pmatrix} I_N \\ \mathbf{0}_{N \times N} \end{pmatrix}, \end{aligned}$$

where $K_h(x) = \frac{K(\frac{x}{h})}{h}$ is a kernel function of choice with bandwidth h . The local-linear estimator for the coefficient matrix curve $A(\tau)$ at rescaled time point τ is

$$\hat{A}_\tau = YW_\tau Z'_\tau (Z_\tau W_\tau Z'_\tau)^{-1} E_1. \quad (2.12)$$

First, note that the form of the estimator is similar to the least squares (LS) case. Since LS minimizes the residual sum of squares (weighted or not), the estimator above incorporates LS for the case of local-constant estimation, as in rolling windows. Since we additionally optimize over the derivative, this approach states an extension of rolling windows. A Taylor expansion of order one of $A(\tau)$ around τ approximates the state of the coefficient and its derivative. The optimized derivative can be backed out by using $E_2 = (\mathbf{0}_{N \times N}, I_N)'$ instead of E_1 in (2.12).

Second, the closed form solution gives us an easy-to-apply estimator. Even high dimensional settings are still in reach. In general, however, these settings may suffer under the concept of time-variation in the entries since the estimator assigns the same bandwidth for all of them. Further, we can use model-based bootstrap methods to carry out inference on the parameters, which is novel in the setting of financial spillovers.

In analogy to rolling windows, the degree of time-variation determined by h_1 still needs to be assessed. In the following sections, we introduce an entirely data-driven approach of bandwidth selection.

2.3.2 Local-Linear Estimator for a Time-Varying Innovation Covariance Matrix

After the estimation of \hat{A}_τ for the VAR coefficient matrices in (2.3), we denote the residuals series, with a slight abuse of notation, as $\hat{u}_1, \dots, \hat{u}_T$. That is, we use the companion form,

$$\hat{U}_t = Y_{t,T} - \hat{A}_\tau X_{t-1,T},$$

to extract \hat{u}_t as the first N values of \hat{U}_t . Since we assume that the innovation covariance matrix is time-varying as well, a local-linear estimator would be the obvious choice here. However, the local-linear estimator, as we introduce it later, does not guarantee positive semi definiteness. Thus, we also introduce the local-constant kernel weighted estimator of the innovation covariance matrix. For rescaled time point τ , it is defined by

$$\tilde{\Sigma}_\tau = \sum_{t=1}^T K_{h_2}\left(\frac{t}{T} - \tau\right) \hat{u}_t \hat{u}_t', \quad (2.13)$$

$$= \mathbf{V}(W \mathbf{1}_{(T \times 1)} \otimes I_N), \quad (2.14)$$

where

$$\mathbf{V} = (\hat{u}_1 \hat{u}_1' | \dots | \hat{u}_T \hat{u}_T') \quad (N \times NT),$$

and $\mathbf{1}$ is a vector of ones. Equation (2.14) gives the matrix notation for the estimation step and will prove useful in comparison to the local-linear covariance estimator. In most applications, the degree of time-variation h_2 is heuristically chosen or it can be assumed that the coefficients and covariance estimator share the same degree, i.e. $h_1 = h_2$. However, a different degree of time-variation in the innovation covariance matrix is conceivable and, thus, we use a second bandwidth h_2 for greater flexibility.

Next, we extend the estimator to the local-linear version. In Appendix B.3, we derive the respective local-linear estimator for $\Sigma(\tau)$. That is, we approximate the parameter curve of $\Sigma(\tau)$ by a derivative process, i.e., a Taylor expansion of order one.

With $Z'_{\Sigma,\tau} = [\mathbf{1}_{(T \times 1)} | D_\tau \mathbf{1}_{(T \times 1)}]$ the estimator reads

$$\hat{\Sigma}_\tau = \mathbf{V} \left(W_\tau Z'_{\Sigma, \tau} (Z_{\Sigma, \tau} W_\tau Z'_{\Sigma, \tau})^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes I_N \right). \quad (2.15)$$

Note, that (2.15) is equivalent to (2.13) for τ close to 0.5. This peculiarity roots from the squared minimization, which cancels out the first order of the Taylor Extension $(t/T - \tau)\dot{\Sigma}_\tau$ when squaring it. In the boundaries, however, (2.15) takes into account the local change. Intuitively, it outperforms the local-constant estimator (2.13) at the beginning and the end of the sample.

Again, the estimator is dependent on a second bandwidth, which allows users to calibrate the models more at their discretion. To remedy such behavior, we provide a data-driven machine learning approach to select the bandwidths: cross-validation.

2.3.3 Bandwidth Selection

In this subsection, we deal with the question of bandwidth selection. Similar to rolling windows, in which the econometrician has to choose a window length, the choice of bandwidth implicitly determines the dynamics. While most applications choose a bandwidth which is in line with the economic story, we want to focus on a data-driven bandwidth. We find this step as vital as the estimation itself since the evolution of the parameters is as important as its level. Without a data-driven bandwidth the overall shape of the results is exposed to the author's discretion.

Cross-validation (CV) has proven to be a useful method to select bandwidths. Intuitively, CV divides the full sample into a training data set and test data set and then, based on the best predictive power, selects a tuning parameter. For most econometric applications, CV is heuristically easy to apply. However, in the context of time series analysis, this task becomes particularly challenging since the sample is a single draw of a multivariate time series. This peculiarity complicates the classification of the sample in training and test data, such that some cross-validation methods are computationally infeasible. Hence, we focus on leave-one-out cross-validation (LOOCV).

Leave-one-out cross-validation refers to a training data set which includes all but one observation and, consequently, a test data set which consists of this very observation. The aim is to predict one observation by training on the rest of the

sample. We estimate the coefficients at time point t without observation y_t in order to predict it. However, the estimator in (2.12) generally performs best using a two-sided kernel. Note that using the two-sided 'holey' estimator uses future values and does not 'predict' the value. Thus, it becomes a projection error. The bandwidth with the best fit is the one which minimizes the sum of squared projection errors. By leaving out the central observation, the minimization uses all valuable information around it.

Let $\hat{A}_{\tau,h}^\circ$ be the 'holey' estimator as in (2.12) where we set $K_h(\tau) := 0$, i.e. the estimator without observation y_t with $t = T\tau$.³ The corresponding fitted value of y_t is denoted as $\hat{y}_{t,h}^\circ$. Then,

$$\hat{u}_{t,h}^\circ = y_t - \hat{y}_{t,h}^\circ \quad (2.16)$$

is the 'holey' residual or projection error for bandwidth h . Clearly, this projection error shares a lot of characteristics with the innovation term u_t in (2.1). For example, the innovation covariance matrix Σ_t is in general non-diagonal and thus the covariance matrix of the projection errors is non-diagonal as well.

In the context of financial networks, we believe that links get formed by new contracts getting signed and old contracts maturing. In return, one would expect slowly changing entries in the dependency structure. However, this formation possibly has endogenous aspects. In crisis periods, on the one hand, firms act more cautious and aim to reduce their dependency on others as swift as possible. In more tranquil investment periods, on the other hand, firms slowly build up links over time and become increasingly intertwined. Hence, time-variation also plays a role in how fast dependencies change and, consequently, it is desirable to implement a dynamic bandwidth in this setup. In a nutshell, a dynamic bandwidth counteracts the potential misspecification of dynamics due to the endogenous formation of links.

In analogy to [Niedźwiecki et al. \(2016\)](#), we use a cross-validation procedure for multivariate locally stationary processes. Even though it is desirable to incorporate the non-diagonal covariance matrix in the sum of squared projection errors, the estimation of such requires a residual series and a bandwidth in the first place. Hence, we stick to the plain CV. That is, for time instance τ , we choose the bandwidth $h_{1,\tau}$

³Note the slight abuse of notation. We need the τ notation to denote the observation we want to estimate.

with

$$h_{1,\tau}^{CV} = \arg \min_h \sum_t K_{h_3} \left(\frac{t}{T} - \tau \right) \hat{u}_{t,h}^\circ{}' \hat{u}_{t,h}^\circ. \quad (2.17)$$

By minimizing this criterion, we check the quality of the model estimates with bandwidth h around τ . The degree of time-variation in $h_{1,t}$ is then determined by h_3 , which we assume to be given.

Note that (2.17) gives us a series of bandwidths for the coefficients. Given this series, we construct a residual series \hat{u}_t , which is then used for the covariance estimates. Similar to the sparse covariance estimation in [Cai and Liu \(2011\)](#), we extend the idea of cross-validation to the second moment of residuals. That is, we employ a distance criterion to validate the goodness of fit of an estimator given a bandwidth. Thus, define the 'holey' estimator for the covariance matrix (2.13) with bandwidth h at τ as

$$\tilde{\Sigma}_{\tau,h}^\circ = \sum_{t=1; t \neq \tau T}^T K_h \left(\frac{t}{T} - \tau \right) \hat{u}_t \hat{u}_t', \quad (2.18)$$

and, respectively for the local-linear estimator (2.15),

$$\hat{\Sigma}_{\tau,h}^\circ = \mathbf{V} \left(W_\tau^\circ Z'_{\Sigma,\tau} (Z_{\Sigma,\tau} W_\tau^\circ Z'_{\Sigma,\tau})^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes I_N \right), \quad (2.19)$$

where W_τ° is set to give weight zero at τ , i.e. $K_h(\tau) := 0$. Again, with a slight abuse of notation, we use index t on these 'holey' covariance estimates. Then, $\hat{\Sigma}_{t,h}^\circ$ is the covariance estimate for the innovations u_t without using the residual \hat{u}_t . Since we assume a smooth evolution of Σ_t , we again want to find the bandwidth with the most predictive power. In particular, we test the 'trained' estimator on the test data set $\hat{u}_t \hat{u}_t'$. Denote the distance to the squared residuals at each time point as

$$d_t(\hat{\Sigma}_{t,h}^\circ) = \|\hat{u}_t \hat{u}_t' - \hat{\Sigma}_{t,h}^\circ\|_F. \quad (2.20)$$

The weighted leave-one-out criterion at τ reads

$$h_{2,\tau}^{CV} = \arg \min_h \sum_t K_{h_4} \left(\frac{t}{T} - \tau \right) d_t(\hat{\Sigma}_{t,h}^\circ), \quad (2.21)$$

where $\|\cdot\|_F$ denotes the Frobenius norm and h_4 again determines the degree of time-variation of the bandwidth for the covariance estimator. This step evaluates which bandwidth has the best explanatory power for the covariance of \hat{u}_τ around τ . Thus, this step also yields a dynamic bandwidth as we let the kernel $K_{h_4}(\frac{t}{T} - \tau)$ sweep through the sample, i.e. for all $\tau \in (0, 1]$. Following all proposed steps yields two series of bandwidths, $h_{1,\tau}$ and $h_{2,\tau}$. Needless to say, these two series give us further insights into the dynamics of our model, as we are also able to see changes in the degree of time-variation of the coefficients and the covariance.

Note that dynamic bandwidths imply that the dynamics of the parameter estimates are less constrained and, thus, it might look odd compared to a regular fixed bandwidth regression. Consequently, we have to ensure that the bandwidth series are smooth enough for neat regression results. The previous step is, however, performed over a grid of bandwidths and, hence, we will kernel smooth the two bandwidth series in the application.

2.3.4 Time-Varying Forecast Error Variance Decompositions

To link the implied dependency structure of coefficients and the innovation covariance matrix to networks, [Diebold and Yilmaz \(2009\)](#) interpreted Forecast Error Variance Decompositions (FEVD) as spillover tables. In this section, we shed light in the abstractness of FEVDs and show how they behave with time-varying parameter estimates. We assume a time-varying VAR(1) representation of (2.1). This VAR(1) can be the companion form of a VAR(p) estimation as in (2.3). In the following, we drop the notation of locally stationary processes to generalize FEVDs for all time-varying VAR models.

Take (2.7), which presents the canonical representation of an N -dimensional local stationary VAR process $y_{t,T}$,

$$y_t - \mu_t = \sum_{j=0}^{\infty} \Psi_{j,t} u_{t-j}, \quad \Psi_{0,t} = I_N, \quad \forall t = 1, \dots, T, \quad (2.22)$$

where u_t is defined as above with covariance Σ_t . The tvMA parameters are recursively defined as

$$\Psi_{i,t} = \left(\prod_{j=0}^{i-1} A_{t-j} \right). \quad (2.23)$$

For a lag order higher than 1, the matrix of interest is the companion form estimate of (2.3) without the intercept (first) column. We can rewrite Σ_t as $\Sigma_t = P_t P_t'$, with P_t as a lower triangular matrix, i.e. with Cholesky decomposition. Consequently, (2.22) reads

$$y_t = \mu_t + \sum_{j=0}^{\infty} \Psi_{j,t} P_t P_t^{-1} u_{t-j} = \mu_t + \sum_{j=0}^{\infty} \Phi_{j,t}^o \epsilon_{t-j}$$

where $\Phi_{j,t}^o := \Psi_{j,t} P_t$ and $\epsilon_t := P_t^{-1} u_t$ is white noise with the identity as covariance matrix.

The responses functions of an orthogonal impulse at time t on y_{t+h} are in the columns of the $(N \times N)$ matrix:

$$\Phi_{h,t}^o = \Psi_{h,t} P_t \quad (2.24)$$

Since the Cholesky decomposition is not invariant to reordering, [Pesaran and Shin \(1998\)](#) proposed a different method. They set a single impulse to variable j with size δ_j , such that the expectation of the innovation vector becomes

$$E[u_t | u_{j,t} = \delta_j] = (\sigma_{1j,t}, \dots, \sigma_{Nj,t}) \sigma_{jj,t}^{-1} \delta_j = \Sigma_t e_j \sigma_{jj,t}^{-1} \delta_j,$$

with σ_{jj} being the j -th diagonal value of Σ . We can interpret the latter as a generalized impulse as it controls for correlated effects. The response functions of generalized impulses with the size of one standard deviation $\delta_j = \sqrt{\sigma_{jj,t}} \forall j$ at time t on y_{t+h} are in columns of the $(N \times N)$ matrix:

$$\Phi_{h,t}^g = \Psi_{h,t} \underbrace{\Sigma_t \text{diag}(\Sigma_t)^{-\frac{1}{2}}}_{=P_t^g}, \quad (2.25)$$

where $\text{diag}(\Sigma_{w,t})$ denotes a diagonal matrix with the diagonal values of Σ_t and P_t^g is the matrix 'decomposition' similar to P_t in the orthogonal case. Note, however, that P_t^g is not a decomposition of Σ_t since $\Sigma_t \neq P_t^g P_t^{g'}$. Thus, this case does not create orthogonal innovations, but, in fact, is invariant to reordering.⁴

When squaring the elements of (2.24) and (2.25) we obtain the orthogonal forecast

⁴The resulting innovations $\epsilon_t^g = P_t^{g-1} u_t$ are by no means shocks, since their covariance matrix is the inverse correlation matrix $\text{diag}(\Sigma_t)^{\frac{1}{2}} \Sigma_t^{-1} \text{diag}(\Sigma_t)^{\frac{1}{2}}$, which is non-diagonal. Entries in this matrix can be linked to partial correlations, see e.g. [Raveh \(1985\)](#).

error variance and the generalized forecast error variance respectively. The H -step forecast error variances at time t read as⁵

$$\Gamma_t^{l^2}(H) = \sum_{h=0}^{H-1} (\Phi_{h,t}^l)^2 + \dots + \Phi_{h,t}^l \quad l = \{o, g\}. \quad (2.26)$$

where M^2 denotes the element-wise squared matrix M .

Following Diebold and Yilmaz (2014), we denote the H -step orthogonalized and the generalized Forecast Error Variance Decompositions as $D_t^{oH} = [d_{ij}^{oH}]$ and $D_t^{gH} = [d_{ij}^{gH}]$, respectively, where

$$d_{ij,t}^{oH} = \frac{e_i' \Gamma_t^{o^2}(H) e_j}{e_i' \Gamma_t^{o^2}(H) \mathbf{1}_{N \times 1}}, \quad d_{ij,t}^{gH} = \frac{e_i' \Gamma_t^{g^2}(H) e_j}{e_i' \Gamma_t^{g^2}(H) \mathbf{1}_{N \times 1}} \quad (2.27)$$

with e_i being the i -th column of the $(N \times N)$ identity matrix and thus $e_i' \Gamma_t^{o^2}(H) \mathbf{1}_{N \times 1}$ being the mean squared forecast error of variable i . This ratio then explains the j -th percentage contribution on the total forecast variance of all variables on variable i . Remember that since P_t^g is not a decomposition of Σ_t , rows of the generalized forecast error variance decompositions do not sum to one.

Normalizing these entries leads us to our network interpretation, where the normalized spillover table \tilde{D}_t^{gH} marks our adjacency matrix. In particular, we want to see contagion of shocks to firms relative to their own shock affection. In Diebold and Yilmaz (2014), the spillover table comes with measures of connectedness

$$\begin{aligned} C_{i \leftarrow \cdot} \left(\tilde{D}_t^{gH} \right) &= \sum_{j \neq i} \tilde{d}_{t,ij}^{gH}, & (\text{From Connectedness to } i) \\ C_{\cdot \leftarrow j} \left(\tilde{D}_t^{gH} \right) &= \sum_{i \neq j} \tilde{d}_{t,ij}^{gH}, & (\text{To Connectedness from } j) \\ C \left(\tilde{D}_t^{gH} \right) &= \frac{1}{N} \sum_i \sum_{j \neq i} \tilde{d}_{t,ij}^{gH}. & (\text{Average Connectedness}) \end{aligned}$$

While the first two measures depict variable specific attributes, Average Connectedness is the average of one of the two. That is, it shows how much effect one

⁵Even though these can be technically described as decompositions, we denote them as forecast error variances to abstract from the later construct, which is a relativized version.

variable takes on average from others. From Connectedness, in contrast, shows how much i takes from others, and vice versa for To Connectedness. Note here, that the connectedness matrix \tilde{D}_t^{gH} is row normalized, i.e., all numbers in the matrix are relative to the total effect taken. For a more detailed discussion about the measures, we refer to [Diebold and Yilmaz \(2014\)](#).

2.3.5 Inference with Bootstrap Confidence Intervals

The previously introduced concepts give a clear picture of the extent to which variables contribute to each other and hence allow for a network interpretation. Ideally, we would like to underpin the estimation of those with some distributional results. In the upcoming subsection, we focus on providing confidence intervals for the estimated networks.

With distributional results of the estimators, one can directly apply them to the estimates. However, network tables are complex functions of the estimation results, and therefore limiting distributions are hard to calculate. Moreover, limiting distributions do not work for all bandwidths, e.g., for small bandwidths, our estimators' variances may be falsely specified. Hence, we focus on bootstrapping methods due to their simplicity and universality. That is, we use the residual bootstrap, which is known to be a flexible and sophisticated approach for linear models.

For a given number of bootstrap repetitions, B , and all the estimates from the previous sections, we create confidence intervals with the following steps:

1. Calculate residuals $\hat{U}_t = Y_t - \hat{A}_t X_{t-1}$ and extract \hat{u}_t ⁶
2. Define orthogonal innovation terms $\tilde{\epsilon}_t := \hat{\Sigma}_t^{-\frac{1}{2}} \hat{u}_t$ that are approximately identically distributed. $\hat{\Sigma}_t^{-\frac{1}{2}}$ is any decomposition of the innovation covariance estimate in (2.13) or (2.15)
3. Draw randomly (with replacement) from the set of \tilde{u}_t to generate a new times series with the orthogonal innovations

$$Y_t^* = \hat{A}_t X_{t-1}^* + \underbrace{\hat{\Sigma}_t^{\frac{1}{2}} \tilde{\epsilon}_{t^*}}_{iid \text{ draw}}, \quad \forall t = 1, \dots, T,$$

$$X_0^* = X_0.$$

⁶Note that these are independent but not identically distributed, due to the fact that we have time-variation in the residual series as well.

4. Calculate the bootstrap coefficient and covariance estimates on the generated time series with (2.12), and (2.13) or (2.15). Save the bootstrapped parameters of interest in $\hat{\theta}_b^*$.
5. Repeat steps 2, 3 and 4 B times, where B is large.
6. Calculate differences for the parameter of interest $\hat{\theta}_b^* - \hat{\theta}$ and subtract the quantile from the estimate. Denote the α -quantile of the empirical distribution of $\hat{\theta}_b^*$ by $q_{\alpha/2}^*$ and $q_{1-\alpha/2}^*$, then the bootstrap confidence interval for $\hat{\theta}$ is given by

$$[2\hat{\theta} - q_{1-\alpha/2}^*, 2\hat{\theta} - q_{\alpha/2}^*].$$

Note that the intervals, as constructed in step 6, are known as the basic Hall bootstrap intervals. Alternatively, one could also use the Efron intervals, which are the percentile quantities of the empirical distribution.

In this paper, we focus on θ as an entry of the spillover table as in (2.27). By applying the confidence intervals to spillover tables, we provide an idea of how connectedness changes over time. This information helps to abstract from differences caused by estimation errors.

2.4 Simulation Study

In the upcoming section, we focus on the proficiency and applicability of the method. In particular, we are interested in the three critical steps of the method. First, a simulation of the estimators for the coefficients and the innovation covariance matrix shows the applicability of local-stationary approaches. Second, we calculate coverage rates to show the performance of the previously introduced bootstrap. Third, since this paper does not provide a theoretical background for the cross-validation, we test its performance to the ideal bandwidth.

All three simulations follow the same parameter generating process. In particular, we focus on $N = 3$ with time-varying coefficients and innovation covariance matrix. In order to show their fit, we randomly generate stationary time-varying coefficients and covariance matrices with entries from a sine function having random magnitude, frequency, and shift for all entries. A more detailed description of the parameter

generating process can be found in Appendix B.4. Note that generated parameter series which are not stable are discarded due to the assumption of local stationarity.

Illustration of the Estimators

This subsection intends to give an understanding of the local-linear estimator. We show one example to give insights into the performance of the estimator in the time-varying setup. This example, in particular, is a negative example as it shows the estimator under challenging conditions, e.g., different frequencies for the entries. However, we hope that this example can highlight the estimator's advantages. We randomly generate a parameter series, which, in return, generates a time-series for each Monte-Carlo repetition with 1000 observations and respective Forecast Error Variance Decompositions.

Figures 2.1, 2.2 and 2.3 show the mean of the Monte-Carlo estimation of (2.12) for the (3×3) matrix function A_t , Σ_t and D_t^{gH} respectively. We select the bandwidth which minimizes the Frobenius norm to the true parameters and denote it as the oracle bandwidth.

First, note that the local-linear and local-constant estimators for the coefficient curves A_t look similar. The local-constant kernel estimator (dashed blue) performs slightly worse at the boundaries but slightly better in this example (see B.1 in Appendix B.3). Furthermore, we observe that the estimator is not entirely unbiased. While it gets an idea of the overall form and level of the entries, it is not able to match the peaks of the true curve. However, the derivative structure of the local-linear estimator lets it perform better at the boundaries of the sample.

A similar picture, but with a better overall estimation result is shown in Figure 2.2. We observe similar performance with a slightly better estimation at the boundaries by the local-linear estimator.

Last, the connectedness matrix D_t^{gH} shows that the boundary effect enhances. That is, Figure 2.3 shows that the local-linear estimator has a significantly better understanding of where the trend is going. This result is particularly essential since forecasting methods require accurate estimations at the end of the sample. For example, institutions are mostly interested in the 10-day capital shortfall conditional on another variable, which is the Forecast Error Variance Decomposition at the end of the sample.

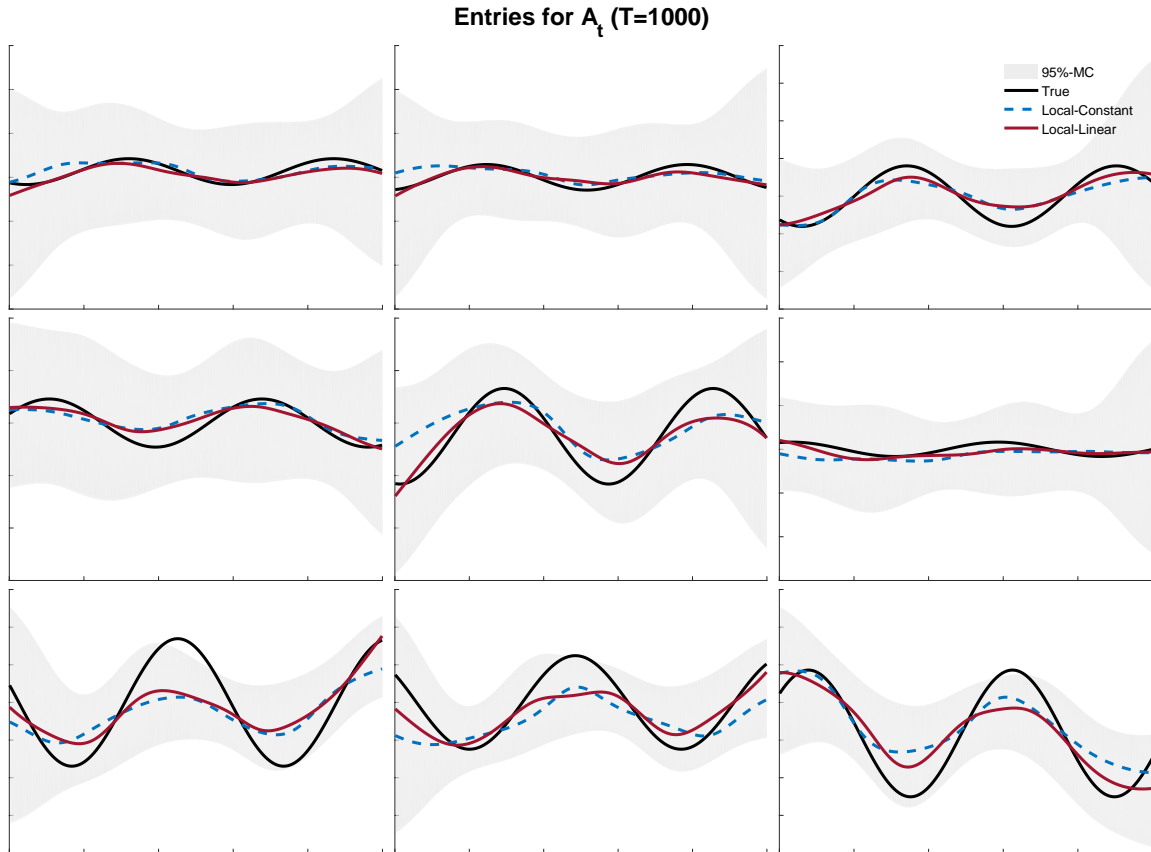


Figure 2.1: Monte Carlo Simulation of A_t with 500 repetitions. Mean comparison of the local-linear estimator (2.12) (red) and the local-constant estimator *aka* kernel-weighted LS-estimator (dashed blue). The black line depicts the true DGP parameters. The grey area shows the Monte-Carlo bands.

In summary, this particular negative example slightly favors the local-constant estimator for the coefficient curves. However, one has to keep in mind that the more essential periods are the ones at the end of the sample. Thus, we recommend using the local-linear estimator as it outperformed the local-constant in the boundaries.

Bootstrap Performance

Table 2.1 shows the coverage rates of the simulation for 200 Monte-Carlo and 1000 bootstrap repetitions. First, it is evident that the coverage rates are permissive for both the basic Hall bootstrap interval (first row) and the Efron percentile bootstrap interval (second row). Second, the intervals perform similarly with virtually no differences even though the first one is bias corrected. Consequently, we most likely

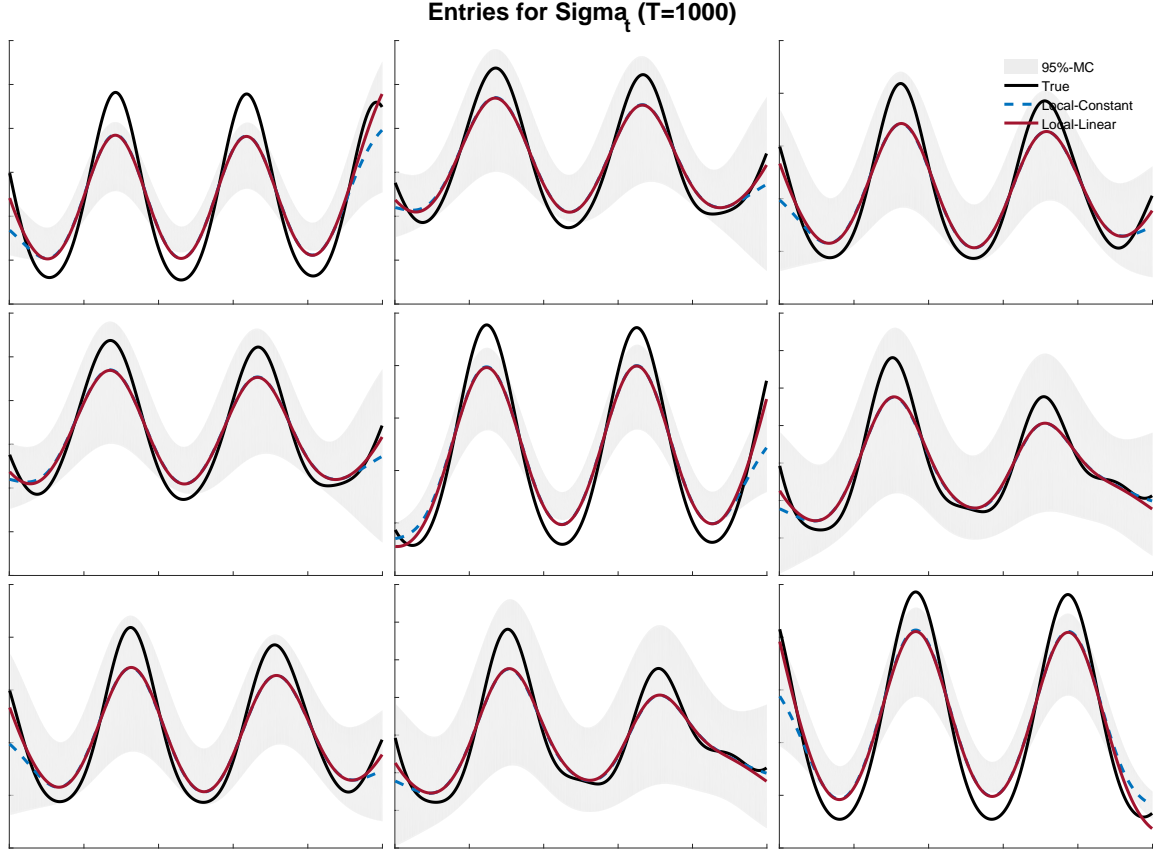


Figure 2.2: Monte Carlo Simulation of Σ_t with 500 repetitions. Mean comparison of the local-linear estimator (2.12) (red) and the local-constant estimator *aka* kernel-weighted LS-estimator (dashed blue). The black line depicts the true innovation covariance entries. The grey area shows the Monte-Carlo bands.

cannot correct the bias with a bootstrap, albeit the fact that the permissive behavior might be inherited from a bias. In contrast to the application, these simulations are calculated with a static oracle bandwidth.⁷

Cross-Validation

We now conduct a simulation analysis of the cross-validation steps as in section 2.3.3. In particular, we use (2.19) as an estimator of the covariance matrix since this seems to perform slightly better boundaries. Taking the other covariance estimator for increased computation performance gave similar results. Figure 2.4 shows the distribution for ' $h^{oracle} - h^{CV}$ ', where the oracle bandwidth is the bandwidth, which

⁷This is the bandwidth that minimizes the Frobenius norm to the true value

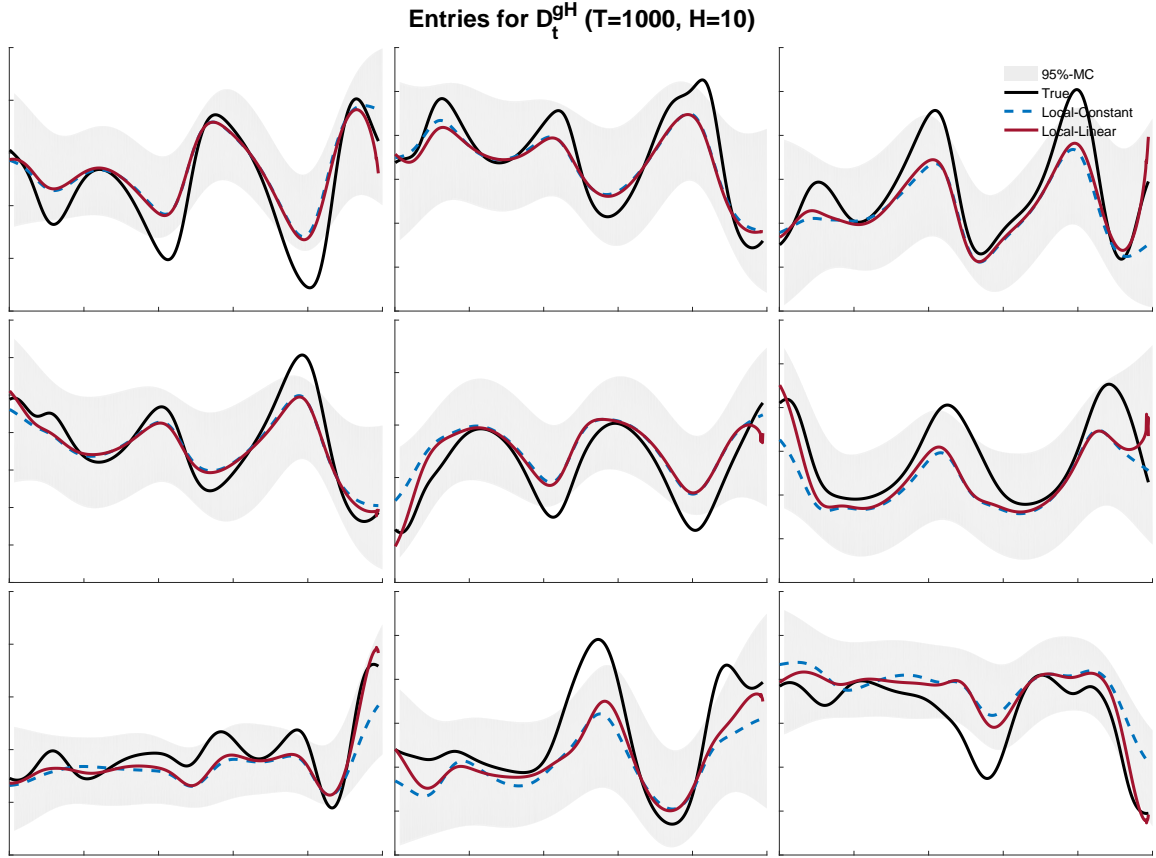


Figure 2.3: Monte Carlo Simulation of D_t^{gH} with 500 repetitions. Mean comparison of the local-linear estimator (2.12) (red) and the local-constant estimator *aka* kernel-weighted LS-estimator (dashed blue). The black line depicts the true innovation covariance entries. The grey area shows the Monte-Carlo bands.

minimizes the Frobenius norm to the true parameters.

| | α | | | |
|--|----------|--------|--------|--------|
| | 0.2 | 0.1 | 0.05 | 0.01 |
| $[2\hat{\theta} - q_{1-\alpha/2}^*, 2\hat{\theta} - q_{\alpha/2}^*]$ | 0.7146 | 0.8235 | 0.8889 | 0.9575 |
| $[q_{\alpha/2}^*, q_{1-\alpha/2}^*]$ | 0.7225 | 0.8244 | 0.8848 | 0.9470 |

Table 2.1: Monte-Carlo simulation results for 200 repetitions and a 1000 bootstrap repetitions. Coverage rates are shown for the basic Hall bootstrap interval and the percentile interval. $\hat{\theta}$ denotes the estimation target \hat{A}_t and quantiles are determined by α .

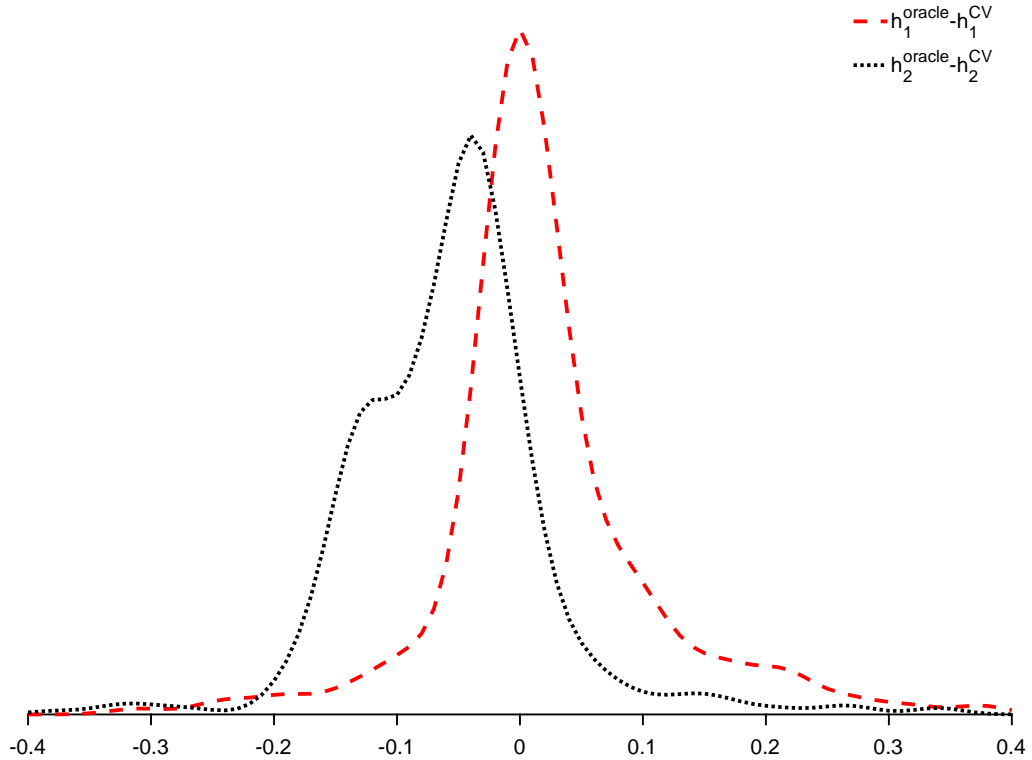


Figure 2.4: Distributional results for the distance between the oracle bandwidth and the cross validation selected bandwidth. The oracle bandwidth is the bandwidth which minimizes the norm to the true parameter curves. The CV bandwidths are chosen with the cross-validation criteria in (2.17) and (2.21). Distributions are kernel smoothed with a gaussian kernel.

First note that the cross-validation approach for the coefficients as in (2.17) appears to have a Gaussian type distribution. In average, it finds the best bandwidth, and its distribution appears to be rather symmetric. The cross-validation for the

covariance estimates as in (2.21) performs slightly worse as it overestimates the bandwidth relative to the oracle one. Nonetheless, it also shows something similar to a Gaussian bell with a small recess on the left side. This result is not surprising since the second step cross-validation (2.21) carries estimation errors from the first cross-validation step and the estimation of the residuals. Thus, this cross-validation has a higher standard deviation. Since bandwidths cannot be negative, the difference mostly appears on the left side of this distribution. In a nutshell, we can see that the cross-validation steps give consistent results with acceptable errors and, hence, are good indicators for the best fit in a time-varying setup.

2.5 Time-Varying Financial Connectedness

Financial connectedness as introduced by [Diebold and Yilmaz \(2014\)](#) is a concept to monitor various key figures for risk management at once. In this section, we proceed at an intuitive level and refer back to Section 2.3.4, or to [Diebold and Yilmaz \(2009\)](#), for more rigorous definitions. First, we want to emphasize the obvious and link this concept to systemic risk. The fact that the connectedness table shows contributions of one firm to the other gives a clear hint about the long-run dependency structure. Thus, it helps to identify systemic issues, and also shows potential gridlock risk.

Furthermore, financial connectedness links to other risk concepts such as market risk and portfolio risk. Time-variation is not only a crucial feature in detecting trends in market risk, but portfolio risk becomes particularly consequential when we observe time-varying connectedness. For example, in a crisis, connectedness may rise and, thus, imply worse diversification opportunities. Consequently, portfolio managers who have insights into the real dynamics of time-varying connectedness can exploit the additional knowledge to generate extra risk-adjusted excess returns (see [Fleming et al. \(2001, 2003\)](#)). In the upcoming section, we hope to shed more light in the dynamics of financial connectedness by empirically assessing volatility spillovers in the financial market.

2.5.1 Data

Our empirical section hovers around the setup of volatility spillovers of [Diebold and Yilmaz \(2014\)](#). We find this application, in particular, appealing since it gives

an idea of how strongly financial firms are entangled. Return volatility is known to show investors fear about the financial health of firms. Clearly, risk figures as probability of default also highlight firms which are more likely to get into financial distress. However, prices of credit default swaps as market-based variables have less trading volume than stocks, and we leave this application for future research. In our application on time-varying parameters, we believe that it is vital to have a big trading volume to measure spillovers consistently. Hence, we focus on realized volatility as a measure for the variance of returns. That is, we use the extreme-value estimation measure by [Parkinson \(1980\)](#),

$$\tilde{\sigma}_{i,t}^2 = 0.361(p_{i,t}^{high} - p_{i,t}^{low})^2.$$

Where $p_{i,t}^{high/low}$ denotes the daily maximum and daily minimum of intra-day log-prices respectively. Similarly, we can use the realized volatility estimation of [Andersen et al. \(2003\)](#), which is based on high-frequency data. However, we find the extreme value estimation particularly appealing due to its simplicity and its robustness to recording errors.

Ideally, the analysis considers all available variables to understand connectedness as a whole. Unfortunately, computational constraints limit our analysis to a relatively low number of financial firms. Thus, we focus on the most prominent institutions. In particular, we include banks, insurance companies (AIG), and other financial services (American Express). Moreover, since the analysis hovers around the idea of responses, we have to ensure that all firms have the same trading times. Hence, we can only include non-American institutions, which get traded on the New York Stock Exchange (NYSE). That is, we focus on the biggest banks from Europe and Japan. Due to missing data, we are not able to include Banks such as BNP Paribas or Chinese Banks. We use CRSP data and fill missing data points (for Deutsche Bank and Credit Suisse) with Yahoo Finance.⁸ The data starts at 03/01/2000 and ends at 29/06/2018.

Table 2.2 describes the 15 institutions and their growth in total assets and market capitalization over the years. Unfortunately, the fourth quarter financial report of 1999, which is the closest to the start of our sample, has too many missing values. Table 2.2 thus shows the 2000Q1 values and interpolates missing values.

⁸We observed that both institutions have weak data from 2000Q1 to 2001Q3. Thus, their estimation for this period is fuzzy.

Note that for every single institution, the number of total assets increased. In particular, the big banks had a significant increase in total assets from 2000Q1 to 2018Q2. For market capitalization, the common trend is upwards. Apart from AIG, Deutsche Bank, and Credit Suisse, all institutions experienced an increase in market capitalization. At the end of the sample, MUFG, HSBC, JPMorgan Chase, and Bank of America are the biggest institutions.

| Institution | Ticker | Country | Total Assets | | Market Cap. | |
|---------------------|--------|---------|--------------|--------|-------------|--------|
| | | | 2000Q1 | 2018Q2 | 2000Q1 | 2018Q2 |
| American Int. Grp | AIG | US | 279.3 | 496.8 | 168.8 | 47.3 |
| American Express | AXP | US | 150.7 | 184.9 | 66.2 | 84.4 |
| Bank of America | BAC | US | 656.1 | 2291.7 | 86.9 | 282.3 |
| Bank of NY Mellon | BK | US | 76.0 | 352.9 | 30.6 | 53.9 |
| Citigroup | C | US | 738.2 | 1912.3 | 201.8 | 168.4 |
| Credit Suisse** | CS | SUI | 613.1 | 817.2 | 57.1 | 37.2 |
| Deutsche Bank | DB | GER | 912.5 | 1657.9 | 40.8* | 21.9 |
| Goldman Sachs | GS | US | 276.9 | 968.6 | 44.8 | 85.6 |
| HSBC | HSBC | UK | 574.7* | 2607.3 | 103.5* | 187.3 |
| JPMorgan Chase | JPM | US | 391.5 | 2590.1 | 71.9 | 350.2 |
| Mitsubishi UFG*** | MUFG | JPN | 750.5 | 2698.5 | 44.5 | 73.9 |
| Morgan Stanley | MS | US | 408.1 | 875.9 | 79.9 | 82.9 |
| Santander Group | SAN | ES | 258.1 | 1673.0 | 43.0 | 86.2 |
| Toronto-Dominion B. | TD | CAN | 233.9 | 1283.8 | 22.5* | 133.0 |
| Wells Fargo | WF | US | 66.4 | 293.6 | 222.3 | 9.9 |

Table 2.2: Key figures of financial institutions in bn. US\$ (source: Compustat, *interpolated missing values, **CS reported for 2000Q4 and 2017Q2***MUFG reported for 2001Q2 and 2018Q1,)

We conduct our analysis on Generalized Forecast Error Variance Decompositions from Section 2.3.4 for a forecast horizon $H = 10$. The respective matrices are estimated with the local-linear estimates (2.12) and (2.15) and bandwidths of (2.17) and (2.21). We set the dynamics of the cross-validation to

$$h_3 = h_4 = \frac{250}{2q_{0.95}T},$$

where T is the sample length and $q_{0.95}$ denotes the 95% percentile of the respective kernel distribution. This bandwidth gives 95% of the weights to 250 trading days (one calendar year). We use the Gaussian kernel due to its good asymptotic

behavior, and its viability with the covariance estimate.⁹

2.5.2 Empirical Results

This section intends to show time-varying estimates of financial connectedness. In contrast to [Diebold and Yilmaz \(2014\)](#), we provide insights into spillovers to non-US institutions, reaction times of investors, and confidence bands for the estimator. Although the new dynamics of the estimators and its confidence bands add information on an institutional level (e.g., From and To Connectedness), we zoom out to the root of the problem, i.e., connectedness as such. In the following, we first present the implied network as a time-average of the FEVD and the last observational period. Second, we analyze the dynamics of Average Connectedness by showing the cross-validation results, Average Connectedness, and its confidence sets.

Figure 2.5 shows the 25% strongest connections of the mean network, which is the average of the networks at each time point. The nodes are arranged with force-directed graph drawing. This method plots the most central nodes in the middle and arranges nodes closer to each other when they share more connections. However, presenting all edges in one graph would yield a muddle to the reader. Thus, for this figure, the smallest 75% edges are set to zero and, consequently, the force layout arranges the nodes neatly.

It is salient that the institutions are formed in clusters by regions. The US financial institutions form a cluster, which has mainly banks in the center. The European banks form a cluster on its own, where every single one has a connection in the top 25%. For the Canadian Toronto-Dominion and the Japanese Mitsubishi UFJ Financial Group, we do not observe any connections in the top 25%. Note, however, that this does not imply that they are less critical for the financial system. It only states that for this group of financial institutions, the impact is significantly lower than for the rest.

On a similar note, we analyze the most recent network in Figure 2.6, the estimated network as of 30/06/2018. Again, only the 25% strongest connections are shown in a force-directed network graph.

While it is clear that we can again observe the clusters by regions, this network shows that the clusters do share more spillovers. Since we only show the 25% most

⁹Other kernels might improve the computation time, but possibly produce non-positive definite matrices.

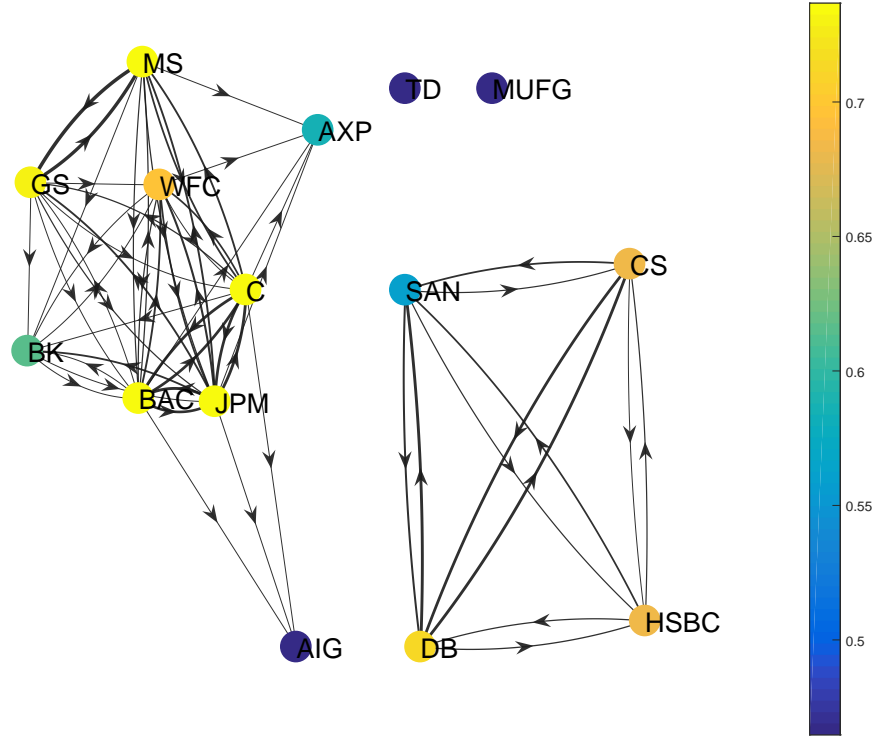


Figure 2.5: Network graph of the mean network over time. The strongest 25% dependencies are shown. Colors indicate the out-degree of the respective institution.

substantial dependencies, it is particularly interesting that we do not observe any isolated nodes anymore. This observation implies that institutions have gotten relatively more connected to institutions from other regions. Even though globalization plays a significant role, it is evident that in this sample the Swiss Credit Suisse and the British HSBC contributed to more global connectedness. They increased their spillovers to others (see the change in color of the node relative to the mean network in Figure 2.5). Furthermore, the scale on the right side shows the out-degrees, which, when comparing to the mean network out-degrees, are higher at the end of the sample.

In order to analyze how dependencies changed over time, we show the average connectedness in Figure 2.7. Before getting into detail, it is worth noting that, in contrast to the Bayesian time-varying VAR of [Korobilis and Yilmaz \(2018\)](#), this estimation has a smoother development of connectedness. The assumptions on the estimation targets require the evolution to be sufficiently smooth. Intuitively, the

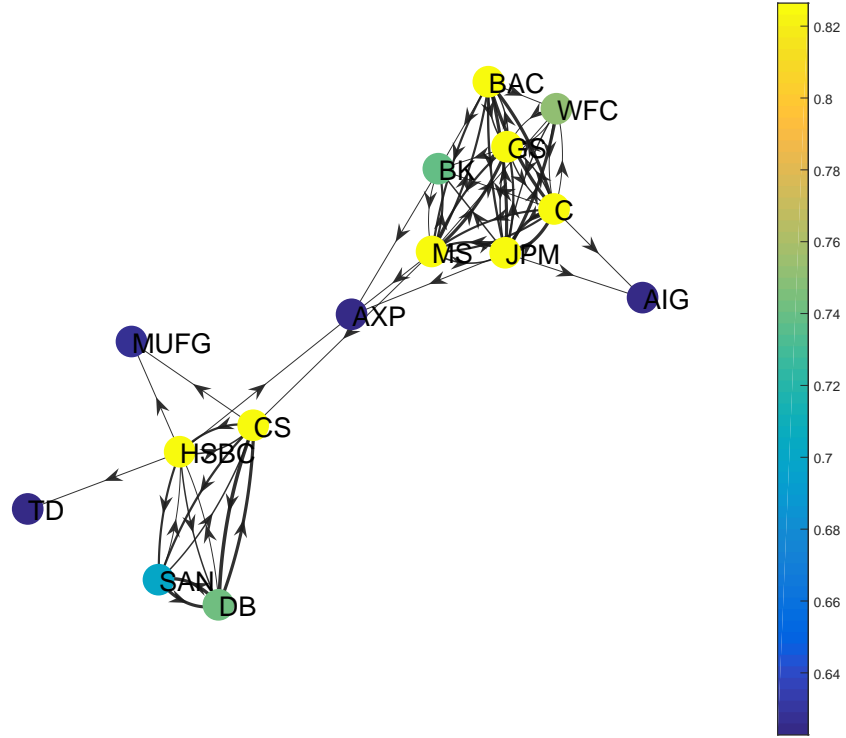


Figure 2.6: Network graph of the network at the 30/06/2018. The strongest 25% dependencies are shown. Colors indicate the out-degree of the respective institution.

smoothness assumes that the network forms exogenously. A Bayesian time-varying VAR allows them to be affected by a shock and, thus, it shows sharper jumps in the connectedness. This behavior is imaginable if we assume endogenous network formation. That is, an extremely negative or positive realization affects the dependency structure instantaneously. Note that the two-sided local-linear estimation approach somehow prohibits that since it smoothly adapts to bigger volatilities of before adverse shocks to make them likely enough to happen.

Figure 2.7 shows Average Connectedness in blue with its confidence intervals in grey. The red lines depict the cross-validation bandwidths with values on the right scale. Most notably, there is a clear upwards trend over nearly the whole sample. Even though we allow for time-variation of all sorts, the result seems to have a nearly linear trend. Solely from 2010 until 2015, the connectedness did not increase. The end of the sample shows an alarming average connectedness of financial firms. More precisely, the amount of average spillovers to other firms exceeds the spillovers

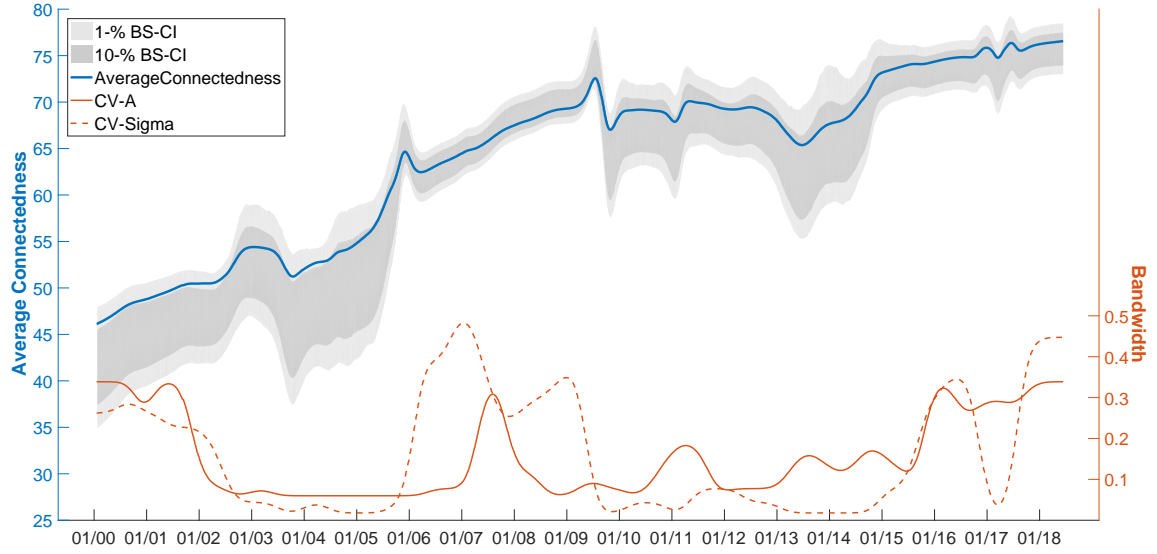


Figure 2.7: Average Connectedness (blue) over time. The grey areas depict the 90% and 99% bootstrap confidence interval. The red lines depict the selected bandwidth for the coefficient (solid) and the covariance (dashed) estimate.

during the financial crisis.

A technically compelling result is the seeming independence of the level of Average Connectedness of adverse events. We see this finding as support for a more exogenously formed network. Specifically, the series of extreme events during the great recession from 2007 to 2009 only shows changes in the selected bandwidths.

The bandwidth of the covariance estimate increased to an all-time high before the crisis and shrank with the first negative news at the beginning of 2007. The coefficient's bandwidth increases at that time but shortly after drops again. If we see the covariance matrix as the immediate co-effects rooting from mutual exposures, such as portfolio holdings or increased market risk, a lower bandwidth implies that most of these effects are experiencing more change. A higher bandwidth, as we observe it in that period, stands for a more persistent mutual exposure to the same risks — namely, the bad loans from the subprime crisis.

Beyond, we interpret the coefficients as directed and more causal dependencies of institutions, e.g., causal relations issuing from asset and liability structures, or options and derivatives. Thus, higher bandwidths for the coefficient's estimates relate to more rigidity in the asset and liability structure of the institutions. The

peak at the end of 2007 can then be interpreted as a phase of greater mistrust of the firms.

Eventually, the confidence intervals give us evidence that all these changes in the average connectedness are significantly different. In times when the confidence sets are not centered, it shows that the mean estimate could also be different. In most phases, the bootstrap is more confident that the true line is below that estimate. While in stressful periods, the bootstrap instead has more confidence of higher levels.

2.6 Concluding Remarks

In the time-series literature, rolling-windows are arguably the most applied time-varying parameter models. We generalized the idea of rolling windows with a semi-parametric estimation approach. We introduced a bootstrap method that applies to the connectedness measure from [Diebold and Yilmaz \(2014\)](#). A bandwidth selection approach complemented the estimation and inference steps. In an extensive simulation study, we showed their efficiency and effectiveness.

The application analyzed a multivariate time series of volatility spillovers on the financial market. In particular, our results highlighted the general trend of connectedness in the financial market. Moreover, we detected a persistent trend between 2002 and 2008, which, we all know, lead to the financial crisis. Comparing networks at different time points, we highlighted the formation of clusters by regions. However, the network estimate on the 30/06/2018 gave evidence of increasing spillovers between regions.

Chapter 3

Estimating Large Dimensional Connectedness Tables

This paper is joint work with Matteo Barigozzi and Christian Brownlees.

3.1 Introduction

In most quantitative research fields, applied problems deal with uncovering dependencies between variables. While dependencies are mostly understood and quantified as contemporaneous covariances and correlations, these linear measures hide directionalities and potential lagged relations. To visualize and summarize dependencies in a more causal way, the recent economics literature pictures such dependencies as spillovers and connectedness. Many papers consult the idea of forecast error variance decompositions (FEVDs) to measure such structures. That is, FEVDs give a coherent understanding of how much the variation of one variable explains the variation of others and, thus, illuminate causal relations with relative contributions. Whereas this concept works for all autocorrelated settings, connectedness measures are most popular in the literature on measuring systemic risk. In particular, [Diebold and Yilmaz \(2014\)](#) successfully reinterpreted this tool and introduced a new way of measuring financial connectedness. Their conception of interpreting FEVDs as networks enables rigorous analyses and insightful visualizations of dependencies, which results in new key figures.

FEVDs condense contemporaneous and lagged dependencies into one connectedness table, which has to be estimated. As for any other econometric method,

accuracy is pivotal. The main equational form of FEVDs follows a vector autoregressive (VAR) structure, such that the estimation of these tables inherits the VAR's estimation uncertainty. While the estimation of VARs has several weaknesses, we are convinced that it suffers particularly in the setup of high dimensionality. That is, the estimation error blows up when the number of variables or nodes (N) approaches the number of time observations (T). This issue is particularly grave because of two reasons. First, most applications seek for time-variation in the parameters and apply rolling window approaches, which requires to reduce the number of observations used in the estimation. Second, to have a pristine unified network interpretation, authors need to ensure that no relevant variables were omitted. Accordingly, they desire a large number of variables.

In this setup, it is widely known that regularization methods for the respective estimators of the coefficient and covariance matrix reduce the uncertainty and the bias. Although the literature on regularization regressions and covariance matrices is rich with the vast availability of different techniques, it is yet unclear how these approaches perform by comparison. Hence, we summarize the most prominent methods and compare them in an extensive simulation study for the setup of sparse relations. Moreover, we deal with the task of regularizing the innovation covariance matrix and how this step affects the overall estimation uncertainty. In short, the simulation study focuses on the clean estimation of VARs and measure their performance on FEVDs. A result of this analysis is that we are not able to elect a clear winner since it appears that their performances are similar overall.

In an empirical analysis, we investigate how spillovers on industrial production of the US between 1972 to 2007 have changed with the Great Moderation. In so doing, we want to answer the question of whether the change in the volatility of the industrial production index is due to a change in some sectors or many. The estimation of FEVDs proves to be challenging, since the split into pre- and post-Great Moderation periods reduces the effective sample size. However, we successfully uncover the spillover networks and the respective dependencies of sectors. We find that there is a significant difference in the average spillover between the pre- and post-Great Moderation periods. An analysis of the distribution of sectoral spillovers highlights that a handful of sectors may be partially responsible for the high volatility before the Great Moderation.

This study connects to the literature on regularization. In particular, we stay in

the strands of regularizing regressions and covariance matrices, respectively. The regularization of regressions has become more popular, because the increasing availability of data made variable-selection more critical. While shrinkage regression methods, such as ridge regression by [Hoerl and Kennard \(1970\)](#), serve as useful tools to decrease the estimation bias, variable selection methods, such as LASSO by [Tibshirani \(1996\)](#), caught more attention in the time series literature. A combination of shrinkage and variable selection is, by the strands of the current research, the most advanced regularization practice applied to time series models. A prominent example is the adaptive elastic-net by [Zou and Zhang \(2009\)](#), which generalizes the ideas of ridge and adaptive LASSO in one approach. Economic applications usually perform these methods in a cross-sectional context. In contrast, time series applications of modern techniques are implemented for elastic-net in a general VAR-case in [Kascha and Trenkler \(2015\)](#) and for the setting of FEVDs in [Demirer et al. \(2017\)](#). The latter is the impetus of this work as it remains unclear how estimation uncertainty affects the overall result of FEVDs.

Complementary to the approach of [Demirer et al. \(2017\)](#), we consider regularizing the innovation covariance matrix likewise. The statistics literature encompasses various methods of regularizing covariance matrix estimates. [Friedman et al. \(2008\)](#) considered a sparse estimate by applying the LASSO penalty to the inverse covariance matrix. Along with this estimator, the optimal shrinkage method by [Ledoit and Wolf \(2004\)](#) reigned as an easy-to-apply regularization for sample covariances. Just recently, thresholding rules like [Bickel and Levina \(2008\)](#), [Rothman et al. \(2009\)](#) and [Cai and Liu \(2011\)](#) introduced the general idea of hard-thresholding, soft-thresholding, and adaptive thresholding, respectively. These methods follow the idea that the sample covariance matrix is subject to calculable estimation errors, which can easily be cut and shrunk. While most thresholding rules apply the same threshold to all entries of the sample covariance, the adaptive approach allows for entry-specific threshold levels.

The rest of the paper organizes as follows. In Section 3.2, we introduce the concepts of FEVDs and give an overview of various regularization methods. We assess these methods with a simulation study in Section 3.3. Section 3.4 applies the regularization of FEVDs to the setup of industrial production to answer the question of sectoral spillovers. Finally, Section 3.5 concludes.

3.2 Methodology

Ultimately, the goal of this study is to estimate large networks in a time series context. For that purpose, we make use of two different methodologies. First, we consider the estimation of networks, which is enabled by the theoretical considerations of forecast error variance decompositions. That is, they explain the amount of spillover between variables. Second, due to the impairment in the estimation of high-dimensional VARs, we deal with regularizations suitable for the estimation of FEVDs. This section sums up generalized FEVDs and introduces suitable regularization methodologies to overcome their estimation uncertainty.

3.2.1 Generalized Forecast Error Variance Decompositions in a Nutshell

As in [Pesaran and Shin \(1998\)](#), we consider a N -dimensional stable VAR(p) process,

$$y_t = \mu + \sum_{i=1}^p A_i y_{t-i} + u_t, \quad \forall t = -p+1, \dots, T, \quad (3.1)$$

where u_t has multivariate normal distribution with covariance matrix Σ_u . Due to the stability assumption, the process can be written in canonical MA representation

$$y_t = \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i}. \quad (3.2)$$

Where the MA parameters are recursively defined as $\Phi_i = \sum_{l=1}^p A_l \Phi_{i-l}$ for $i > 0$ and $\Phi_0 = I_N$. Note that the components of u_t are generally not orthogonal such that structural interpretations are economically meaningless. To this extent [Koop et al. \(1996\)](#), [Pesaran and Shin \(1998\)](#), define the (unscaled) generalized impulse response (GIR) function at horizon h to an impulse δ_j on the j th entry of the reduced form innovation u_t . They do so in integrating out the effect of all the remaining shocks in the shock vector:

$$\text{IR}(h, \delta_j, j) = E(y_{t+h} | u_{j,t} = \delta_j) - E(y_{t+h} | u_{j,t} = 0). \quad (3.3)$$

Under the Gaussian assumption, we can use

$$E[u_t|u_{j,t} = \delta_j] = (\sigma_{1j}, \dots, \sigma_{Nj})\sigma_{jj}^{-1}\delta_j = \Sigma_u e_j \sigma_{jj}^{-1} \delta_j,$$

and substitute it in (3.3). The impulse response function can be rewritten as

$$\text{IR}(h, \delta_j, j) = \Phi_h \Sigma_u e_j \sigma_{jj}^{-1} \delta_j,$$

with e_j as the j th column of the identity matrix. It is customary to set $\delta_j = \sqrt{\sigma_{jj}}$, which yields the scaled generalized impulse response functions $\text{IR}(h, \sqrt{\sigma_{jj}}, j)$. We assemble the scaled IRs in the $(N \times N)$ matrix

$$\Psi^g(h) = [\psi_{ij}^g(h)] = [\text{IR}(h, \sqrt{\sigma_{11}}, 1), \dots, \text{IR}(h, \sqrt{\sigma_{NN}}, N)] = \Phi_h \Sigma_u \text{diag}(\Sigma_u)^{-\frac{1}{2}}, \quad (3.4)$$

where $\text{diag}(M)$ denotes a diagonal matrix with diagonal values of a square matrix M .

Analog to standard impulse response analysis, we gain further insights rewriting y_t as impulse response functions multiplied by an innovation vector. Let $P = \Sigma_u \text{diag}(\Sigma_u)^{-\frac{1}{2}}$ and define the generalized shock u_t^g as

$$u_t^g := P^{-1}u_t \sim \mathcal{N}(\mathbf{0}_{N \times 1}, \Omega),$$

where $\Omega = \text{diag}(\Sigma_u)^{\frac{1}{2}} \Sigma_u^{-1} \text{diag}(\Sigma_u)^{\frac{1}{2}}$. Then it is possible to express (3.2) as

$$y_t = \mu + \sum_{i=0}^{\infty} \Psi^g(i) u_{t-i}^g. \quad (3.5)$$

Henceforth we are allowed to interpret u_t^g as the innovation vector in the generalized impulse response analysis which gets one element shocked with unity and all others set to zero.

Note that the i th forecast error variance at h explained by innovations in variable j is $(\psi_{ij}^g(h))^2$. Then, the effect of a generalized impulse of variable j at time t on the H -step ahead forecast error variance of variable i is

$$MSE[y_{i,t+H}|u_t^g = e_j] = \sum_{h=0}^{H-1} (\psi_{ij}^g(h))^2. \quad (3.6)$$

Also the H -step total forecast error variance from all variables to i is just the mean squared error of variable i :

$$MSE[y_{i,t+H}] = \left(\sum_{h=0}^{H-1} (\Phi_h \Sigma_u \Phi_h') \right)_{ii}. \quad (3.7)$$

Pesaran and Shin (1998) divide (3.6) by (3.7) and thus get a table showing the contributions accounted by innovations in variable j to the H -step forecast error variance of variable i . Like Diebold and Yilmaz (2014), we denote the H -step ahead generalized forecast error variance decompositions $D^{gH} = [d_{ij}^{gH}]$ with entries

$$d_{ij}^{gH} = \frac{MSE[y_{i,t+H} | u_t^g = e_j]}{MSE[y_{i,t+H}]}. \quad (3.8)$$

Note that the numerator implicitly shocks single entries of u_t^g and the denominator shocks single entries of u_t . With the relation given by Ω , we see that the rows of D^{gH} just sum up to one if Σ_u is diagonal. Diebold and Yilmaz (2014) row-normalize these tables for a cleaner network interpretation. However, row-normalization distorts the entries such that the distribution of estimation errors gets more complex. Thus, if not explicitly stated, we do not perform this normalization.

3.2.2 Estimating Large Vector Autoregressions

Naturally, we want a broad set of variables in the economic model, and, thus, we have to deal with high-dimensionality. Recall that (3.8) is a function of the coefficients and innovation covariance matrix in (3.1). Thus, to estimate FEVDs, we have to estimate the unknowns in (3.1). For that purpose, the literature already provides various approaches for estimating large VARs. Since our goal is to improve on estimating connectedness tables in a high-dimensional VAR setup, we compare different regularization approaches. We hope to address the most prominent examples, but we are aware that we might miss out on some others.

First, LASSO-techniques are known to perform well in an autoregressive setup. Those methods, in particular, aim to regularize the coefficient matrix A but do not regularize the covariance matrix Σ_u , which also suffers from estimation uncertainty. From (3.4), it should be evident that a poor estimation of the innovation covariance matrix spoils the overall FEVD. Later in this paper, we point out that,

in particular for the sparsity setup, it is negligent not to regularize the innovation covariance matrix. Hence, we combine regularization methods for the unknown regression coefficient and covariance matrix to achieve the best possible estimate of the FEVD. Eventually, we introduce a new regression model to regularize the long-run dependency.

Regularizing Vector Autoregressive Coefficients

As in [Kascha and Trenkler \(2015\)](#), we first transform the VAR setup such that our coefficient matrix A can be estimated in vector form.

$$y = (Z' \otimes I_N) \beta + u \quad (3.9)$$

where $y = \text{vec}([y_1, \dots, y_T])$, $Z_{t-1}^{0'} = (y'_{t-1}, \dots, y'_{t-p})'$, $Z_{t-1} = (1, Z_{t-1}^{0'})'$, $Z = [Z_0, \dots, Z_{T-1}]$, $\beta = \text{vec}(A)$ and $u = \text{vec}([u_1, \dots, u_T])$. We set $X := (Z' \otimes I_N)$ to obtain the general regression form. For a general regression, the ordinary least squares (unpenalized) estimator reads

$$\hat{\beta}_{OLS} = \underset{\beta}{\operatorname{argmin}} ||y - X\beta||_2^2.$$

Based on this objective function, we aim to regularize the coefficient matrix A .

(Adaptive) elastic net, LASSO and ridge We outline the most general concept following [Zou and Zhang \(2009\)](#)'s adaptive elastic net. This penalized estimator is a compound of the general concepts of elastic net and adaptive LASSO. In particular, it simultaneously shrinks and selects entries in the coefficient matrices and, moreover, has the oracle property, which ensures optimal large sample performance. We adapt

$$\hat{\beta}_{AEnet} = \underset{\beta}{\operatorname{argmin}} \left[||y - X\beta||_2^2 + \lambda \sum_{i=1}^{N^2 p} w_i \left(\alpha |\beta_i| + (1 - \alpha) \frac{1}{2} \beta_i^2 \right) \right], \quad (3.10)$$

where $w_i = |\hat{\beta}_{i,ini}|^{-\gamma}$ is an initial guess with $\gamma > 0$. λ is the penalty term, which has to be chosen, e.g. by cross-validation. Then, λ controls for the strength of the elastic-net penalty. Note that the original paper proposed to only use the weights on the LASSO penalty. However, similar to [Demirer et al. \(2017\)](#) we put the weight

before the shrinkage penalty and use the `glmnet` routine from [Friedman et al. \(2010\)](#).

The regression in (3.10) is an enhanced version of the penalty regression and, thus, generalizes many other concepts. For example, the elastic-net penalty with $\alpha \in (0, 1)$ is the bridge between the LASSO and the ridge estimator and inherits both their desirable properties, e.g., it removes the degeneracies of the LASSO estimator caused by extreme correlations. Moreover, in case of large N and small T , LASSO selects at most NT non-zero entries before it saturates. The mixture with the ridge penalty eradicates this behavior. In a nutshell, the absolute penalty term automatically selects variables while the quadratic penalty shrinks and stabilizes the solution paths; see [Zou \(2006\)](#).

Now, choosing $w_i = 1$ gives the naive elastic-net,¹ $\alpha = 1$ the adaptive LASSO, $\alpha = 0$ the ridge regression, and $\alpha = w_i = 1$ the classical LASSO from [Tibshirani \(1996\)](#). In Section 3.3, we compare performances of the LASSO, ridge and the adaptive elastic net in a simulation of FEVD. If not stated differently, we use $w_i = |\hat{\beta}_{i,OLS}|^{-1}$.

Regularizing the Innovation Covariance Matrix

Recall, that forecast error variance decompositions are functions of the coefficient and covariance estimates. In particular, all impulse response functions include a zero period response, i.e., some decomposition of the covariance matrix. Even though the generalized variance decomposition is not a decomposition of the covariance matrix estimate itself, it solely uses this estimate as an input. Because the covariance matrix estimate suffers similarly under high-dimensionality, we also have to regularize the covariance matrix to obtain the best possible estimate for FEVDs. The literature on the regularization of covariance matrices has originated a variety of methods. In the upcoming section, we introduce three of them. We assume that the innovation series u_t is known. That is, we have to assume that the regularization of the VAR coefficients works reasonably well. For this overview, we change the notation of the innovation series u_t to a general random variable \mathbf{X} .

Adaptive thresholding Thresholding methods are designed for sparse covariance matrices and mimic the idea of shrinking and dropping entries (i.e. selection) of the

¹[Zou and Zhang \(2009\)](#) rescale the naive estimator by $(1 + (1 - \alpha)\lambda/T)$. But similar to [Friedman et al. \(2010\)](#), we drop this distinction.

sample covariance matrix. Assume an N -variate random vector $\mathbf{X} = (X_1, \dots, X_N)'$ with covariance matrix $\Sigma = (\sigma_{ij})_{N \times N}$. Further, assume an iid random sample $\{\mathbf{X}_1, \dots, \mathbf{X}_T\}$ from the distribution of \mathbf{X} . The aim is to estimate the covariance matrix Σ with big N and small T . Start with the sample covariance matrix,

$$\hat{\Sigma} = (\hat{\sigma}_{ij})_{N \times N} := \frac{1}{T-1} \sum_{t=1}^T (\mathbf{X}_t - \bar{\mathbf{X}})(\mathbf{X}_t - \bar{\mathbf{X}})',$$

where $\bar{\mathbf{X}} = T^{-1} \sum_{t=1}^T \mathbf{X}_t$. Further define the variance of the sample covariance's entries as

$$\theta_{ij} := \text{Var}((X_i - \mu_i)(X_j - \mu_j)) = E[(((X_i - \mu_i)(X_j - \mu_j) - \sigma_{ij})^2)], \quad (3.11)$$

with $\mu_i = E[X_i]$. We can now interpret the sparse covariance estimation as a mean vector estimation. That is, an individual entry can be described with

$$\frac{1}{T} \sum_{t=1}^T (X_{t,i} - \mu_i)(X_{t,j} - \mu_j) = \sigma_{ij} + \sqrt{\frac{\theta_{ij}}{T}} z_{ij}, \quad (3.12)$$

with z_{ij} asymptotically standard normal. On this basis, it is straightforward to create an individual threshold for each entry of the covariance matrix. Yet, the variability of an individual entry, θ_{ij} , needs to be estimated with its sample counterpart $\hat{\theta}_{ij} = T^{-1} \sum_{t=1}^T [(X_{t,i} - \bar{X}_i)(X_{t,j} - \bar{X}_j) - \hat{\sigma}_{ij}]^2$. The adaptive (entry dependent) threshold estimator,

$$\hat{\Sigma}^* = (\hat{\sigma}_{ij}^*)_{N \times N} = (s_{\lambda_{ij}}(\hat{\sigma}_{ij}))_{N \times N}, \quad (3.13)$$

allows for different threshold levels for the entries and incorporates the variability of the entries. The function $s_{\lambda}(z)$ then describes the threshold rule applied to the entries of the sample covariance. [Cai and Liu \(2011\)](#) propose the threshold level

$$\lambda_{ij} := \lambda_{ij}(\delta) = \delta \sqrt{\frac{\hat{\theta}_{ij} \log N}{T}}, \quad (3.14)$$

where the regularization parameter $\delta > 0$ can be selected with cross-validation (CV) or, as the authors suggest, be set to 2.

Note now, that the thresholding functions $s_{\lambda}(z)$ can be chosen by the econometrician's needs. For a specific class of thresholding functions, the estimator achieves

optimal convergence and performs better than the universal thresholding estimator, which thresholds all entries with the same threshold. This class has to satisfy the following conditions:

- (i) $|s_\lambda(z)| \leq c|y| \quad \forall \quad z, y \text{ satisfying } |z - y| \leq \lambda \text{ and some } c > 0$
- (i)* $|s_\lambda(z)| \leq |z|$ (shrinkage)
- (ii) $s_\lambda(z) = 0 \text{ for } |z| \leq \lambda$ (thresholding)
- (iii) $|s_\lambda(z) - z| \leq \lambda \quad \forall \quad z \in R$ (limits of shrinkage)

Cai and Liu (2011) conduct analyses for the class satisfying (i),(ii),(iii), but state that it is also possible to adapt this to (i)*. The hard-thresholding rule is ruled out by (i), but other thresholding functions, such as the soft-thresholding² $s_\lambda(z) = \text{sgn}(z)(|z| - \lambda)_+$ and the adaptive-lasso rule $s_\lambda(z) = z(1 - |\lambda/z|^\eta)_+$ with $\eta \geq 1$, are included. Again, λ determines the strength of the penalty. But, unlike before, this parameter is a function described by (3.14) and thus not freely selectable in the adaptive threshold estimator. Here, δ determines the degree of penalization and has to be selected carefully.

Note, that the authors suggest a thresholding rule which applies to all entries in the covariance matrix. In simulations, we experienced zero entries on the diagonal of the thresholded estimator. Since this behavior contradicts the idea of covariances, we treat diagonal values slightly different. Heuristically, whenever a diagonal value $\hat{\sigma}_{ii}$ of the sample covariance is smaller than λ_{ii} , we opt for shrinking this very value to a number bigger than zero. In particular, we employ a shrinkage with

$$s_{\lambda_{ii}}(\hat{\sigma}_{ii}) = \hat{\sigma}_{ii} \sqrt{\frac{\sum_j \hat{\sigma}_{ij}^* \sum_j \hat{\sigma}_{ji}^*}{\sum_j \hat{\sigma}_{ij} \sum_j \hat{\sigma}_{ji}}}, \quad \forall i \quad \text{with} \quad \hat{\sigma}_{ii} \leq \lambda_{ii}.$$

This shrinkage rescales the sample covariance entry with the geometric mean taken over the average shrinkage of the i th row and column.

Ledoit and Wolf In contrast to sparse covariance estimations, we also consider an unblended shrinkage estimator. The estimator of Ledoit and Wolf (2004) is well-conditioned (inverting it does not amplify estimation errors) and more accurate than the sample covariance matrix. In particular, the estimator is the optimal convex

²Cai and Liu (2011) forgot the absolute value of z .

linear combination of the sample covariance and the identity. Whereas optimality is achieved asymptotically with respect to a squared loss function. We consider this regularization method, since it is easy-to-compute and *bona fide*, i.e., it is not affected by the choice of a regularization term and does not require additional knowledge. The Ledoit-Wolf shrinkage estimator reads

$$\hat{\Sigma}^{LW} = \frac{b^2}{d^2} m I_N + \left(1 - \frac{b^2}{d^2}\right) \hat{\Sigma}, \quad (3.15)$$

with

$$\begin{aligned} m &= \text{tr}(\hat{\Sigma})N^{-1}, & d^2 &= \|\hat{\Sigma} - mI_N\|^2, \\ b^2 &= \min \left[T^{-2} \sum_{t=1}^T \|(\mathbf{X}_t - \bar{\mathbf{X}})(\mathbf{X}_t - \bar{\mathbf{X}})' - \hat{\Sigma}\|^2, d^2 \right], \end{aligned}$$

where $\|\cdot\|$ and $\hat{\Sigma}$ denote the Frobenius norm and sample covariance matrix respectively. \mathbf{X}_t and $\bar{\mathbf{X}}$ are defined as above and m is the average of the diagonal values of the sample covariance matrix. In contrast to [Ledoit and Wolf \(2004\)](#), we use the scaled sample covariance matrix, i.e., we divide the sum by $T - 1$ instead of T . However, this scaling should be negligible with the T 's we use.

This estimator is a linear shrinkage estimator, which optimally mixes the 'all-bias no-variance' estimator mI_N with the 'all-variance no-bias' estimator $\hat{\Sigma}$. The term b^2/d^2 automatically assigns more weight to the 'no-variance' estimator if the variance of the sample's second order, measured by b^2 , is large. Thus, similar to the shrinkage in [James and Stein \(1992\)](#), $\hat{\Sigma}^{LW}$ trades between bias and variance of the estimator aiming to minimize the mean squared error. Note that the shrinkage weight can also be chosen with cross-validation.

GLASSO [Friedman et al. \(2008\)](#) propose to estimate sparse graphs by penalizing the inverse covariance matrix. That is, they estimate the inverse covariance with a LASSO penalty. Let $\Theta = \Sigma^{-1}$ and optimize

$$\log \det \Theta - \text{tr}(\hat{\Sigma} \Theta) - \|\Theta\|_1, \quad (3.16)$$

over all nonnegative definite matrices Θ , where $\|\cdot\|_1$ denotes the L_1 norm, i.e., the sum of the absolute values of the entries. As before, $\hat{\Sigma}$ is the sample covariance.

The first two terms describe the multivariate Gaussian likelihood, and the latter is the LASSO penalty, which selects entries in Θ and sets others to zero. The matrix $\hat{\Theta}$, which minimizes this objective function is the respective estimator of the inverse covariance matrix. Since the approach intends to estimate undirected graphical models, they call this estimation the graphical LASSO or, in short, GLASSO.

The fact that the penalization is on the inverse covariance makes this approach particularly appealing in the setup of FEVDs. For example, [Barigozzi and Brownlees \(2013\)](#) highlight the relation of the inverse covariance (concentration matrix) to partial correlations. In particular, if entry $[\Sigma^{-1}]_{ij}$ in the inverse covariance is zero, then variables i and j are conditionally uncorrelated. In economic setups, it is often plausible to rather assume sparsity in the partial correlations than in the overall correlation structure. In other words, it is easier to rule out direct than indirect effects.

Regularizing long-run effects

While most methods assume sparsity in the regression coefficients or the covariance matrix, economic setups sometimes motivate sparsity for a different construct. Namely, some regressions only serve as a tool for a final object of interest. As FEVDs represent the overall dependency with different response times between variables, the idea of sparsity may also apply here. In this subsection, we briefly introduce a method, which models sparsity in the long-run dependency.

We motivate the idea of restricting long-run dependencies in the setup of a VAR(1). This step works without loss of generality since we can always rewrite a VAR(p) in companion form as a VAR(1). Then, for a VAR(1), the moving average matrices $\Phi_k = [\phi_{ij,k}]$ are recursively defined and they read

$$\Phi_k = A^k.$$

If the economic story implies that many variables do not affect each other, we ideally also presume sparsity in the lagged responses. For example, if we assume sparsity in the long-run responses, then we should regularize all Φ_k 's. This regularization proves to be difficult since the matrix potential is a complex function of the coefficient matrix A .

To overcome this issue, we take the forecast error, i.e., the response to a one-

standard-deviation impulse. The long-run (lagged) response of an impulse is

$$\begin{aligned} FE(H) &= \sum_{h=0}^{H-1} \Phi_h = \sum_{h=0}^{H-1} (A^h), \\ \lim_{H \rightarrow \infty} FE(H) &= \sum_{h=0}^{\infty} (A^h) = (I_N - A)^{-1}, \end{aligned} \quad (3.17)$$

where the last equation holds due to the stability condition and is the result of the geometric series. If there is no spillover of one variable to another we assume that the respective entry in $FE(\infty) = (I_N - A)^{-1}$ is zero.

It is evident that zeros in the forecast error most likely imply that the respective forecast error variance, i.e. the element-wise squared version, is also zero. So, in order to impose sparsity on the spillover network, this matrix is a potent candidate for the regularization of FEVDs. Take (3.17) and plug it into the model (3.1),

$$\begin{aligned} y_t &= (I_N - FE(\infty)^{-1})y_{t-1} + u_t, \\ \Delta y_t = y_t - y_{t-1} &= -FE(\infty)^{-1}y_{t-1} + u_t, \\ y_{t-1} &= FE(\infty)(-\Delta y_t) + FE(\infty)u_t. \end{aligned} \quad (3.18)$$

Estimating the last equation with the previously introduced regularization methods permits to regularize $FE(\infty)$ and, with $A = I_N - FE(\infty)^{-1}$, we can back out the autoregressive coefficients. Whereas this version quite likely performs worse if sparsity is on A instead on the FEVDs, it may be better in the contrary case. Henceforth, we denote the LASSO regularization on this target as the geometric regularization.

3.2.3 Data-Driven Choice of λ and δ

Except for the Ledoit-Wolf shrinkage estimator, all estimators above require the choice of a penalization term. In particular, techniques related to elastic-net and graphical LASSO require the choice of λ and the adaptive thresholding rule the choice of δ . Technically, one can also choose the shrinkage weight in Ledoit-Wolf, but since the authors complement their estimator with an optimized weight, this step is unnecessary. To validate the goodness of the estimate for the single penalty terms, machine-learning approaches such as cross-validation (CV) prove functionality. We

briefly discuss how CV can be applied in this context here and suggest slightly modified versions of the ones provided.

In short, CV divides the sample into training and test data. It estimates a trained estimate on the training data and validates its performances on the rest. That is, the quality of the estimate from the training dataset is assessed by its ability to explain the test dataset. The value selected by CV is the version which explains the test sample the best. A common practice is the K -fold CV, which divides the sample into K test samples. For each of these K samples, CV trains the estimator on the rest and then validates it on this sample. A loss-function, such as the mean-squared error or the negative log-likelihood, validates the performance. If K equals the sample size, it becomes the leave-one-out version, which validates the estimator on a single observation.

While CV works well for most setups, it is worth highlighting the predicament in high dimensions. Since its main idea is the division into test and training data, this split decreases the number of observations in the estimation. The estimator faces a different degree of estimation uncertainty and, thus, CV is bent on choosing a stronger penalization term. Consequently, leave-one-out CV is the ideal choice since the training set has a sample size close to T , i.e., $T - 1$, and thus it uses nearly the full sample size to train. On the other hand, a lower K , e.g., 10, might be expected to estimate the prediction error well, since it averages over more diverse training sets (Hastie et al., 2009, Chapter 7.12). In a nutshell, there is a trade-off between a good training sample and a reasonable estimate of the prediction error.

In a typical setup, CV is valid, but for dependent data, such as time-series regressions, it is not always theoretically justified. When dividing the sample into two sets, it is the case that observations at the boundaries show up in both samples, i.e., as the dependent and as the independent variable. Leaving out p observations between the samples circumvents that issue. However, for randomly drawn subsamples, this comes with a significant loss of observations. Thus, it is advisable to use block sampling, similar to the block bootstrap. Moreover, Bergmeir et al. (2018) show that the K -fold version is permissible for a purely autoregressive model with uncorrelated innovations. Consequently, we recommend sticking to the established 10-fold CV with non-random sampling.

Unlike CV for coefficient estimates, there exists no dominantly established version for the covariance estimation. To understand why this is the case, it is worth

spending a look at the loss functions. CV for the coefficient estimates, here β , minimizes the mean-squared error of the implied residuals. This loss-function is straight forward since it is also the minimization target in the estimation step. However, the validation of the covariance matrix proves to be more abstract due to the unobservability of the covariance matrix. [Cai and Liu \(2011\)](#) propose to use the sample covariance estimate of the test sample to validate the goodness of fit. The loss function for the k th test sample is

$$\ell_k(\delta) = \|\widehat{\Sigma}_{\{1:T\}\setminus k}(\delta) - \widehat{\Sigma}_k\|,$$

where the first term of the difference is the trained estimate for the covariance of \mathbf{X} , Σ , with penalization parameter δ . The second term is the sample covariance of the test data sample. This loss-function, however, is by the motivation of this paper imprecise since the sample covariance is a bad estimate of the test sample. Thus, we prefer to validate the trained estimate $\widehat{\Sigma}_{\{1:T\}\setminus k}(\delta)$ on the residuals directly. Naturally, we want to use a similar loss-function as in the optimization of the estimator. For example, we can use a likelihood-based loss-function similar to the GLASSO approach. Unfortunately, we experienced issues with this loss function since the likelihood is not always well defined for the regularized version.

Alternatively, we propose to use a modified version of [Cai and Liu \(2011\)](#). For the k th test sample it is calculated by

$$\ell_k(\delta) = \frac{1}{T_k} \sum_{t \in k} \|\widehat{\Sigma}_{\{1:T\}\setminus k}(\delta) - \mathbf{X}_t \mathbf{X}_t'\|,$$

where T_k denotes the sample size of k . In contrast, this loss-function shows the mean of the distance between the squared observations and the trained estimate.

On a final note, we want to highlight an obvious extension. Since CV selects penalty terms on predefined sample splits, we can also think of plugging in different regularization techniques for the estimators. In case of doubt which estimator is the best, we can let CV select the estimator with the best predictive power.

3.3 Simulation Study

3.3.1 Data Generating Processes

Before we present the simulation results, it is worth spending a thought on the data generating processes (DGPs). In the context of FEVDs, we are convinced that sparsity accompanies with increasing dimension N . However, it remains unclear whether sparsity appears in the VAR coefficients, the innovation covariance matrix, the FEVD or in all of them. Thus, we introduce various data generating processes (DGPs) and hope to address the most relevant issues. This simulation study is limited to VAR(1) models, as higher lag orders only increase the factor of the high-dimensionality problem. The DGPs contain sparse coefficients A and innovation covariances Σ_u .

DGP 1: White noise (in the observations)

$$A = \mathbf{0}_{N \times N}$$

$$\Sigma_u = I_N$$

DGP 2: Diagonal coefficient (auto-correlation with no spillovers)

$$A = 0.5I_N$$

$$\Sigma_u = I_N$$

DGP 3: Diminishing diagonal (only approximately sparse)

$$A \text{ is a banded diagonal matrix, with entries } a_{ij} = 0.5^{|i-j|}$$

$$\Sigma_u = I_N$$

DGP 4: Random graph (without ordering):

A is random sparse matrix. That is, it has entries with probability $P(a_{ij} \neq 0) = \tau$. $\tau = 0.3$ denotes the degree of sparsity. Non-zero entries are temporarily set to one. Then, A is rescaled such that its maximum eigenvalue is in modulus $\mathcal{U}(0.2, 0.5)$.

$\Sigma_u = S^{1/2} D S^{1/2'}$. $S^{1/2}$ is a sparse matrix with the same sparsity structure as A , i.e., $s_{ij}^{1/2} \neq 0 \Leftrightarrow a_{ij} \neq 0$. Its non-zero values are drawn from $\mathcal{U}(-0.5, 0.5)$. D is a diagonal matrix with values drawn from $\mathcal{U}(1, 2)$.

DGP 5: Block-Diagonal FEVD (sparse FEVD)

This DGP mimics a network relation for time series with a sparse FEVD. First, a diagonal-block structure is chosen for $FE(\infty) = (I - A)^{-1}$ with $\sqrt{1/\tau}$ equal-sized quadratic blocks. $\tau = 0.3$ denotes the degree of sparsity. Non-zero entries are temporarily set to one.

$A = I - FE(\infty)^{-1}$ is rescaled such that its maximum eigenvalue is in modulus $\mathcal{U}(0.2, 0.5)$.

$\Sigma_u = S^{1/2} D S^{1/2'}$. $S^{1/2}$ is a sparse matrix with the same sparse block-structure as A , i.e., $s_{ij}^{1/2} \neq 0 \Leftrightarrow a_{ij} \neq 0$. Its non-zero values are drawn from $\mathcal{U}(-0.5, 0.5)$. D is a diagonal matrix with values drawn from $\mathcal{U}(1, 2)$.

Note that *DGP 1*, *2*, and *5* always produce a sparse FEVD and *DGP 3* and *4* do not. Moreover, all but *DGP 4* and *5* are deterministic and have the identity as the covariance matrix. *DGP 4* and *DGP 5* will have 25 different random realizations in the simulation. The methods performances are compared for FEVDs with forecast horizon 10.

3.3.2 Bias, Accuracy and ROC

First, we take a look at the mean of the entries and the norms for regularized and least squares FEVDs. For that, we construct time series with *DGP 5*, $N = 100$ and $T \in [100, 150]$. The analysis compares OLS, adaptive elastic-net (ENET), and ridge. Additionally, we also regularize the covariance matrix with the soft-threshold estimator. Penalty terms are chosen with respect to the best performance (minimization of the Frobenius norm to the true FEVD). Figure 3.1 shows the mean distance, i.e., the entries of $\hat{D}^{gH} - D^{gH}$, its Frobenius norm $\|\hat{D}^{gH} - D^{gH}\|$, and the Frobenius norm of the estimate to the one matrix, i.e., $\|\hat{D}^{gH} - \mathbf{1}_{N \times N}\|$.

The left panel depicts the magnitude of the bias when T goes to N . We see that on average the non-regularized estimator profoundly overestimates entries in the FEVDs. That is, it faces a strong positive bias the closer N is to T . Any regularization already takes away most of the bias. Elastic-net slightly overshoots the regularization and underestimates entries. However, it goes up again, with the additional threshold estimator, because of the division by the standard deviations.

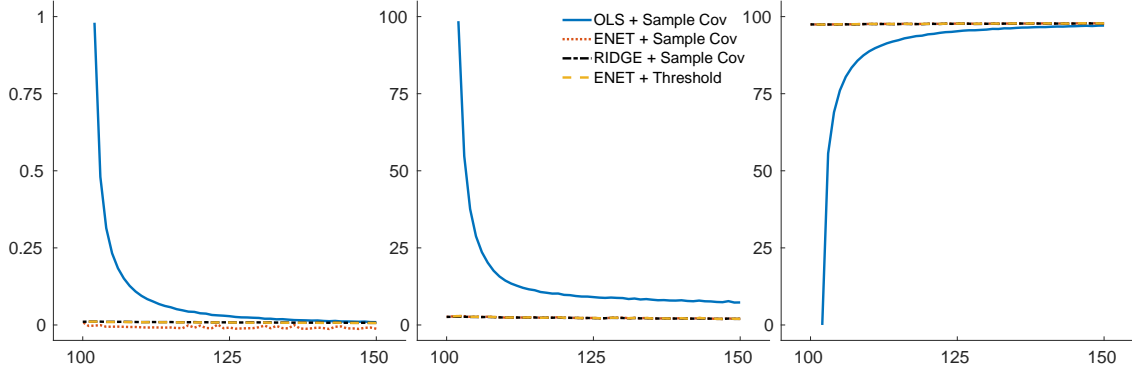


Figure 3.1: Simulation results for 500 Monte-Carlo repetitions of *DGP 5* and $N = 100$. The first two panels show the mean difference and the norm of $\hat{D}^{gH} - D^{gH}$ respectively. The third panel shows the Frobenius norm of the estimate to the one matrix, i.e., $\|\hat{D}^{gH} - \mathbf{1}_{N \times N}\|$. The sample size T is on the x-axis.

The bias vanishes for $T = 150$, which corresponds to 50% more observations than variables.

The center panel shows the accuracy of the estimation plotted as the Frobenius norm. The non-regularized version has to face heavy inaccuracies from the bias. However, when its bias vanishes ($T = 150$), the norm is still significantly higher than the norms of the regularization versions. For all regularization methods, we see a better estimation on average. In particular, we see that the best estimator's accuracy is unaffected by the samples size. Surprisingly, though, the additional regularization of the covariance visually yields no improvement. Finally, the right panel shows the direction of the bias. The norm to the one matrix $\mathbf{1}_{N \times N}$ is nearly zero when N is close to T .

In the context of networks, we explicitly care about the diagnostic ability of the estimator as well. That is, we are interested in how accurate it predicts a non-zero entry in the network matrix. For that purpose, we show the probability of detection, also known as the true-positive rate (TPR), the probability of false alarm, also known as the false-positive rate (FPR), and the receiver operator characteristic (ROC). A value is considered to be true-positive (TP) in case of a hit and false-positive (FP) in case of a false alarm. The TPR is the number of TPs divided by the total positive entries in the actual network. The FPR is the quotient of the number of false alarms and the total amount of values to predict (N^2).

The challenge of this analysis is to show a meaningful graph, even though the

true network is not binary and the estimations are mostly non-zero. To overcome this issue, we set the 50% smallest entries in the estimate to 0 and the rest to 1. The true FEVD constructed by *DGP 5* already contains 50% zero entries, such that we set the rest to 1. For the ROC curve, we vary the cut-off percentile from 0 to 1. ROC then plots the FPR vs. the TPR. Each entry on the curves stands for one cut-off percentile. A curve which lies on the diagonal line is equivalent to a completely random estimate. Figure 3.2 shows the three curves for 500 Monte-Carlo simulations. It contains the TPR, FPR and the ROC respectively in each panel.

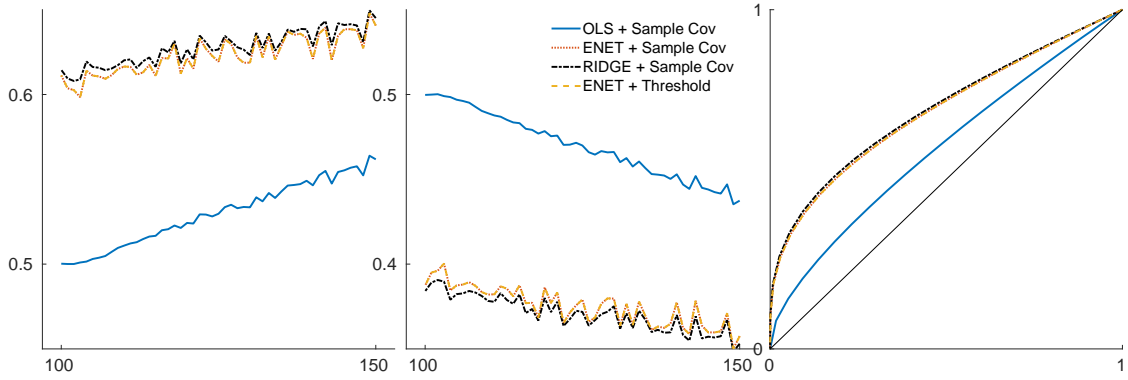


Figure 3.2: Simulation results for 500 Monte-Carlo repetitions of *DGP 5* and $N = 100$. The first two panels show the true positive rate (TPR) and false positive rate (FPR) respectively. The sample size T is on the x-axis. The right panel shows the receiver operator characteristic (ROC) for $T = 150$, which is FPR on the x-axis and TPR on the y-axis. The diagonal thin black line is the equivalent of a random estimate.

It is evident that the unregularized version performs significantly worse than the others. However, it appears to be the case that all regularization methods perform similarly well when it comes to these measures. The first two panels show that the TPR and the FPR of the regularized versions are significantly better than the one of OLS plus the sample covariance. With increasing sample size T , we see that this gap reduces marginally. For $T = 100$, the blue line is exactly 0.5, which means that the estimate is as good as a random matrix. The ROC curve in the last panel confirms that the regularized versions are superior to the unregularized one. Their curves more than double the distance to the diagonal and, thus, they are more precise. In summary, it appears that all regularized versions have similar performance results.³

³Note, however, that the elastic-net and the threshold estimator select entries, i.e., they set some entries to zero. Their advantage lies in the true-negative values. Unfortunately, this quotient only makes sense for estimators who set values to zero. Thus, it does not make sense to have a

When it comes to predicting positive values, the elastic-net performs slightly worse than the ridge regression. Adding the threshold estimator does not change that finding.

3.3.3 Comparison of Different Regularization Methods

This section intends to show the relative performance gain of using regularization methods for the estimation of FEVDs. That is, we compare the accuracy of the various regularized versions to the OLS plus sample covariance case. In particular, we run simulations for the aforementioned DGPs and calculate the Frobenius norm of the regularized case versus the OLS plus sample covariance one: $||\hat{D}_{reg}^{gH} - D^{gH}|| / ||\hat{D}_{OLS}^{gH} - D^{gH}||$ with $H = 10$. We compare the performance gain over $N = \{50, 150, 250\}$ and $T = \{75, 175, 500\}$. We split the simulation into two parts since the estimation is a two-step procedure. Note that OLS breaks down for $N > T$ such that we are not able to calculate any value for these cases.

First, we regularize A paired with the sample covariance for Σ_u . We compare the ridge, LASSO, adaptive elastic-net, and the geometric long-run regularization. The latter is using the LASSO penalty only, such that it performs variable selection. Penalty parameters λ are chosen such that they minimize the respective norm $||\hat{D}_{reg}^{gH} - D^{gH}||$. Thus, the values show the best possible performance gain.

Table 3.1 contains the simulation results for the regularization of A for 500 Monte-Carlo repetitions. First, it is evident that there is an overall big efficiency gain by regularizing. *DPG 1-3* all show a similar gain for all estimators. The best performance gain for all DGPs and regularization methods is at $N = 250$. However, for $N = 50$ and $T = 500$, we still observe a remarkable efficiency gain. That is, the norm to the true value shrinks to 22 – 25% for *DGP 1-3* and 56 – 67% for *DGP 4-5*.

Surprisingly, all traditional regularization methods perform similarly well. It is unclear which one performs the best since the differences in performances are marginal. If at all, the performance of the ridge estimator is better by a tiny margin for all DGPs. This result is in line with the finding from the ROC curve. In contrast, the geometric regularization performs slightly worse for *DGP 3* and *4*, since they only have approximative sparsity in the FEVD. Unsurprisingly, it performs better for *DGP 5*, which has a sparse long-run Forecast Error matrix. The performance

comparison to other estimators here.

| DPG1 | | | | | | | | | | | | |
|------------------|-------|-------|-------|-------|-------|-------|-------------|-------|-------|-----------|-------|-------|
| $N \backslash T$ | ridge | | | LASSO | | | elastic-net | | | geometric | | |
| | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 |
| 50 | 21.5% | 22.6% | 22.6% | 21.6% | 22.7% | 22.6% | 21.6% | 22.7% | 22.6% | 22.4% | 23.2% | 22.8% |
| 150 | | 12.3% | 13.8% | | 12.2% | 13.8% | | 12.3% | 13.8% | | 12.6% | 13.9% |
| 250 | | | 10.7% | | | 10.8% | | | 10.8% | | | 10.9% |
| DPG2 | | | | | | | | | | | | |
| $N \backslash T$ | ridge | | | LASSO | | | elastic-net | | | geometric | | |
| | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 |
| 50 | 21.5% | 22.6% | 22.6% | 21.6% | 22.7% | 22.6% | 21.6% | 22.7% | 22.6% | 22.4% | 23.2% | 22.8% |
| 150 | | 12.3% | 13.8% | | 12.2% | 13.8% | | 12.3% | 13.8% | | 12.6% | 13.9% |
| 250 | | | 10.7% | | | 10.8% | | | 10.8% | | | 10.9% |
| DPG3 | | | | | | | | | | | | |
| $N \backslash T$ | ridge | | | LASSO | | | elastic-net | | | geometric | | |
| | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 |
| 50 | 23.5% | 23.2% | 23.9% | 24.2% | 23.2% | 26.4% | 24.2% | 23.2% | 26.4% | 29.9% | 30.0% | 35.4% |
| 150 | | 13.7% | 14.5% | | 13.7% | 14.3% | | 13.8% | 14.3% | | 17.2% | 19.2% |
| 250 | | | 11.7% | | | 11.2% | | | 11.2% | | | 15.2% |
| DPG4 | | | | | | | | | | | | |
| $N \backslash T$ | ridge | | | LASSO | | | elastic-net | | | geometric | | |
| | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 |
| 50 | 35.2% | 47.7% | 65.4% | 35.1% | 47.6% | 65.4% | 34.9% | 47.4% | 64.9% | 37.0% | 49.2% | 67.5% |
| 150 | | 19.8% | 30.9% | | 19.7% | 30.9% | | 19.7% | 30.9% | | 20.3% | 31.5% |
| 250 | | | 20.4% | | | 20.4% | | | 20.4% | | | 20.7% |
| DPG5 | | | | | | | | | | | | |
| $N \backslash T$ | ridge | | | LASSO | | | elastic-net | | | geometric | | |
| | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 |
| 50 | 33.9% | 41.2% | 56.2% | 34.1% | 41.9% | 58.6% | 34.0% | 42.5% | 58.7% | 32.8% | 42.4% | 63.4% |
| 150 | | 20.4% | 27.7% | | 20.2% | 27.7% | | 20.9% | 27.5% | | 19.2% | 27.7% |
| 250 | | | 19.7% | | | 19.5% | | | 20.1% | | | 18.7% |

Table 3.1: Simulation results for the regularization of A paired with the sample covariance. $\|\hat{D}_{reg}^{gH} - D^{gH}\| / \|\hat{D}_{OLS}^{gH} - D^{gH}\|$ are shown for different N and T with 500 Monte-Carlo repetitions. *DGP 4* and *DGP 5* have 25 different random realizations of A and Σ_u .

gain over the traditional regularization techniques appears negligible.

As a second step, we compare regularization methods for Σ_u . That is, we calculate residuals with the adaptive elastic-net and estimate the FEVD with the (regularized) estimate of Σ_u . The adaptive elastic-net penalty is the one, which minimizes the Frobenius norm to the true parameter A . We use the soft-thresholding rule for the adaptive Threshold estimator.

Table 3.2 shows simulations with different regularizations. The first estimator, the naked sample covariance, sets the benchmark for the performance of the covariance regularization. Again, we measure the performance of the regularization methods with the norm of the estimated FEVD to the true one. The most salient result is the 0% norm for *DGP 1-2*. In particular, the FEVDs get perfectly estimated, since the covariance matrix in the DGPs is the identity. The strongest penalization finds this matrix. A shrinkage towards a diagonal matrix as in the Ledoit-Wolf

| DPG1 | | | | | | | | | | | | |
|------------------|------------|-------|-------|-----------|-------|-------|-------------|-------|-------|--------|-------|-------|
| $N \backslash T$ | Sample-Cov | | | Threshold | | | Ledoit-Wolf | | | GLASSO | | |
| | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 |
| 50 | 21.6% | 22.5% | 22.6% | 0% | 0% | 0% | 0.1% | 0% | 0% | 0% | 0% | 0% |
| 150 | | 12.3% | 13.8% | | 0% | 0% | | 0% | 0% | | 0% | 0% |
| 250 | | | 10.8% | | | 0% | | | 0% | | | 0% |
| DPG2 | | | | | | | | | | | | |
| $N \backslash T$ | Sample-Cov | | | Threshold | | | Ledoit-Wolf | | | GLASSO | | |
| | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 |
| 50 | 21.6% | 22.5% | 22.6% | 0% | 0% | 0% | 0.1% | 0% | 0% | 0% | 0% | 0% |
| 150 | | 12.3% | 13.8% | | 0% | 0% | | 0% | 0% | | 0% | 0% |
| 250 | | | 10.8% | | | 0% | | | 0% | | | 0% |
| DPG3 | | | | | | | | | | | | |
| $N \backslash T$ | Sample-Cov | | | Threshold | | | Ledoit-Wolf | | | GLASSO | | |
| | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 |
| 50 | 25.5% | 23.1% | 27.5% | 4.1% | 8.9% | 19.9% | 4.2% | 9.0% | 20.0% | 4.1% | 8.9% | 19.9% |
| 150 | | 14.6% | 14.8% | | 2.9% | 7.7% | | 3% | 7.7% | | 2.9% | 7.7% |
| 250 | | | 11.5% | | | 4.6% | | | 4.6% | | | 4.6% |
| DPG4 | | | | | | | | | | | | |
| $N \backslash T$ | Sample-Cov | | | Threshold | | | Ledoit-Wolf | | | GLASSO | | |
| | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 |
| 50 | 35.3% | 48.0% | 65.6% | 27.0% | 43.9% | 64.6% | 22.9% | 39.5% | 61.8% | 24.0% | 41.0% | 65.0% |
| 150 | | 19.8% | 30.9% | | 13.5% | 28.0% | | 11.1% | 24.4% | | 11.8% | 25.7% |
| 250 | | | 20.5% | | | 16.6% | | | 14.0% | | | 14.8% |
| DPG5 | | | | | | | | | | | | |
| $N \backslash T$ | Sample-Cov | | | Threshold | | | Ledoit-Wolf | | | GLASSO | | |
| | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 | 75 | 175 | 500 |
| 50 | 36.8% | 43.0% | 59.8% | 29.7% | 41.1% | 60.7% | 29.5% | 47.5% | 69.1% | 28.1% | 40.9% | 74.2% |
| 150 | | 21.7% | 28.0% | | 15.9% | 26.4% | | 14.9% | 29.6% | | 14.5% | 26.2% |
| 250 | | | 20.1% | | | 17.6% | | | 18.7% | | | 17.0% |

Table 3.2: Simulation results for the regularization of Σ_u paired with the best performing adaptive elastic-net estimator. $\|\hat{D}_{reg}^{gH} - D^{gH}\|/\|\hat{D}_{OLS}^{gH} - D^{gH}\|$ are shown for different N and T with 500 Monte-Carlo repetitions. *DGP 4* and *DGP 5* have 25 different random realizations of A and Σ_u .

estimator perfectly estimates the identity here. Similarly, for the Threshold and GLASSO estimator, which both select entries in the covariance matrix. However, the perfect estimations are also due to the perfect estimation of the coefficient matrix in the first step.

In *DGP 3*, elastic-net is not able to perfectly estimate this matrix since it is not sparse. Thus, the estimation of the overall FEVD is not completely perfect. However, it appears that sometimes the methods perfectly estimate the covariance matrix, which is again the identity. The performance is similar for all of the regularization methods. Different performances for the methods are observable for non-diagonal covariance matrices in *DGP 4-5*. A significantly increased performance is observable for all combinations of N and T except for $N = 50$ and $T = 500$. For this case, the sample covariance performs reasonably well, such that its regulariza-

tion does not enhance estimation accuracy. Again, we do not find any particular difference in the performance between these estimators and, thus, we cannot give a clear advice which one to use.

In summary, we find that the regularization of the single steps leads to a significant performance gain in the estimation of FEVDs. If we have to recommend the use of one specific regularization technique, we recommend using the ridge combined with the Ledoit-Wolf estimator. Both of them slightly outperform the others and are also easy-to-apply. However, since we already have to find the penalization term, it suggests itself to use cross-validation techniques to validate the performances of these methods for specific data.

3.4 Empirical Application

In the mid-1980s, the Great Moderation exhibited a change in the growth of real gross domestic product growth, industrial production, monthly payroll employment, and unemployment. For instance, the Federal Reserve Board's Index of Industrial Production (IP) shows a significant decrease in volatility after 1984. [Foerster et al. \(2011\)](#) investigate this break concerning sectoral and common shocks. Since the aggregate IP index is a weighted average of IPs across sectors, the volatility of the sectors should average out. The authors search for roots of this puzzle. They find that common shocks cause most of the break in IP's volatility. Consequently, a decrease in the volatility of aggregate shocks explains why the IP index is less volatile post-1984. At one go, they also find that the contribution of common shocks decreased.

The question arises what the descent of the decrease in aggregate shock volatility is. In particular, it would be interesting to see if a handful of large sectors lowered their links to others and thus reducing common shock volatility and pairwise correlation. For sufficient insights into these links, we want to know who affects whom. Thus, this question links to the comprehension of spillovers. This section estimates FEVD tables to study sectoral spillovers of shocks. These connectedness tables then reveal the share of a sector's variation explained by others. We examine pre-and post-1984 periods in greater detail to give a fair comparison of how spillover have changed. In a nutshell, the purpose of this application is to quantify the spillovers of sectors and uncover changes in the propagation of shocks.

3.4.1 Data

As in Foerster et al. (2011), we use sectoral data on IP throughout 1972-2008. This data spans up to $N = 138$ sectors, which corresponds to the six-digit classification of the North American Industry Classification System (NAICS). The sectoral indices are available on a monthly basis. Since the pre- and post-Great Moderation periods have different sample sizes, they face distinct degrees of estimation uncertainty. Hence, we split the whole sample into three equally sized subsamples with $T = 144$ months each. The subsamples are from 03/1972 to 02/1984, 03/1984 to 02/1996, and 03/1996 to 02/2008. The boundary between the first and the second sample marks the Great-Moderation. For simplicity, we label these samples as 1972-1983, 1984-1995, and 1996-2007, respectively.

Let $IP_{i,t}$ denote the value of industrial production of sector i at date t . We take monthly growth rates and annualize the respective percentage points, $g_{i,t} = 1200 \times \ln(IP_{i,t}/IP_{i,t-1})$. The aggregate level of IP growth is the weighted average over the sectors, $g_t = \sum_{i=1}^N w_{i,t} g_{i,t}$, with given weights $w_{i,t}$. Figure 3.3 plots the growth rate of IP on an aggregate level. The first subsample, from 1972 to 1983, coincides with the pre-Great Moderation period and the other two to the post-Great Moderation. It is evident, that average monthly volatility of aggregate IP diminished with the Great Moderation and stayed fairly constant thereafter.

3.4.2 Results

We interpret the spillover constituent in the data as a VAR(1) model, which coincides roughly with the ARMA(1,1) model in Foerster et al. (2011). They implemented connectedness with the help of quarterly input-output tables. In contrast, we are now implementing the connectedness in our model solely via the autocorrelation component of $g_{i,t}$ for monthly data. The higher frequency allows setting a forecast horizon of one quarter, i.e., three months, which is the same horizon as for the shock covariance matrix of the quarterly data. This connection between monthly and quarterly frequency gives insights into the contagion within a quarter. In particular, Foerster et al. (2011)'s average pairwise correlations and aggregate shocks may be

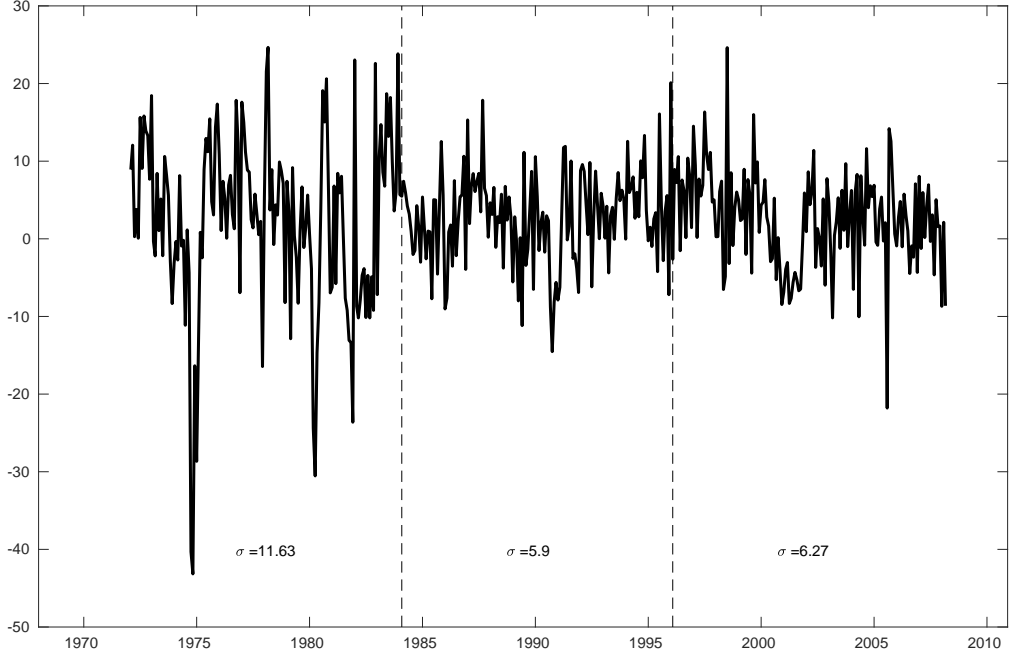


Figure 3.3: Growth rates of monthly industrial production with annual rates.

better understood if we break up the three-month volatility. Now,

$$\begin{aligned}
 y_t &= \mu + Ay_{t-i} + u_t, \quad \forall t = 0, \dots, T, \\
 y_t &= [g_{1,t}, \dots, g_{N,t}]' \\
 u_t &\sim \mathcal{N}(0, \Sigma_u)
 \end{aligned}$$

states the base regression. We regularize A and Σ_u with the techniques mentioned earlier. Since the simulations didn't hint a clear winner, we validate those on the data. Namely, we run a 10-fold cross-validation to select the best performing regularization. This step also includes the selection of α in the elastic-net and the penalty terms λ and δ . Then, forecast error variance decompositions are calculated with forecast horizon $H = 3$. Recall that we consider the generalized version of FEVDs, i.e., the shocks are not idiosyncratic and can be correlated. Effects which propagate via the autocorrelation matrix A increase correlations between the sectors.

We row-normalize D^{g^H} to show the percentage contribution to volatility. The row-normalized matrix is denoted as \tilde{D}^{g^H} with entries $\tilde{d}_{ij}^{g^H}$. Additionally, we present key figures related to the network literature. In particular, we use the same measures as [Diebold and Yilmaz \(2014\)](#): From-, To-, and Average Connectedness. These

measures are defined respectively as the row and column sums without the diagonal entries and the average row sum.

$$\begin{aligned}
C_{i \leftarrow \cdot} \left(\tilde{D}^{gH} \right) &= \sum_{j \neq i} \tilde{d}_{ij}^{gH}, & (\text{From-Connectedness to } i) \\
C_{\cdot \leftarrow j} \left(\tilde{D}^{gH} \right) &= \sum_{i \neq j} \tilde{d}_{ij}^{gH}, & (\text{To-Connectedness from } j) \\
C \left(\tilde{D}^{gH} \right) &= \frac{1}{N} \sum_i \sum_{j \neq i} \tilde{d}_{ij}^{gH}. & (\text{Average Connectedness})
\end{aligned}$$

The first two measures are sector-specific measures. The latter sums up the overall explanatory power of connectedness, which gives us an idea of how much variation is explained on average by spillovers and not directly by shocks. Note now, that we are mainly interested in the distribution of the outgoing spillovers of the sectors measured by the To-Connectedness. Precisely, if a handful of large sectors had a high To-Connectedness, the volatility of those sectors' IP would not average out in the aggregate IP.

First, we validate the performances of the different methods on the data set. We use the three-digit sectoral disaggregation with 88 sectors. Figure 3.4 and Figure 3.5 show the mean squared errors for the regularization of A and Σ , respectively. While we compare different methods for the estimation of Σ , the estimation of A is technically only done with the adaptive elastic-net. That is, we compare the minimal mean squared errors for different values of α . Recall, that $\alpha = 1$ is the adaptive LASSO and $\alpha = 0$ is the (adaptive) ridge estimator. However, we find it desirable to have the good behavior of, both, the LASSO and ridge at all times in the estimation. Thus, α is allowed to take values in $[0.025, 0.975]$.

Figure 3.4 compares the best performing λ 's for different values of α . The two curves use different initial guesses for the weight $w_i = |\hat{\beta}_{i,ini}|^{-1}$. OLS corresponds to the approach of [Demirer et al. \(2017\)](#), and the elastic-net (selected by 10-fold CV, $\alpha = 0.5$, and $w_i = 1$) corresponds to the original idea of [Zou and Zhang \(2009\)](#). Two results are evident here. First, the minimal MSE appears to be a monotonously increasing function in $\alpha \in (0, 1)$. Consequently, we select $\alpha = 0.025$ for all periods. Second, the elastic-net initial guess dominantly gives lower MSE. Hence, CV suggests an estimator with penalty close to ridge and the elastic-net as an initial guess.

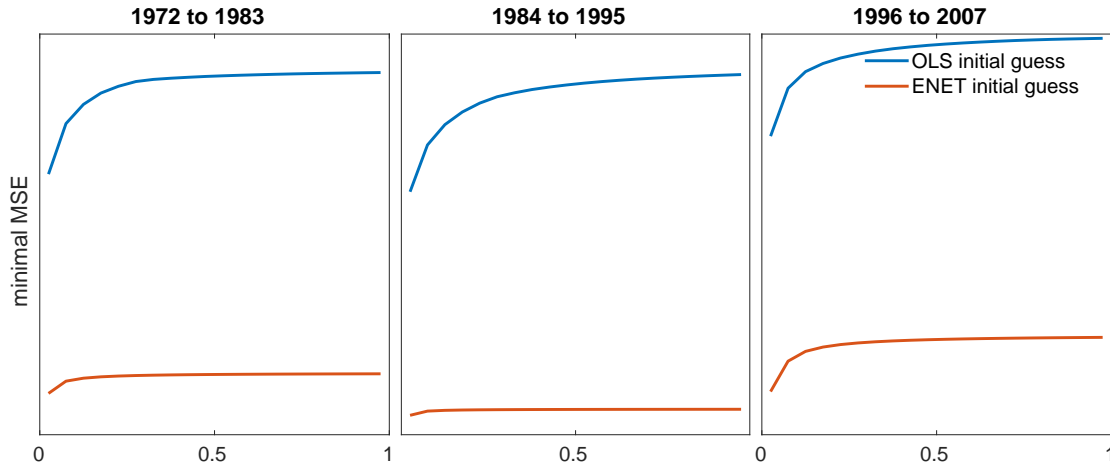


Figure 3.4: 10-fold cross-validation results for the tuning parameter α in the adaptive elastic-net. The blue curve shows for different α 's the minimal MSE with \hat{A}_{OLS} as the initial estimate in the weights of the individual penalties for different α s. The red curve shows the same for \hat{A}_{ENET} , as suggested by [Zou and Zhang \(2009\)](#). Both curves show the results of the first 10-fold CV over values of λ .

In the second step, CV for the covariance estimators includes the soft-thresholding estimator, the adaptive-lasso thresholding estimator, the GLASSO estimator, the bona fide Ledoit-Wolf estimator, and the Ledoit-Wolf estimator with manually selectable shrinkage weight. Figure 3.5 shows their performance with different penalization and shrinkage parameters on the abscissa.

CV selects the soft-thresholding estimator for all periods. This version thresholds the values entry-wise and appears to outperform the others slightly. The other entry-wise regularization, the adaptive-lasso thresholding, comes seconds and, the manual Ledoit-Wolf estimator third. Even though CV suggests the manual Ledoit-Wolf, the bona-fide estimator performs well. In fact, it suggests using this estimator before using any GLASSO regularization. In particular, it might be unrealistic to set partial correlations between two series to zero for this data set. However, it is not advisable to discard GLASSO for other applications.

Table 3.3 shows some selected summary statistics. Two results are evident in this table. First, the non-regularized version has a significantly higher Average Connectedness and also stays remarkably constant over the periods. That is, they do not detect any change with the Great Moderation. In contrast, the regularized versions can capture a difference between the pre- and post-Great Moderation periods. Moreover, for all dimensions, the connectedness level bounced back. Sec-

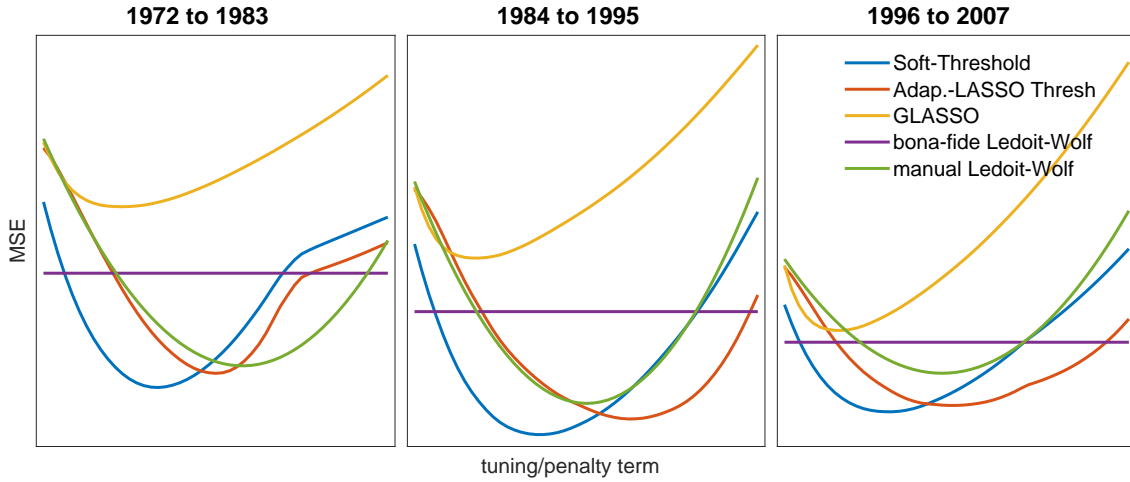


Figure 3.5: 10-fold cross-validation results for the tuning parameter of different covariance regularization methods. Values of α are on the x-axis and the minimal mean squared error on the y-axis. The two thresholding estimators tune δ in equation (3.14), the GLASSO tunes λ in equation (3.16), and the manual Ledoit-Wolf the shrinkage term.

ond, the average results of the regularized versions are robust over different levels of disaggregation. The non-regularized estimators, however, have higher Average Connectedness when we increase dimensions. This observation clearly emphasizes the need of regularizations in this context and also exemplifies the bias of FEVDs (see Section 3.3).

Eventually, we want to emphasize the degree of freedom of the estimators in Table 3.3. As expected, CV decreases the average degree of freedom in the coefficient estimates \hat{A}_{reg} with higher dimensions since it has to deal with more estimation uncertainty. The four- and six-digit levels have nearly always 0% non-zero entries in the coefficient matrix. We do not see any gain in using more dimensions in this analysis, and thus, we further analyze the three-digit level to include reasonably many lagged responses. For this level of disaggregation, we use the elastic-net estimator as an initial guess, since it outperformed the least squares version in the CV. In the analysis, we observe average spillovers of 39.5%, 20%, and 26.7% for the three periods respectively.

To give a synoptic view of the spillovers, we summarize the estimated connectness tables in Figure 3.6. It shows the tables as a network graph, for which the force-layout arranges the nodes. That is, nodes appear closer in the graph if they are more connected to each other. The size of the node relates to the respective From-

| | | 1972-1983 | | 1984-1995 | | 1996-2007 | |
|-----------------------------|---------------------------------------|-------------|-------|-------------|-------|-------------|-------|
| | | $C(D^{gH})$ | df | $C(D^{gH})$ | df | $C(D^{gH})$ | df |
| Three-digit (88 sectors) | non-regularized | 83.1% | | 82.2% | | 81.2% | |
| | reg. $\hat{A}_{ini} = \hat{A}_{OLS}$ | 43.5% | 10.7% | 24.9% | 6.5% | 27.4% | 8.1% |
| | reg. $\hat{A}_{ini} = \hat{A}_{ENET}$ | 39.5% | 3.9% | 20.0% | 0.6% | 26.7% | 2.2% |
| Four-digit (117 sectors) | non-regularized | 92.9% | | 92.3% | | 92.0% | |
| | reg. $\hat{A}_{ini} = \hat{A}_{OLS}$ | 40.3% | 0.01% | 23.2% | 0.2% | 30.9% | 10.5% |
| Six-digit (138 sectors) | non-regularized | 98.4% | | 98.2% | | 98.3% | |
| | reg. $\hat{A}_{ini} = \hat{A}_{OLS}$ | 42.5% | 0% | 28.1% | 0.02% | 33.7% | 0% |

Table 3.3: Summary of the estimation results for different levels of sectoral disaggregation. The columns labeled $C(D^{gH})$ show the estimated Average Connectedness. The columns labeled df show the sparsity level or percentage degree of freedom of the autoregression coefficient A .

Connectedness of the sector. The colors illustrate the To-Connectedness. Finally, we highlight the sector with the highest To-Connectedness.

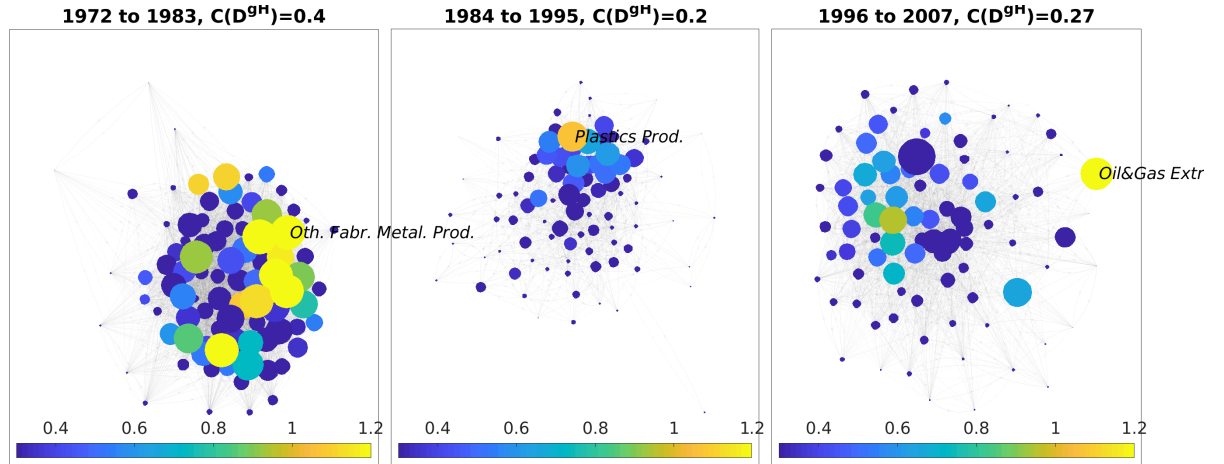


Figure 3.6: Connectedness network for the respective periods. The size of the node relates to the respective From-Connectedness of the sector. The colors depict the To-Connectedness. The sector with the highest To-Connectedness is labeled.

From eye-balling, it is evident that the network has changed significantly after 1984. Whereas the pre-Great Moderation period shows a closely connected graph with many powerful nodes in the center, the two consecutive periods have more widespread graphs with only a couple of sound nodes. Clearly, this also shows in the Average Connectedness. Furthermore, it appears that the distribution changed

for the To- and From-Connectedness. As mentioned earlier, a plausible explanation for the high volatility in the aggregate index is that a handful of sectors spilled a lot of volatility before the Great Moderation. We further investigate this hypothesis with the distribution of the To- and From-Connectedness.

First, we check if sectors with bigger weights in the aggregate IP index receive more spillovers than sectors with small weights. Figure 3.7 shows the distribution of the From-Connectedness measure over the sectors. The red makers highlight the 50% most significant sectors by weights.

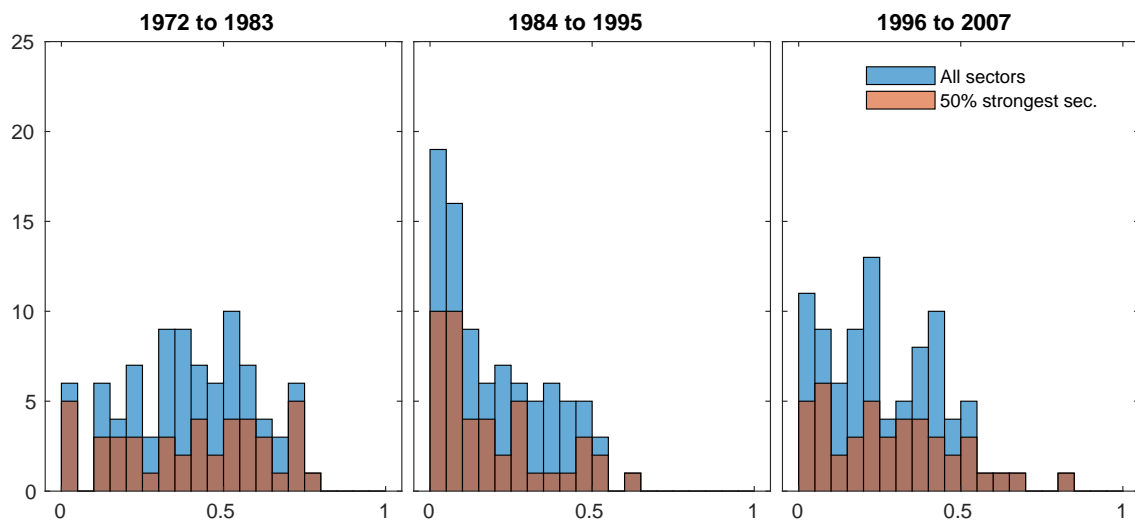


Figure 3.7: Histogram of the From-Connectedness of the 88 three-digit level sectors. The 50% strongest sectors by weight are highlighted in red.

Perhaps surprisingly, the distribution of weights appears to be independent of the From-Connectedness. Also, the overall distributions reflect the same picture as the network graphs. A higher level of incoming spillovers indicates that other sectors' volatility has relatively more effect on the volatility of a single sector. This finding, however, is no clear evidence for the generally higher volatility. Hence, the distribution of the spillover receiving measure is of little importance in the explanation of the higher volatility of aggregate shocks.

Although higher spillovers do not cause higher volatility at an aggregate level, the roots may be in the distribution of the opposing side. Outgoing spillovers, here measured as To-Connectedness, determine how much a single sector's volatility explains the volatility of others. For example, if a handful of sectors has a very high

level of To-Connectedness, their volatility determines the volatility of other sectors more. Consequently, the volatility of the aggregate IP index is indirectly affected more by a couple of sectors, and their shocks do not average out.

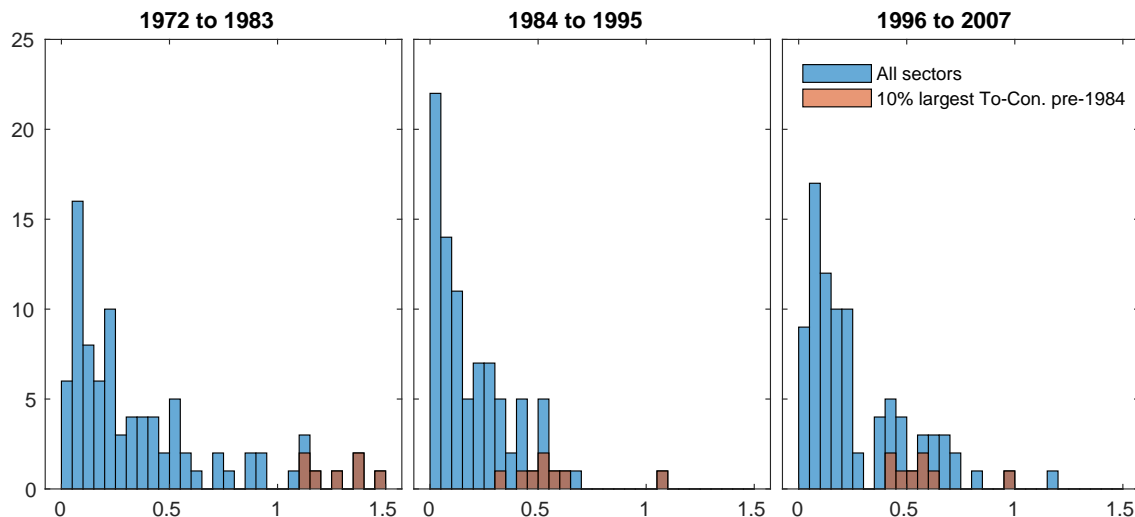


Figure 3.8: Histogram of the To-Connectedness of the 88 three-digit level sectors. The 10% largest sectors by To-Connectedness in the pre-Great Moderation period are highlighted in red.

Figure 3.8 shows the histogram of the To-Connectedness for the respective periods. The first panel shows a clear heavy tail on the right of the distribution. Roughly 15% of the sectors have a To-Connectedness higher than 1. These sectors explain on average more than 100% of the volatility of other sectors. So their volatility shows up on average more than twofold in the aggregate index. The heavy tail disappears mostly after the Great Moderation, i.e., the distributions of the consecutive periods have significantly fewer values in the right tail. Moreover, the sectors with the 10% largest To-Connectedness before the Great Moderation mostly have reduced the values to around 0.5 in the last panel. Comparing only the first and the last panel, they have a similar distribution except for the heavy right tail. A similar result is observable in the cumulative distribution functions (see Appendix Figure C.1). This result supports the hypothesis that the Great Moderation is only a temporary phenomenon. To test this hypothesis, though, further research is needed.

In summary, the results give insights into the change observable during the Great Moderation. Foerster et al. (2011) emphasize that the decrease in aggregate volatility is not due to some sectors, but instead rooted in the change of aggregate shocks,

i.e., shocks to multiple sectors at once. However, our analysis of propagation of shocks within a three-month period highlights that just a handful of sectors may have been responsible for the change in quarterly aggregate shocks. That is, these sectors had higher spillovers such that they increased the volatility of the aggregate shocks.

3.5 Conclusions

In this study, we investigated the estimation of high-dimensional vector autoregressive models. In a simulation study, we compared different regularization methods for the coefficient and the covariance matrix. We evaluated performances on the estimation of forecast error variance decompositions, and we found that the regularization of both matrices leads to better estimations. Since there was no clear cut winner in the simulation, we suggested validating the estimators with cross-validation. In an application on US industrial production, we were able to uncover the change of spillovers at the Great Moderation. Specifically, we aimed to answer the question of whether a handful of sectors was responsible for the decrease in aggregate volatility. We found that a couple of sectors had a particular high outgoing spillover before the Great-Moderation. After the Great-Moderation these outgoing spillovers were unmatched.

Appendix A

Appendix to Chapter 1

A.1 Propositions and proofs

Proposition 1. *Let $\Sigma_\tau^{-1} = A'_0 B_\tau^{-2} A_0$ and $\dot{\Sigma}_\tau^{-1} = -2A'_0 B_\tau^{-3} \dot{B}_\tau A_0$ be the existing inverse of a $(N \times N)$ covariance matrix and its differential, with A_0 having a unit diagonal and $B_\tau = \text{diag}(b_1, \dots, b_N)$ and $\dot{B}_\tau = \text{diag}(\dot{b}_1, \dots, \dot{b}_N)$. Then, the $(N \times N)$ matrix A_0 is unique and the $N \times N$ matrices B_τ and \dot{B}_τ are unique up to sign reversal of its entries if $\dot{b}_i/b_i \neq \dot{b}_j/b_j$ for all $i \neq j \in 1, \dots, N$.*

Proof: The proof follows directly from [Lanne et al. \(2010\)](#). The following will be a copy of their proof in a notation aligned with this paper. Define $S = A_0 B_\tau^{-1}$ and $\Lambda = -2\dot{B}_\tau B_\tau^{-1}$. Then there exists $\tilde{S} = SQ$ leading to the same observations where Q is an orthogonal matrix with $QQ' = I_N$, such that

$$\Sigma_\tau^{-1} = SS' = SQQ'S' \quad (\text{A.1})$$

$$\Rightarrow \dot{\Sigma}_\tau^{-1} = S\Lambda S' = SQ\Lambda Q'S' \quad (\text{A.2})$$

Since B_τ is diagonal and A_0 has unit diagonal, A_0 is uniquely identified if S is. To prove the proposition we need to rule out all orthogonal matrices except the Q 's which are diagonal with only 1's and -1 's.

From (A.2) follows that $Q\Lambda Q' = \Lambda$ and therefore $Q\Lambda = \Lambda Q$ and $(-2\dot{b}_i/b_i)q_{ij} = (-2\dot{b}_i/b_i)q_{ji}$. Hence, Q is diagonal. Because all eigenvalues of a diagonal real orthogonal matrix are ± 1 and the diagonal entries of a diagonal matrix are its eigenvalues, Q is a diagonal matrix with 1's and -1 's. \square

A.2 Complementary Figures

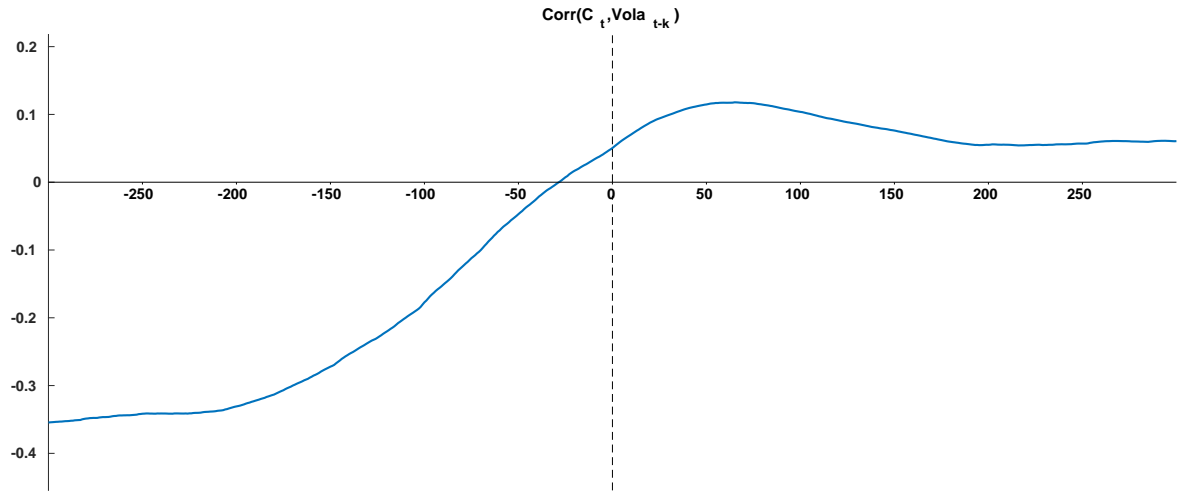


Figure A.1: Lead-lag correlation: $\text{corr}(C(G_t), \overline{\text{vola}_{t-k}})$. The average volatility at time t is taken over the firms.

Appendix B

Appendix to Chapter 2

B.1 Proofs

Proof of Lemma 2.2.1

Proof. From continuity of $\Sigma(\tau)$ on $[0, 1]$ and its extension to \mathbb{R} , we get that $\sup_{t \in \mathbb{Z}} E(u_t u_t') = \sup_{t \in \mathbb{Z}} \Sigma < \infty$ holds. Then, with the companion form matrix (1.15) (without intercept) we follow the same arguments from the proof of [Bühlmann and Künsch \(1995\)](#) and get the claimed result; see also Example 2.3(ii) in [Dahlhaus \(2000\)](#). \square

Proof of Lemma 2.2.2

Proof. Since the u_t 's are independent, we have a Cramér representation

$$\Sigma^{-1/2}(\frac{t}{T})u_t = \int_{-\pi}^{\pi} \frac{1}{\sqrt{2\pi}} \exp(i\lambda t) d\xi(\lambda), \quad t \in \mathbb{Z}$$

with $\xi(\lambda)$ as in Definition 2.2.1. Together with Lemma 2.2.1, this gives

$$\begin{aligned} y_{t,T} &= \mu(\frac{t}{T}) + \sum_{j=1}^{\infty} \Psi_{t,T,j} u_{t-j} = \mu(\frac{t}{T}) + \sum_{j=1}^{\infty} \Psi_{t,T,j} \Sigma^{1/2}(\frac{t-j}{T}) \Sigma^{-1/2}(\frac{t-j}{T}) u_{t-j} \\ &= \mu(\frac{t}{T}) + \int_{-\pi}^{\pi} \exp(i\lambda t) A_{t,T}^0(\lambda) d\xi(\lambda), \end{aligned}$$

where

$$A_{t,T}^0(\lambda) = \frac{1}{\sqrt{2\pi}} \left(\sum_{j=0}^{\infty} \Psi_{t,T,j} \Sigma^{1/2} \left(\frac{t-j}{T} \right) \exp(-i\lambda j) \right)$$

Then, following the proof of Theorem 2.3(i) in Dahlhaus gives the claimed result with $\mathbf{A}(u, \lambda)$ as in (2.8) such that

$$\sup_{t,\lambda} \left| \mathbf{A}_{t,T}^0(\lambda) - \mathbf{A}\left(\frac{t}{T}, \lambda\right) \right|_1 \leq \frac{N^2 B}{T}. \quad (\text{B.1.1})$$

See also Example 2.3 in Dahlhaus (2000). From the uniform stability condition, we have that for all $\tau \in [0, 1]$ (and actually all $\tau \in \mathbb{R}$)

$$I_D = \left(\sum_{j=0}^{\infty} \Psi_j(\tau) \exp(-ij\lambda) \right) \left(I_D - \sum_{k=1}^p A_k(\tau) \exp(-ik\lambda) \right),$$

where the smoothness properties of the tvVAR coefficient curves $A_i(\cdot)$, $i = 1, \dots, p$ transfer to the tvVMA coefficient curves $\Psi_j(\cdot)$, $j \in \mathbb{N}_0$. Further, we get from (B.1.1) that

$$\begin{aligned} \sup_{t \in \mathbb{Z}, t \leq T} |y_{t,T} - \tilde{y}_{t,T}|_1 &= \sup_t \left| \int_{-\pi}^{\pi} \exp(i\lambda t) \left(A_{t,T}^0(\lambda) - A\left(\frac{t}{T}, \lambda\right) \right) d\xi(\lambda) \right| \\ &\leq \sup_{t,\lambda} \left| A_{t,T}^0(\lambda) - A\left(\frac{t}{T}, \lambda\right) \right| \int_{-\pi}^{\pi} d\xi(\lambda) \\ &= O_P\left(\frac{1}{T}\right). \end{aligned}$$

From the uniform stability condition (2.5), we obtain that for all $\tau \in (0, 1]$ that

$$|\Psi_j(\tau)|_1 \leq C(\tau) \rho^j(\tau),$$

where the constants $C(\tau) < \infty$ and $\rho(\tau) \in (0, 1)$ can be chosen such that $\sup_u C(\tau) < \infty$ and $\rho = \sup_{\tau \in [0,1]} \rho(\tau) < 1$ hold. This immediately gives

$$\sup_{\tau} \sum_{j=0}^{\infty} |\Psi_j(\tau)|_1 < \infty.$$

□

B.2 Derivation of local linear least square tvVAR

Define

$$\begin{aligned}
\mathbf{y} &= \text{vec}(y_{1,T}, \dots, y_{T,T}), & (NT \times 1) \\
Z_t &= (1, (\text{vec}(y_{t,T}, \dots, y_{t-p+1,T}))')', & ((Np+1) \times 1) \\
Z^{(0)} &= (Z_0, \dots, Z_{T-1}), & ((Np+1) \times T) \\
\boldsymbol{\beta}(\tau) &= \text{vec}(\nu(\tau), A_1(\tau), \dots, A_p(\tau)), & ((N^2p+N) \times 1) \\
\mathbf{w} &= \text{vec}(u_1, \dots, u_T), & (NT \times 1) \\
\boldsymbol{\sigma}(\tau) &= \text{vech}(\Sigma(\tau)), & (N(N+1) \times 1) \\
\mathbf{B}(\tau) &= (\nu(\tau), A_1(\tau), \dots, A_p(\tau)), & (N \times (Np+1)) \\
\mathbf{Y} &= [y_{1,T}, \dots, y_{T,T}], & (N \times T) \\
Z' &= [Z^{(0)'} | D_\tau Z^{(0)'}], & (T \times 2(Np+1)) \\
D_\tau &= \text{diag}(\frac{1}{T} - \tau, \dots, \frac{T}{T} - \tau), & (T \times T) \\
\dot{\boldsymbol{\beta}}(\tau) &= \text{vec}(\dot{\nu}(\tau), \dot{A}_1(\tau), \dots, \dot{A}_p(\tau)), & ((N^2p+N) \times 1) \\
\dot{\mathbf{B}}(\tau) &= (\dot{\nu}(\tau), \dot{A}_1(\tau), \dots, \dot{A}_p(\tau)), & (N \times (Np+1)) \\
W_\tau &= \text{diag}(K_h(\frac{1}{T} - \tau), \dots, K_h(\frac{T}{T} - \tau)), & (T \times T) \\
\mathbf{W} &= [u_1, \dots, u_T], & (N \times T)
\end{aligned}$$

where ‘vec’ denotes the column stacking operator. The ‘vech’-operator is defined to stack columnwise the elements on and below the main diagonal of a square matrix. Further, we denote $A \otimes B = (a_{ij}B)_{ij}$ the Kronecker product of matrices $A = (a_{ij})$ and $B = (b_{ij})$. The kernel function is assumed to be symmetric and non-negative.

Now, we derive the local linear least-squares estimator for the VAR coefficient curves $A_i(\cdot)$, $i = 1, \dots, p$ and the intercept curve $\nu(\cdot)$. After having established suitable estimators for these quantities, we will make use of corresponding residuals to estimate the innovations’ variance covariance curve $\Sigma(\cdot)$ as well. As the VAR coefficient curves are assumed to be sufficiently smooth (continuously differentiable on $(0, 1)$), we can use a Taylor expansion of $A_i(\frac{t}{T})$ around u to get

$$A_i(\frac{t}{T}) = A_i(\tau) + \dot{A}_i(\tilde{\tau})(\frac{t}{T} - \tau),$$

where $\tilde{\tau}$ lies between $\frac{t}{T}$ and τ . \dot{A}_i denotes the (matrix-valued entry-wise derivative) of the function A_i with respect to time. For each $\tau \in (0, 1)$, this exercise together with the recursively defined tvVAR(p) model equations in (2.4) leads to the following local linear least-squares estimator for the VAR coefficient curves, the intercept curves plus their first derivatives. Precisely, we define

$$\begin{aligned} [\hat{\mathbf{B}}(\tau)|\hat{\dot{\mathbf{B}}}(\tau)] &= \left[\hat{\nu}_\tau, \hat{A}_1(\tau), \dots, \hat{A}_p(\tau), \hat{\dot{\nu}}_\tau, \hat{\dot{A}}_1(\tau), \dots, \hat{\dot{A}}_p(\tau) \right] \\ &= \arg \min_{(\nu, A_k, \dot{\nu}, \dot{A}_k), k=1, \dots, p} \sum_{t=1}^T |y_{t,T} - \hat{y}_\tau(t)|_2^2 K_h\left(\frac{t}{T} - \tau\right), \\ \hat{y}_\tau(t) &= \left(\nu + \dot{\nu}\left(\frac{t}{T} - \tau\right) \right) - \sum_{k=1}^p \left(A_k + \dot{A}_k\left(\frac{t}{T} - \tau\right) \right) y_{t-k,T} \end{aligned}$$

where $|\cdot|_2$ denotes the Euclidean norm. In detail, we assume to have p pre-sample values, such that the latter equation is fulfilled for all t . However, for the sake of notation, we avoid to mention pre-sample values. In vectorized form, the above minimization problem transfers to

$$\begin{aligned} \begin{pmatrix} \hat{\boldsymbol{\beta}}(\tau) \\ \hat{\dot{\boldsymbol{\beta}}}(\tau) \end{pmatrix} &= \text{vec} \left(\left[\hat{\nu}_\tau, \hat{A}_1(\tau), \dots, \hat{A}_p(\tau), \hat{\dot{\nu}}_\tau, \hat{\dot{A}}_1(\tau), \dots, \hat{\dot{A}}_p(\tau) \right] \right) \\ &= \arg \min_{\text{vec}([\mathbf{B}|\dot{\mathbf{B}}])} \left(\mathbf{y} - (Z' \otimes I_N) \text{vec}([\mathbf{B}|\dot{\mathbf{B}}]) \right)' (W \otimes I_N) \\ &\quad * \left(\mathbf{y} - (Z' \otimes I_N) \text{vec}([\mathbf{B}|\dot{\mathbf{B}}]) \right), \end{aligned}$$

where $\mathbf{B} = [\nu, A_1, \dots, A_p]$, and $\dot{\mathbf{B}}$ respectively. Solving the minimization problem above by standard arguments, we get the local linear LS estimator

$$\begin{aligned} \begin{pmatrix} \hat{\boldsymbol{\beta}}(\tau) \\ \hat{\dot{\boldsymbol{\beta}}}(\tau) \end{pmatrix} &= ((Z' \otimes I_N)' (W \otimes I_N) (Z' \otimes I_N))^{-1} (Z' \otimes I_N)' (W \otimes I_N) \mathbf{y} \\ &= (((ZWZ')^{-1}ZW) \otimes I_N) \mathbf{y}, \end{aligned}$$

where standard rules for Kronecker products have been used to get the last identity. As we will be interested in the coefficient and intercept curves itself and not in the first derivatives, we introduce the $(Np+1 \times 2(Np+1))$ matrix $E'_0 = [I_{Np+1}|O_{Np+1}]$, where I_N and O_N denote the $(N \times N)$ identity and zero matrices, respectively. This

matrix E_0 allows us to write $\widehat{\beta}(\tau)$ as

$$\begin{aligned}\widehat{\beta}(\tau) &= (E'_0 \otimes I_N) ((Z' \otimes I_N)' (W \otimes I_N) (Z' \otimes I_N))^{-1} (Z' \otimes I_N)' (W \otimes I_N) \mathbf{y} \\ &= (E'_0 \otimes I_N) (((ZWZ')^{-1}) \otimes I_N) (ZW \otimes I_N) \mathbf{y} \\ &= ((E'_0(ZWZ')^{-1}ZW) \otimes I_N) \mathbf{y}.\end{aligned}\tag{B.2.1}$$

Again by using the usual rules for the Kronecker product, we end up with an alternative representation that avoids vectorization, that is,

$$\widehat{\mathbf{B}}(\tau) = YWZ'(ZWZ')^{-1}E_0.\tag{B.2.2}$$

This estimator is the same estimator as in (2.12).

B.3 Residual-based estimation of $\Sigma(\tau)$

After having defined the local linear LS estimator $\widehat{\mathbf{B}}(\tau)$ for the VAR coefficient matrices and the intercept in (2.2), we can make use of it to get the residuals $\widehat{u}_1, \dots, \widehat{u}_T$

$$\begin{aligned}\widehat{u}_t &= y_{t,T} - \widehat{\mathbf{B}}(\tau)Z_{t-1} \\ &= y_{t,T} - \widehat{\nu}\left(\frac{t}{T}\right) - \widehat{A}_1\left(\frac{t}{T}\right)y_{t-1,T} - \dots - \widehat{A}_p\left(\frac{t}{T}\right)y_{t-p,T}\end{aligned}$$

and to define the local linear LS estimator for Σ Precisely, following the same ideas as above, we define

$$[\widehat{\Sigma}(\tau)|\widehat{\Sigma}(\tau)] = \arg \min_{\Sigma, \dot{\Sigma}} \sum_{t=1}^T \left| \widehat{u}_t \widehat{u}'_t - \left(\Sigma + \dot{\Sigma}\left(\frac{t}{T} - \tau\right) \right) \right|_2^2 K_h\left(\frac{t}{T} - \tau\right),$$

In vectorized form, the above minimization problem transfers to

$$\begin{aligned}\begin{pmatrix} \text{vec}(\widehat{\Sigma}(\tau)) \\ \text{vec}(\dot{\widehat{\Sigma}}(\tau)) \end{pmatrix} &= \arg \min_{\text{vec}(\Sigma), \text{vec}(\dot{\Sigma})} \left(\mathbf{v} - (Z'_\Sigma \otimes I_{N^2}) \text{vec}([\Sigma|\dot{\Sigma}]) \right)' (W \otimes I_{N^2}) \\ &\quad * \left(\mathbf{v} - (Z'_\Sigma \otimes I_{N^2}) \text{vec}([\Sigma|\dot{\Sigma}]) \right)\end{aligned}$$

where we set $v_t = \text{vec}(\widehat{u}_t \widehat{u}_t')$ and we used the notation

$$\begin{aligned}\mathbf{v} &= \text{vec}(v_1, \dots, v_T) \quad (N^2 T \times 1), \quad Z'_\Sigma = [\mathbf{1}_T | D_\tau \mathbf{1}_T], \quad (T \times 2) \\ \mathbf{V} &= (v_1, \dots, v_T) \quad (N^2 \times T)\end{aligned}$$

where $\mathbf{1}_T$ is the T -dimensional vector of ones. Note that Z'_Σ corresponds to Z' defined above, for the special case of fitting a tvVAR(0) model that estimates only the intercept curve. That is, when only the first column of $Z^{(0)'} is used which is nothing else but $\mathbf{1}_T$. Solving the minimization problem above by standard arguments, we get the local linear LS estimator$

$$\begin{aligned}\begin{pmatrix} \text{vec}(\widehat{\Sigma}(\tau)) \\ \text{vec}(\widehat{\dot{\Sigma}}(\tau)) \end{pmatrix} &= ((Z'_\Sigma \otimes I_{N^2})' (W \otimes I_{N^2}) (Z'_\Sigma \otimes I_N))^{-1} (Z'_\Sigma \otimes I_{N^2})' (W \otimes I_{N^2}) \mathbf{v} \\ &= (((Z_\Sigma W Z'_\Sigma)^{-1} Z_\Sigma W) \otimes I_{N^2}) \mathbf{v}.\end{aligned}$$

As we will be interested in the innovations' covariance matrix itself and not in the first derivatives and as this curve will be symmetric, we make use of the $N(N+1)/2 \times N^2$ elimination matrix L_N which transforms $\text{vec}(\widehat{\Sigma}(\tau))$ into $\text{vech}(\widehat{\Sigma}(\tau))$; see e.g. Lütkepohl (2006). This matrix L_N allows us to write $\widehat{\boldsymbol{\sigma}}(\tau)$ as

$$\begin{aligned}\widehat{\boldsymbol{\sigma}}(\tau) &= L_N ((1, 0) \otimes I_{N^2}) ((Z'_\Sigma \otimes I_{N^2})' (W \otimes I_{N^2}) (Z'_\Sigma \otimes I_N))^{-1} \\ &\quad * (Z'_\Sigma \otimes I_{N^2})' (W \otimes I_{N^2}) \mathbf{v} \\ &= ((1, 0) \otimes L_N) (((Z_\Sigma W Z'_\Sigma)^{-1}) \otimes I_{N^2}) (Z_\Sigma W \otimes I_{N^2}) \mathbf{v} \\ &= ((1, 0) \otimes L_N) (((Z_\Sigma W Z'_\Sigma)^{-1} Z_\Sigma W) \otimes I_{N^2}) \mathbf{v} \tag{B.3.3}\end{aligned}$$

$$\begin{aligned}&= ((1, 0) (Z_\Sigma W Z'_\Sigma)^{-1} Z_\Sigma W \otimes L_N) \mathbf{v} \\ &= ((1, 0) (Z_\Sigma W Z'_\Sigma)^{-1} Z_\Sigma W \otimes I_{N(N+1)/2}) (I_T \otimes L_N) \mathbf{v} \\ &= ((1, 0) (Z_\Sigma W Z'_\Sigma)^{-1} Z_\Sigma W \otimes I_{N(N+1)/2}) \widetilde{\mathbf{v}}, \tag{B.3.4}\end{aligned}$$

where we set $\widetilde{v}_t = \text{vech}(\widehat{u}_t \widehat{u}_t')$ and we used the notation

$$\widetilde{\mathbf{v}} = \text{vec}(\widetilde{v}_1, \dots, \widetilde{v}_T) \quad (TN(N+1)/2 \times 1), \quad \widetilde{\mathbf{V}} = (\widetilde{v}_1, \dots, \widetilde{v}_T) \quad (N(N+1)/2 \times T)$$

Here it worth noting that (B.3.4) is exactly the local linear LS estimator obtained for $\text{vech}(\Sigma(\tau))$ if we start the derivation with the minimization problem by stacking $\text{vech}(\widehat{u}_t \widehat{u}_t')$ instead of $\text{vec}(\widehat{u}_t \widehat{u}_t')$ and using $\widetilde{\mathbf{v}}$ instead of \mathbf{v} .

By dropping the elimination matrix and using the usual rules for the Kronecker product, we end up with an alternative representation that avoids vectorization, that is,

$$\begin{aligned}
& \text{vec}(\widehat{\Sigma}(\tau)) \\
&= \mathbf{V} W Z'_\Sigma (Z_\Sigma W Z'_\Sigma)^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
&= \text{vec} \left(\mathbf{V}_{N \times NT} (W \otimes I_N) (Z'_\Sigma \otimes I_N) ((Z_\Sigma W Z'_\Sigma)^{-1} \otimes I_N) \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes I_N \right) \right) \\
&= \text{vec} \left(\mathbf{V}_{N \times NT} \left(W Z'_\Sigma (Z_\Sigma W Z'_\Sigma)^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes I_N \right) \right), \tag{B.3.5}
\end{aligned}$$

where

$$\mathbf{V}_{N \times NT} = (\widehat{u}_1 \widehat{u}'_1 | \cdots | \widehat{u}_T \widehat{u}'_T) \quad (N \times NT).$$

Finally, this leads to the matrix representation of the local linear LS estimator of $\Sigma(\tau)$, i.e.

$$\widehat{\Sigma}(\tau) = \mathbf{V}_{N \times NT} \left(W Z'_\Sigma (Z_\Sigma W Z'_\Sigma)^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes I_N \right). \tag{B.3.6}$$

B.4 Simulation: Data Generation

In the simulation, we generate stable parameter series for dimension N with the following base matrices:

$$\bar{A} = \mathbf{1}_{(N \times N)} * 0.2 + \text{diag}(\mathcal{U}(0.2, 0.4), \dots, \mathcal{U}(0.2, 0.4))$$

where $\text{diag}(\cdot)$ denotes a diagonal matrix with the respective entries.

$$SE(A) = \mathbf{1}_{(N \times N)} * 0.2$$

$$\bar{\Sigma}^{0.5} = \mathbf{1}_{(N \times N)} * 0.1 + I_{(N \times N)} * 0.4$$

For $t = 1, \dots, T$ and $i, j = 1, \dots, N$, parameter matrix entries are then generated by the following two functions:

$$\begin{aligned}
a_{ij}(t) &= \bar{A}_{ij} + SE(A)_{ij} \left[\alpha_{ij} \sin \left(\beta_{ij} 2\pi + \gamma_{ij} 2\pi \frac{1-t}{T} \right) \right] \\
\alpha_{ij} &\sim \mathcal{N}(0, 1) && \text{(Magnitude)} \\
\beta_{ij} &\sim \mathcal{U}(0, 1) && \text{(Shift)} \\
\gamma_{ij} &\sim \mathcal{U}(0.3, 0.5) + \delta && \text{(Frequency)} \\
\delta &\sim \mathcal{U}(0, 1) && \text{(random frequency)}
\end{aligned}$$

All resulting parameter series are checked for stability. That is, at all time points, the eigenvalue of the parameter matrix $[a_{ij}(t)]$ has to be smaller than one in modulus. If not, the process is generated again.

The entries of the square root innovation covariance matrix $\Sigma^{0.5} = [\sigma_{ij}^{0.5}(t)]$ are generated similarly.

$$\begin{aligned}
\sigma_{ij}^{0.5}(t) &= \bar{\Sigma}_{ij}^{0.5} \left[2 + \sin \left(\beta'_{ij} 2\pi + \gamma'_{ij} 2\pi \frac{1-t}{T} \right) \right] \\
\beta'_{ij} &\sim \mathcal{U}(0, 1) && \text{(Sigma Shift)} \\
\gamma'_{ij} &\sim \mathcal{N}(0, 1) + 2\delta' && \text{(Sigma Frequency)} \\
\delta' &\sim \mathcal{U}(0, 1) && \text{(Sigma rand freq)}
\end{aligned}$$

Note that both δ and δ' are randomly drawn for all i, j , while the other randomly drawn parameters are drawn for the specific entry.

The data generating process is a VAR(1) process with coefficient series $A(t) = [a_{ij}(t)]$ and innovation covariance series $\Sigma^{0.5}(t) = [\sigma_{ij}^{0.5}(t)]$.

$$\begin{aligned}
Y_t &= A(t)Y_{t-1} + \Sigma^{0.5}(t)\epsilon_t, && \forall t = 1, \dots, T \\
Y_0 &= \bar{A}\bar{\Sigma}^{0.5}\epsilon_t \\
\epsilon_t &\sim \mathcal{N}(0, I_N)
\end{aligned}$$

Additional Simulation Results

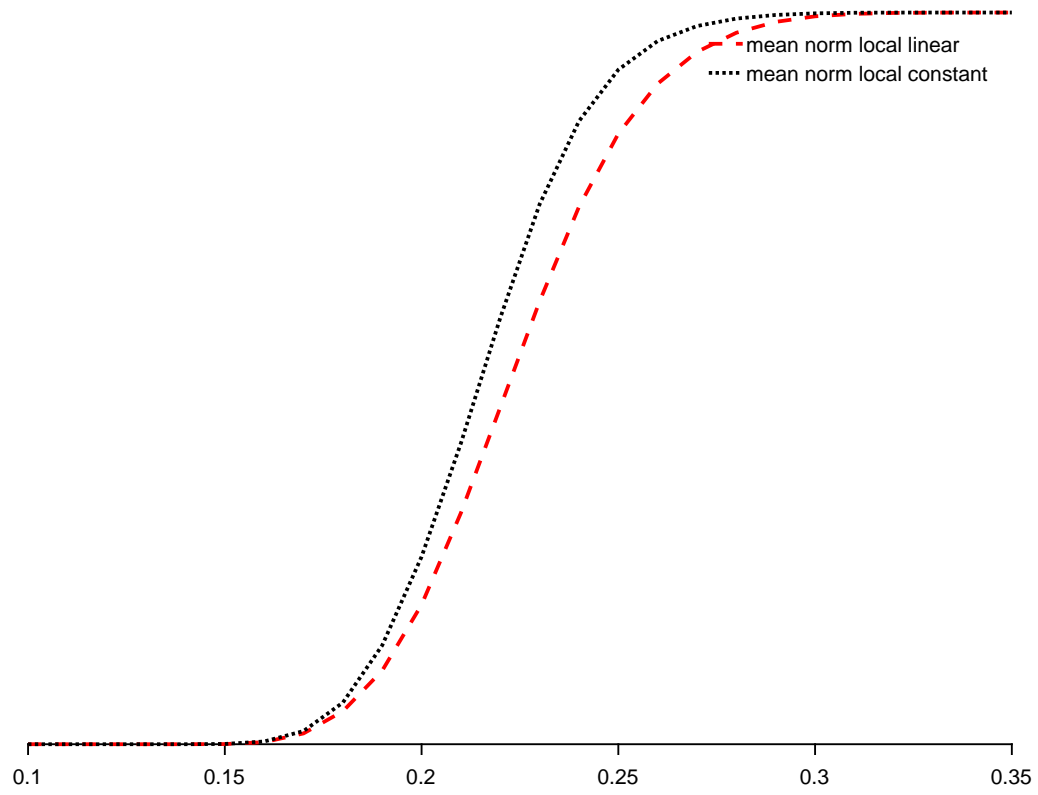


Figure B.1: Monte Carlo Simulation with 500 repetitions of A_t . Cumulative distribution function of the norm of the local-linear estimator (2.12) (red) and the local-constant estimator minus the true parameter.

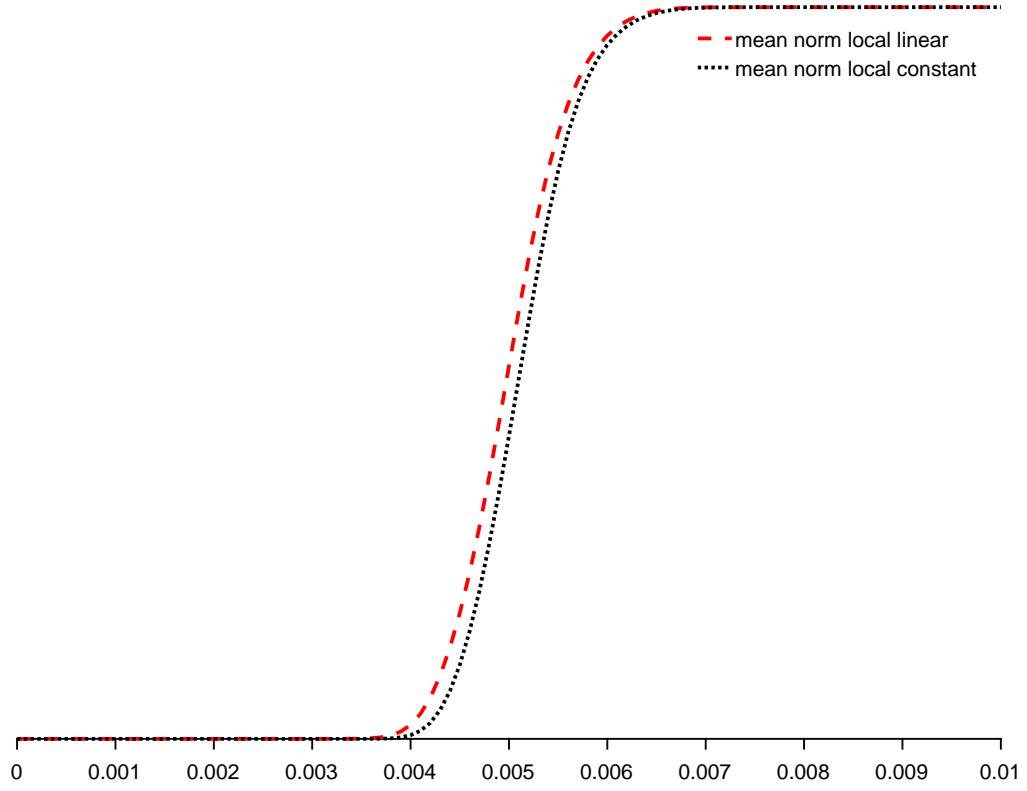


Figure B.2: Monte Carlo Simulation with 500 repetitions of Σ_t . Cumulative distribution function of the norm of the local-linear estimator (2.12) (red) and the local-constant estimator minus the true parameter.

Appendix C

Appendix to Chapter 3

C.1 Complementary Graphs

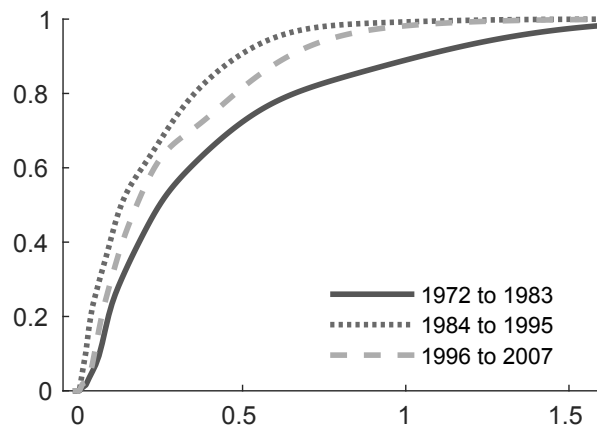


Figure C.1: Estimated CDF of the To-Connectedness of the 88 three-digit level sectors.

Bibliography

- Daron Acemoglu, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. Systemic risk and stability in financial networks. *American Economic Review*, 105(2):564–608, 2015.
- Viral V Acharya, Lasse H Pedersen, Thomas Philippon, and Matthew Richardson. Measuring systemic risk. *The Review of Financial Studies*, 30(1):2–47, 2017.
- Tobias Adrian and Markus K Brunnermeier. Covar. Technical report, National Bureau of Economic Research, 2011.
- Torben G Andersen, Tim Bollerslev, Francis X Diebold, and Paul Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625, 2003.
- Matteo Barigozzi and Christian Brownlees. Nets: Network estimation for time series. *Journal of Applied Econometrics*, 2013.
- Christoph Bergmeir, Rob J Hyndman, and Bonsoo Koo. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120:70–83, 2018.
- Peter J Bickel and Elizaveta Levina. Regularized estimation of large covariance matrices. *The Annals of Statistics*, pages 199–227, 2008.
- Monica Billio, Mila Getmansky, Andrew W Lo, and Loriana Pelizzon. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559, 2012.
- Olivier J Blanchard and Danny Quah. The dynamic effects of aggregate demand and supply disturbances, 1988.
- Mark Bognanni. A class of time-varying parameter structural vars for inference under exact or set identification. *Federal Reserve Bank of Cleveland Working Papers*, (WP 18-11), 2018.

- Tim Bollerslev and Jeffrey M Wooldridge. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11(2):143–172, 1992.
- Christian T Brownlees and Robert Engle. Volatility, correlation and tails for systemic risk measurement. *Available at SSRN*, 1611229, 2012.
- Markus K Brunnermeier. Deciphering the liquidity and credit crunch 2007-2008. *Journal of Economic perspectives*, 23(1):77–100, 2009.
- Peter Bühlmann and Hans R Künsch. The blockwise bootstrap for general parameters of a stationary time series. *Scandinavian Journal of Statistics*, pages 35–54, 1995.
- Tony Cai and Weidong Liu. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106(494):672–684, 2011.
- Rainer Dahlhaus. On the kullback-leibler information divergence of locally stationary processes. *Stochastic Processes and their Applications*, 62(1):139–168, 1996.
- Rainer Dahlhaus. A likelihood approximation for locally stationary processes. *The Annals of Statistics*, 28(6):1762–1794, 2000.
- Rainer Dahlhaus. Locally stationary processes. In *Handbook of statistics*, volume 30, pages 351–413. Elsevier, 2012.
- Rainer Dahlhaus, Wolfgang Polonik, et al. Nonparametric quasi-maximum likelihood estimation for gaussian locally stationary processes. *The Annals of Statistics*, 34(6):2790–2824, 2006.
- Rainer Dahlhaus et al. Fitting time series models to nonstationary processes. *The Annals of Statistics*, 25(1):1–37, 1997.
- Roberto A De Santis and Srecko Zimic. Spillovers among sovereign debt markets: identification by absolute magnitude restrictions. 2017.
- Mert Demirer, Francis X Diebold, Laura Liu, and Kamil Yilmaz. Estimating global bank network connectedness. *Journal of Applied Econometrics*, 2017.

- Francis X Diebold and Kamil Yilmaz. Measuring financial asset return and volatility spillovers, with application to global equity markets*. *The Economic Journal*, 119 (534):158–171, 2009.
- Francis X Diebold and Kamil Yilmaz. On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134, 2014.
- Francis X Diebold and Kamil Yilmaz. *Financial and Macroeconomic Connectedness: A Network Approach to Measurement and Monitoring*. Oxford University Press, USA, 2015.
- Jianqing Fan, Nancy E Heckman, and Matt P Wand. Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of the American Statistical Association*, 90(429):141–150, 1995.
- Jeff Fleming, Chris Kirby, and Barbara Ostdiek. The economic value of volatility timing. *The Journal of Finance*, 56(1):329–352, 2001.
- Jeff Fleming, Chris Kirby, and Barbara Ostdiek. The economic value of volatility timing using “realized” volatility. *Journal of Financial Economics*, 67(3):473–509, 2003.
- Andrew T Foerster, Pierre-Daniel G Sarte, and Mark W Watson. Sectoral versus aggregate shocks: A structural factor analysis of industrial production. *Journal of Political Economy*, 119(1):1–38, 2011.
- Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.
- Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1, 2010.
- Renee Fry and Adrian Pagan. Sign restrictions in structural vector autoregressions: A critical review. *Journal of Economic Literature*, 49(4):938–960, 2011.
- Liudas Giraitis, George Kapetanios, Anne Wetherilt, and Filip Žikeš. Estimating the dynamics and persistence of financial networks, with an application to the sterling money market. *Journal of Applied Econometrics*, 31(1):58–84, 2016.

- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The elements of statistical learning: data mining, inference, and prediction, springer series in statistics, 2009.
- Junichi Hirukawa. Discriminant analysis for multivariate non-gaussian locally stationary processe. *Scientiae Mathematicae Japonicae*, 60(2):357–380, 2004.
- Junichi Hirukawa. Large-deviation results for discriminant statistics of gaussian locally stationary processes. *Advances in Decision Sciences*, 2012, 2012.
- Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.
- William James and Charles Stein. Estimation with quadratic loss. In *Breakthroughs in statistics*, pages 443–460. Springer, 1992.
- Òscar Jordà, Moritz Schularick, and Alan M Taylor. Financial crises, credit booms, and external imbalances: 140 years of lessons. *IMF Economic Review*, 59(2):340–378, 2011.
- Christian Kascha and Carsten Trenkler. Forecasting vars, model selection, and shrinkage. Technical report, Working Paper Series, Department of Economics, University of Mannheim, 2015.
- Gary Koop, M Hashem Pesaran, and Simon M Potter. Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74(1):119–147, 1996.
- Dimitris Korobilis and Kamil Yilmaz. Measuring dynamic connectedness with large bayesian var models. *Available at SSRN 3099725*, 2018.
- Markku Lanne, Helmut Lütkepohl, and Katarzyna Maciejowska. Structural vector autoregressions with markov switching. *Journal of Economic Dynamics and Control*, 34(2):121–131, 2010.
- Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411, 2004.
- Daniel J Lewis. Identifying shocks via time-varying volatility. *manuscript, Harvard University*, 2017.
- George Milunovich and Minxian Yang. On identifying structural var models via arch effects. *Journal of Time Series Econometrics*, 5(2):117–131, 2013.

- Maciej Niedźwiecki, Marcin Ciolek, and Yoshinobu Kajikawa. On adaptive selection of estimation bandwidth for analysis of locally stationary multivariate processes. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4860–4864. IEEE, 2016.
- Michael Parkinson. The extreme value method for estimating the variance of the rate of return. *Journal of Business*, pages 61–65, 1980.
- H Hashem Pesaran and Yongcheol Shin. Generalized impulse response analysis in linear multivariate models. *Economics Letters*, 58(1):17–29, 1998.
- Giorgio E Primiceri. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852, 2005.
- Ruprecht Puchstein and Philip Preuß. Testing for stationarity in multivariate locally stationary processes. *Journal of Time Series Analysis*, 37(1):3–29, 2016.
- Adi Raveh. On the use of the inverse of the correlation matrix in multivariate data analysis. *The American Statistician*, 39(1):39–42, 1985.
- Roberto Rigobon and Brian Sack. Spillovers across us financial markets. Technical report, National Bureau of Economic Research, 2003.
- Adam J Rothman, Elizaveta Levina, and Ji Zhu. Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association*, 104(485):177–186, 2009.
- Juan F Rubio-Ramirez, Daniel F Waggoner, and Tao Zha. Structural vector autoregressions: Theory of identification and algorithms for inference. *The Review of Economic Studies*, 77(2):665–696, 2010.
- Christopher A Sims. Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, pages 1–48, 1980.
- Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.
- Harald Uhlig. What are the effects of monetary policy on output? results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2):381–419, 2005.

- Michael Vogt. Nonparametric regression for locally stationary time series. *The Annals of Statistics*, 40(5):2601–2633, 2012.
- Michael Vogt and Holger Dette. Detecting gradual changes in locally stationary processes. *The Annals of Statistics*, 43(2):713–740, 2015.
- Zhou Zhou. Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association*, 108(502):726–740, 2013.
- Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006.
- Hui Zou and Hao Helen Zhang. On the adaptive elastic-net with a diverging number of parameters. *The Annals of Statistics*, 37(4):1733, 2009.

Curriculum Vitae, Ruben Hipp

| | |
|-------------------|---|
| 09/2013 – 07/2019 | PhD Studies in Economics, Center for Doctoral Studies in Economics, University of Mannheim. |
| 09/2013 – 07/2015 | M.Sc. in Economics, University of Mannheim. |
| 09/2010 – 07/2013 | B.Sc. in Economics, University of Mannheim. |
| 09/2017 – 02/2018 | Visiting PhD student, Universidad Pompeu Fabra, Barcelona. |
| 09/2012 – 01/2013 | Visiting student, Swansea University, Swansea. |
| 09/2000 – 06/2009 | Abitur, Goethe Gymnasium Emmendingen. |