



Universidade Nova de Lisboa

Dissertation, presented as part of the requirements for the
degree of Doctor of Philosophy in Economics/Finance

Connecting the dots: An exploration of spillover networks in economics and finance

Three essays

Felix Brunner

23267

A dissertation carried out for the Ph.D. in Economics/Finance
under the supervision of
Professor Paulo M.M. Rodrigues

October 2023

Connecting the dots: An exploration of spillover networks in economics and finance

Three essays

Abstract

In this thesis, I explore implications of spillover networks in economics and finance. Specifically, spillover networks allow to trace aggregate outcomes to micro-economic sources of innovations. This thesis investigates such mechanisms in various contexts. First, I explore several methods to infer networks from large-dimensional panels. An application to sectoral industrial production yields new insights into the decline in aggregate volatility known as the Great Moderation. Applying the same estimation techniques to volatility spillovers in financial markets reveals that asset-level shocks significantly affect aggregate market outcomes. Finally, I propose to incorporate spillover networks into asset pricing models and collect empirical evidence for the impact of directed network linkages on equilibrium asset prices.

© Nova SBE and Felix Brunner, 2023
brunner.felix@gmail.com

Data and code available upon request.

This thesis was typeset by the author using L^AT_EX.

This work was supported by the Fundação para a Ciência e Tecnologia, Portugal.

Keywords: Networks, statistical learning, time series econometrics, asset pricing.

To Frederik.

Preface

These past years, encompassing the journey of continuously working towards the completion of this thesis, have been nothing short of an intellectual odyssey. As I reflect on the countless hours spent in research, writing, and revision, I am looking back at a process with ups, downs, and occasional unforeseen shocks. I am reminded of days when I felt like I had finally understood, and others when I was convinced I knew nothing at all. Crafting this work has been a profound source of personal and intellectual growth, at times only a product of persistence in the face of numerous challenges, at others a joyful pursuit. This protracted process has been shaped not only by my dedication but also quite significantly by the support and inspiration provided by many remarkable individuals. It is with deep gratitude that I recall what has revealed to me the significance of networks, not only in economic studies, but also in the form of an incredibly powerful supportive network.

My deep gratitude goes to the supervisor of the research that finally went into this thesis, Paulo, for stepping in during a time of unforeseen transition, which ultimately turned out to be a lucky twist of fate. Your trust and encouragement during this period allowed me to not only recover from the sudden disruption but also to thrive in my academic pursuits of unfamiliar areas. I knew I could always count on your advice and willingness to help, although I did not always make use of it. Thank you also for understanding my frequent absence to make ends meet. Your expertise as a researcher and your qualities as a person are truly inspiring, and it is a privilege to have learned under your mentorship.

I owe a huge debt of gratitude to my dear friend and collaborator, Ruben, for seeking my alleged skills in statistical learning to transform of a draft he had conceptualized during his studies into a full-fledged publication. It was at this juncture that my own endeavors took a significant turn for the better, and our collaboration became an indispensable element in the creation of this thesis. I am sincerely thankful for your friendship, as well as your mentorship in areas I might not have ventured into without your guidance. I hope you share my perspective that our collaboration complemented his formidable skill set with a unique contribution of my own, rather than hindering progress through excessive perfectionism. Your significant contributions to this work

have been nothing short of instrumental, and I hope to be there to accompany your future undertakings, preferably in a neighborhood closer to mine.

I must also express my sincere thanks to my early advisor, Martijn, and my early mentor, Fernando, who patiently endured my misguided first attempts at finding and executing meaningful research questions. As I transitioned into the research stage following a bumpy leap into the academic world, I too often not only lacked confidence but also sound knowledge. Through your push and well-intentioned challenge of my work, I learned that I was still missing a trustworthy instinct for research ideas and the necessary skills to realize them. Ultimately, it inspired me to look further than economics into subjects connected to data science and machine learning.

At the start of my PhD studies, I embarked on this academic adventure with a deep motivation to master a quantitative discipline and to build serious research abilities. Ill-equipped with an unfitting skill set from my previous studies, I disorientedly stumbled into a dense curriculum of economic coursework and exams. If it was not for the fellowship of my classmates and their willingness to share insights when I doubted my abilities, this journey could have been over so much sooner. A tip of the hat goes also to numerous colleagues who I spent much time with on and off campus, and who I became friends with. In particular, but not only, this thesis has invaluable benefited from the discussions and friendship with Robbie, Fahiz, and Max. My gratitude also extends to the students of earlier cohorts for offering advice and material, and to the companions at the University of North Carolina for guiding and upgrading my semester-long stay there enormously.

At this point, I would also like to express my appreciation also to many remarkable faculty members who have offered orientation and comments, and whose insights and passion for their respective fields have consistently inspired my academic pursuits. I have learned a great deal from their lectures and discussions, and must thank countless faculty around the world that provide open access to high-quality materials. Further, I am grateful for the helpfulness that I experienced from administrative staff and local connections with regards to the bureaucratic and institutional hurdles in various stages. I acknowledge the financial support the early stages of this work have received by the Fundação para a Ciência e a Tecnologia, Portugal, under grant PD/BD/128120/2016, and the compute resources made available by the Bank of Canada later on.

I am also incredibly grateful to everyone who has contributed to this unpredictable process without even knowing. It has been an challenging journey into the unknown, which I would not have dared without knowing so many wonderful individuals in my network of support reaching far beyond academia. Shout-outs, at this point, especially to many great friends, some in close proximity, some scattered in distant places around

the world. Your friendship has helped to keep me sane when my work gave me no reason to, and I hope we manage to cross paths more often in the future. Without you, my thesis would still have been written but I would have laughed a lot less. In this context, I want to mention many sports coaches and exercise buddies that have prevented the complete disintegration of my physical appearance during much of deskwork.

None of my modest achievements would have been possible without the unwavering support of my loving parents, Doris and Rolf, and their unfaltering belief in me and my judgement. The values you instilled in me during my upbringing, your boundless joy in life, your sincerity, and your down-to-earthness have shaped the person I am today. I consider myself incredibly privileged to be your son. I know that it hasn't been easy to watch me spread my wings and venture into the far reaches of the world in pursuit of my often enigmatic goals. You've done an exceptional job in raising me, and any flaws I possess are entirely my own. I could not have asked for a better start in life, and I carry your love and teachings with me in all my endeavors. Thank you for being the steady branches that carry my nest.

Finally, my heartfelt thanks goes to my beloved partner, Katharina, who had to endure my constant focus on the creation of this work for too long. Your support and encouragement have often been the wind beneath my wings when my own belief wavered. Your insistence on distracting me from my relentless work has been a blessing, and you have inspired countless fateful decisions. Thanks to you, this journey has been much more thrilling, infused with a sense of shared adventure and countless memorable moments. I sincerely hope that my frequent absence in mind, my overly critical thinking, and my bad sense of humor have not caused you too much distress or discouraged you from bearing with my nonsense. Now that my mind is finally liberated from the ballast of this thesis, I vow to be better and to prioritize our time together. You and Frederik are the most compelling reasons imaginable to look forward in excitement to the adventures that lie ahead.

Berlin, October 2023

Felix Brunner

Contents

Introduction	1
1 Estimating large-dimensional connectedness tables: The Great Moderation through the lens of sectoral spillovers	7
1.1 Introduction	8
1.2 Methodology	11
1.3 Simulation study	21
1.4 Empirical application: Production volatility spillovers and the Great Moderation	28
1.5 Conclusions	44
1.A Appendices	45
1.B Bibliography	55
2 Echo-nomics: Does the stock market resonate with granular innovations?	61
2.1 Introduction	62
2.2 Empirical setting	65
2.3 The volatility spillover network	72
2.4 Network granularity and systemic effects	80
2.5 Conclusions	95
2.A Appendices	97
2.B Bibliography	117
3 Granular risk premia	123
3.1 Introduction	124
3.2 Framework	128
3.3 Empirical strategy	133
3.4 Tracing granular risk premia	137
3.5 Granular risks and asset prices	144
3.6 Conclusions	152

3.A Appendices	155
3.B Bibliography	163
Conclusions	169

List of Tables

1.1	Simulation results for the regularization of the regression coefficients . .	23
1.2	Simulation results for the regularization of the covariance matrix	24
1.3	Summary of the model selection with cross-validated hyperparameters .	32
1.4	Comparison between static and dynamic model choices	33
1.5	Estimation results for different levels of sectoral disaggregation	35
1.6	Decomposition of variance contributions	41
2.1	Market impact portfolio sorts	83
2.2	Granular time series regression estimates	88
2.3	Granular time series regressions with control variables	93
2.4	Granular time series regressions with instrumental variables	94
2.5	Tickers in the 31 March 2022 sample	100
2.6	Sector abbreviations	101
2.7	List of peaks in systemic asset importance	115
3.1	Model comparison with various proxy tables and restrictions	138
3.2	Regressions of aggregate moments on network architecture properties .	140
3.3	One-way cross-sectional portfolio sorts.	147
3.4	Two-way cross-sectional portfolio sorts.	148
3.5	Regressions of aggregate returns on model expectations	150

List of Figures

1.1	Estimation overview for high-dimensional FEVDs	12
1.2	Simulated bias and variance of regularized estimators	25
1.3	Simulated link selection of regularized estimators	27
1.4	Drawings of connectedness networks for each period	37
1.5	Cumulative distribution functions of sectoral out-connectedness	38
1.6	Scatter plot of sectoral out-connectedness evolution	38
1.7	Bump chart of sectoral contributions to aggregate IP index	42
1.A.8	Illustration of cross-validation procedure	45
1.A.9	Annualized growth rates of monthly industrial production	46
1.A.10	Cross-validation results for adaptive elastic-net tuning parameter	47
1.A.11	Cross-validation results for covariance regularization tuning parameters	48
1.A.12	Non-regularized connectedness networks	49
1.A.13	Histogram of sectoral in-connectedness	49
1.A.14	Scatterplots of sectoral weights and in-connectedness	50
1.A.15	Histogram of sectoral out-connectedness	50
1.A.16	Regression of out-connectedness on inventories-to-shipments ratio	51
1.A.17	Bump chart of sectoral contributions to aggregate IP index without spillovers	52
1.A.18	Bump chart of sectoral spillover contributions to aggregate IP index	53
2.1	Proportion of the CRSP dataset contained in the sample	67
2.2	Forecast error variance decomposition network graph	73
2.3	Rolling asset-level out-connectedness	77
2.4	Rolling network statistics for the FEVD networks	79
2.5	Rolling spillover impact factors of individual assets	82
2.6	Granular and network residuals	86
2.7	Rolling granular time series regression coefficients	89
2.8	Illustration of the sampling logic	97
2.9	Evolution of the daily volatility index	97
2.10	Firm size distribution 31 March 2022	99

2.11	Histograms of intraday volatility data	101
2.12	Correlation heatmaps of intraday volatility data	102
2.13	Hyperparameter search for the Adaptive Elastic Net	103
2.14	VAR coefficient estimate heatmaps	104
2.15	Histograms of VAR residuals	104
2.16	Correlation heatmaps of VAR residuals	105
2.17	VMA matrix heatmaps	106
2.18	Hyperparameter search for the Graphical LASSO	107
2.19	Residual covariance matrix heatmaps	108
2.20	Impulse response matrix heatmaps	109
2.21	Forecast error variance network graph	110
2.22	Rolling estimation summary statistics	111
2.23	Regularization summary statistics	113
2.24	Rolling size weight impact of individual assets	114
3.1	Graphical representations of factor model and network model	129
3.2	Fraction of the CRSP dataset contained in the rolling samples	134
3.3	Common and granular risk premia estimates	142
3.4	Sectoral sums of granular risk premia estimates	143
3.5	Variance decompositions aggregate and individual returns.	146
3.6	Concentration of market capitalizations in the CRSP universe	160
3.7	Estimate of the scaling coefficient c for each sample	160
3.8	Log-likelihood for each sample	160
3.9	Network connectedness over time	161
3.10	Network asymmetry over time	161
3.11	Network concentration over time	161
3.12	Risk exposures of the aggregate portfolio	162
3.13	Expected aggregate returns and realized aggregate returns.	162

Introduction

The study of spillover networks has emerged as a powerful tool for uncovering the mechanisms that underlie economic dynamics. Networks are a fundamental and ubiquitous aspect of our world, serving as the invisible web that connects disparate elements, whether they are individuals, companies, or countries. In the scientific study of networks, researchers of many disciplines seek to unravel the hidden structures that underpin the intricacies of an increasingly interconnected world. Through the lens of networks, they can understand and navigate the complex interplay of relationships and decipher the patterns and pathways that drive the flow of information and influence. In economics, networks allow researchers to gain novel insights into how economic activity is interlinked and how shocks and innovations propagate. For instance, central ideas in the economic line of reasoning like production, consumption, or the flow of information are inevitably intertwined with supply chain networks, financial networks, or networks of social interaction.

One particular facet of networks that has garnered substantial attention is the notion of spillover effects. A spillover is the phenomenon that an economic event in one context entails the occurrence of an event in a seemingly unrelated context, often with profound consequences. For example, technological innovations in one industry may entail externalities on the output of other industries, or the release of information about a given entity may also resolve the uncertainty surrounding another. Gathering spillovers between many entities creates a network that provides a more holistic view of how events propagate from one entity to another, effectively “connecting the dots” and revealing cascades of what would otherwise be isolated effects. Therefore, spillover networks shed light on how innovations spread in the larger system, and thus, illuminate the relationship of individual entities to aggregate quantities.

The goal of this thesis is to understand the intricate interplay between micro-level innovations and macro-economic outcomes, as well as the systemic effects that arise from the propagation of these granular influences. The overarching question that emerges is whether it is possible to trace aggregate variation through the dynamic framework of the spillover network back to their granular sources. Answering this question is of paramount importance to economic and econometric analysis as it delves

into the depths of causality and interdependence within complex economic systems. Network analysis presents a compelling avenue for such explorations inspired by the seminal contributions of Gabaix (2011) and Acemoglu et al. (2012). Here, granular effects arise if low-level innovations propagate through spillover networks to the extent that they lead to significant perturbations at the aggregate level.

The comprehension of granular effects holds the key to a deeper understanding of the dynamics of economies and financial markets, where seemingly small changes can trigger substantial shifts with systemic repercussions. By investigating the mechanisms and pathways through which granular innovations diffuse and influence aggregate quantities, I aim to contribute to the comprehension of economic networks that allow low-level events to influence not only macro-level outcomes but also the stability and resilience of entire systems. The study of spillover networks is, therefore, not only a fascinating intellectual pursuit but also a crucial avenue for enhancing economists' ability to anticipate, model, and potentially mitigate systemic risk and disruptions in the face of granular events.

This dissertation consists of three chapters, where each chapter is a self-contained essay that addresses a separate economic question in relation to spillover networks and the granular effects that arise from them. Although later chapters do not require the knowledge of prior ones, they do build upon previous findings.

Spillovers often arise when the underlying links cannot be observed directly. For this reason it is of central interest to econometricians to be able to infer spillover networks from available data panels. The standard approach to do so is to estimate forecast error variance decomposition tables, that is, matrices obtained as a function of vector autoregressive coefficients and residual covariance matrices, and interpret them as networks. In Chapter 1, Ruben Hipp and I investigate the question of how such tables can be obtained in settings of large-dimensionality using regularized estimation methods. We conduct a comparative simulation study of various such methods and show that standard estimators are unsuitable in large-dimensional contexts. In an empirical application, we then apply a cross-validated estimator on sectoral spillovers of industrial production in the US from 1972 to 2019. Our results suggest that the decline of granular effects due to a weakening spillover network are a critical explanation for the decrease in aggregate volatility known as the Great Moderation.

Another domain in which the connection of granular spillover networks and aggregate variation matters significantly is that of financial markets, where the propagation of low-level innovations can lead to a domino effect, affecting entire markets. For instance, the financial crisis of 2008 illustrated how interconnectedness in the form of credit and risk networks could lead to the rapid transmission of distress from a single institution

to the global financial system. Chapter 2, again in collaboration with Ruben Hipp, takes the large-dimensional network estimation techniques of the first chapter to the US stock market to understand information contagion and systemic risk in the form of volatility spillovers between listed firms. By scrutinizing repeatedly estimated networks of 100 US stocks, we document the presence and fast-changing dynamics of directed spillovers in equity markets. Given the potential for low-level volatility to spread in the cross-section of stocks, we then probe a version of the “granular hypothesis” where network interconnectedness, combined with a concentrated size distribution, enables micro innovations at the firm level to significantly influence the aggregate market index. The results indicate the presence of granular effects in the stock market, hinting that inter-asset spillovers are a missing link in models of financial markets.

The fact that information spillovers between individual assets impact aggregate quantities implies that they also influence equilibrium asset prices. Hence, the presence of granular effects in financial markets raises the question whether granular effects measurably affect expected returns. To investigate this question, I revisit the asset pricing framework in Chapter 3 and complement it with a spillover network that allows asset-level returns to have an effect on each other. I derive a pricing kernel from a network connectedness model that implies risk premia for systemically relevant granular risks. Such “granular risk premia” arise from a concentration of index weights together with a spillover structure that propagates non-systematic innovations to a large portion of investor wealth. Based on the volatility spillover networks of the second chapter, I then collect empirical evidence in support of various conjectures based on this framework. My analysis suggests that granular risk premia exist temporarily and relate to mean returns in the cross-section of stocks. Thus, consistent with the continuous emergence of anomalies relative to reduced-form factor pricing models, granular effects could be an overlooked determinant of asset prices.

At this point, I want to point out some foundational and continuative textbook references to readers that are not fully familiar with all topics that come together in this thesis. First, Newman (2018) provides an excellent general introduction into the scientific study of networks. For questions connected to the estimation of networks, Lütkepohl (2005) offers a solid basis in multivariate time series econometrics, and Hastie et al. (2009) covers matters connected to the application of statistical learning techniques. Lastly, for anyone looking for insight in relation to the asset pricing topics covered in the last chapter, Cochrane (2009) is a comprehensive reference.

Bibliography

- Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). “The network origins of aggregate fluctuations.” In: *Econometrica* 80.5, pp. 1977–2016.
- Cochrane, J. (2009). *Asset pricing: Revised edition*. Princeton university press.
- Gabaix, X. (2011). “The granular origins of aggregate fluctuations.” In: *Econometrica* 79.3, pp. 733–772.
- Hastie, T., R. Tibshirani, J. H. Friedman, and J. H. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.
- Newman, M. (2018). *Networks*. Second. Oxford university press.

Chapter 1

Estimating large-dimensional connectedness tables: The Great Moderation through the lens of sectoral spillovers

with Ruben Hipp

Abstract

We estimate sectoral spillovers around the Great Moderation with the help of forecast error variance decomposition tables. Obtaining such tables in high dimensions is challenging because they are functions of the estimated vector autoregressive coefficients and the residual covariance matrix. In a simulation study, we compare various regularization methods on both and conduct a comprehensive analysis of their performance. We show that standard estimators of large connectedness tables lead to biased results and high estimation uncertainty, both of which are mitigated by regularization. To explore possible causes for the Great Moderation, we apply a cross-validated estimator on sectoral spillovers of industrial production in the US from 1972 to 2019. We find that the spillover network has considerably weakened, which hints at structural change, for example, through improved inventory management, as a critical explanation for the Great Moderation.

Earlier versions of this chapter circulated under the title “Estimating Large Dimensional Connectedness Tables”. It has meanwhile been published as Brunner and Hipp (2023).

1.1 Introduction

With the onset of the Great Moderation, around 1984, key macroeconomic time-series exhibit sharp decreases in growth rate volatility. Whether this shift in fluctuations is due to a structural change in the economy, improved economic policies, or just good luck has been extensively studied using a variety of approaches. In particular, for industrial production (IP), the literature provides manifold narratives, often using contemporaneous correlations between sectors to approximate dependencies. Recent advancements further allow econometricians to describe directional dependencies in the form of forecast error variance decompositions (FEVDs, see Diebold and Yilmaz (2014)). Namely, FEVDs measure how much the variation of one sector can explain the variation of another. They do so by condensing contemporaneous and lagged dependencies into a single connectedness table based on vector autoregressions (VARs). These tables, however, inherit the VAR's estimation uncertainty in high dimensions: the estimation error blows up when the number of variables (N) approaches the number of time observations (T).

The goal of this paper is to examine sectoral spillovers as a potential driver of the decline in volatility during the Great Moderation. It consists of two parts: First, we explore regularization methods for estimating large networks from time-series observations and detail the properties of different approaches in an extensive Monte Carlo (MC) simulation. Second, we investigate the structural change hypothesis of the Great Moderation by applying this methodology to sectoral spillovers of IP.

Econometricians typically face a high-dimensional setup when estimating FEVDs. For one thing, networks must include all relevant variables to have a unified and precise interpretation; that is, N is large. In addition, time variation in the parameters — for example, a structural break or rolling windows — reduces the number of observations; that is, T is often small. With large N and small T , standard estimation methods produce poor estimates and bad forecasts due to overfitting. Regularization methods for regressions and covariance matrices counteract these ramifications. Yet, it remains unclear which ones to choose. Thus, we provide an extensive MC study of regularization techniques combined with FEVDs to guide researchers in large-dimensional network estimations.

While there are successful applications of FEVDs in high dimensions (for example, see Demirer et al. (2017)), two questions remain unanswered. First, how does estimation uncertainty affect the overall results of FEVDs? Second, would an additional regularization of the innovation covariance matrix improve the results? Our MC simulation results demonstrate the performance gain of each regularization step in the estimation of the overall FEVD network. We show that regularization of both the

coefficients and the covariance matrix not only trade-off bias for variance by reducing the estimation uncertainty of the FEVD, but also mitigates a positive bias in the entries. To the best of our knowledge, this result is novel in the literature and highlights the importance of regularization in the context of FEVDs. Perhaps surprisingly, a horse race of regularization methods applied to FEVDs yields similar results among estimators, with the winner being conditional on the setting.

In our application, we investigate changes to the sectoral dependency structure of IP around the Great Moderation. The central question is whether the large volatility of the aggregate IP index originated from amplifications of sectoral fluctuations by the network in the spirit of Acemoglu et al. (2012). That is, we address the gap in the literature on how structural change in sectoral interconnectedness affects aggregate index volatility. The directed nature of FEVDs allows us to investigate linkages that were previously hidden under the rationale of correlations as a result of common exposure to aggregate shocks. Thus, we motivate our application by the fact that a strong intersectoral network of idiosyncratic shocks is observationally equivalent to the prevalence of aggregate shocks, if connectedness is ignored.

We apply FEVDs to scrutinize IP spillovers between 88 sectors in the US from 1972 to 2019. The estimation is challenging since the split into pre- and post-Great Moderation periods reduces the effective sample size. We use cross-validated regularization to tackle this challenge. The estimates provide time-varying spillover networks that uncover the corresponding directed dependency structure among sectors. We find that sectors that were initially influential decreased their outgoing links after 1984, reducing the importance of spillover effects overall. An analysis of contributions to the aggregate index suggests that inventory-heavy sectors added considerably to the high volatility pre-Great Moderation via their spillovers. We connect individual changes in spillovers to decreases in inventories-to-shipments ratio, which supports a narrative of structural change through improved inventory and supply chain management. After 2007, we observe again higher levels of aggregate volatility, but without out-sized contributions of some sectors. This observation insinuates no reversal of the Great Moderation's structural change.

Our comparative MC study connects to the literature on the regularization of regressions and covariance matrices. Regularizations have become popular not only because the increasing availability of data makes variable selection more critical, but also because technological advances make the application of high-dimensional estimators increasingly feasible. For the regression step, we consider shrinkage estimators, such as ridge regression by Hoerl and Kennard (1970), variable selection methods, such as LASSO by Tibshirani (1996), and combinations of the two, such as adaptive elastic-net

by Zou and Zhang (2009). Whereas these methods find application primarily in cross-sectional contexts, the time-series literature has succeeded in using regularization, for example, in the general case of VARs in Kascha and Trenkler (2015) and in the setting of FEVDs in Demirer et al. (2017). For the regularization of covariance matrices, we examine variable selection methods for the partial correlation matrix, as in the graphical LASSO by Friedman et al. (2008), optimal shrinkage estimators, as in Ledoit and Wolf (2004), and sample covariance thresholding, as in Bickel and Levina (2008), Rothman et al. (2009) and Cai and Liu (2011). We contribute to this diverse literature by comparing the regularized estimators' performances in the context of FEVDs.

Empirically, we contribute to the understanding of the decline in IP volatility that took place during the Great Moderation. A prevalent hypothesis in the context of industrial production is the decline in aggregate shocks, as described by Foerster, Sarte, et al. (2011). This contrasts with the results of Gabaix (2011), who considers overweight index constituents as a central driver of aggregate fluctuations in industrial production. Similarly, Carvalho and Gabaix (2013) argue that idiosyncratic shocks on the sectoral level can account for the shift in macroeconomic volatility. Acemoglu et al. (2012) connect to this idea and stress that strong production networks propagate sectoral productivity shocks to the rest of the economy and hinder diversification in the aggregate index. In contrast to the fast-expanding literature that uses input–output tables as network proxies (for a review see Carvalho and Tahbaz-Salehi (2019)), our analysis directly sheds light on the network implied by sectoral IP correlations. Thus, we offer a unifying view on the two opposing explanations of Foerster, Sarte, et al. (2011) and Gabaix (2011) by highlighting the transition from an economy with strong intersectoral spillovers from a few sectors to an economy with a less transmitting network.

Our empirical results show that the intersectoral network has changed structurally with the Great Moderation, making idiosyncratic shocks less likely to impact the aggregate. In other words, before the Great Moderation, granular shocks were able to propagate through the network such that they amplified to strong aggregate fluctuations. This finding offers a new perspective on the observation that correlations between sectors have significantly changed with the Great Moderation, which is often attributed to the existence of a strong factor structure, that is, shocks to multiple sectors at once. Similar to Galí and Gambetti (2009), our findings support the notion that structural change has abated shock transmission, for example, through improvements to supply chain and inventory management as argued in Kahn et al. (2002), Summers et al. (2005) and Davis and Kahn (2008).

The rest of the paper is organized as follows. In Section 1.2, we introduce the

concept of FEVDs and provide an overview of various regularization methods. We assess their performance in a simulation study in Section 1.3. Section 1.4 applies the regularization of FEVDs to the IP setup to answer the question of sectoral spillovers. Finally, Section 1.5 concludes. Supplementary material, such as mathematical details, practical illustrations, and complementary empirical graphs can be found in the Appendix.

1.2 Methodology

This section provides a general overview of FEVDs and introduces suitable regularization methodologies to mitigate estimation uncertainty in large-dimensional applications.

We start with an N -dimensional stable VAR(1) process,

$$y_t = \nu + Ay_{t-1} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma), \quad \forall t = 1, \dots, T. \quad (1.1)$$

Following Pesaran and Shin (1998), we obtain the FEVD as a function of the VAR coefficient matrix A and the innovation covariance matrix Σ .¹ A detailed derivation is presented in Section 1.2.1.

As in many structurally motivated economic models, we want to include a broad set of variables. However, considering many variables entails the curse of dimensionality. Thus, to estimate high-dimensional FEVDs, we have to estimate the VAR coefficient matrix A and the innovation covariance matrix Σ in large dimensions. For that purpose, we assess the performance of various estimation techniques by comparing different regularization approaches for large VARs.² Figure 1.1 gives an overview of the regularization methods considered in this paper.

The statistical learning literature contains two types of regularization techniques that are beneficial for FEVDs: the regularization of coefficients and the regularization of covariance matrices. Concerning the former, LASSO techniques tend to perform well in autoregressive setups. Such techniques apply regularization to the coefficient matrix A but do not imply any regularization for the covariance matrix Σ . The combination

¹Note that any VAR(p) translates into a VAR(1), hence, our specification provides the companion form to higher-order lag numbers without loss of generality. For details see Lütkepohl (2005, page 15).

²We address the most prominent examples, but we are aware that there are additional approaches that are beyond the scope of this paper. L0 penalties in high-dimensional regressions are computationally infeasible for classical machines: the nonconvexity of the L0 penalty creates a combinatorial problem, which is NP-hard. Notable examples of variable selection are the information criteria AIC/BIC and the Variational Garotte in Kappen and Gómez (2014). We experimented with the latter approach in small scale simulations, which turned out to be computationally too expensive and did not yield any improvement. The covariance matrix estimation literature also deals with the regularization of the eigenvalues; for example, in Lam et al. (2016).

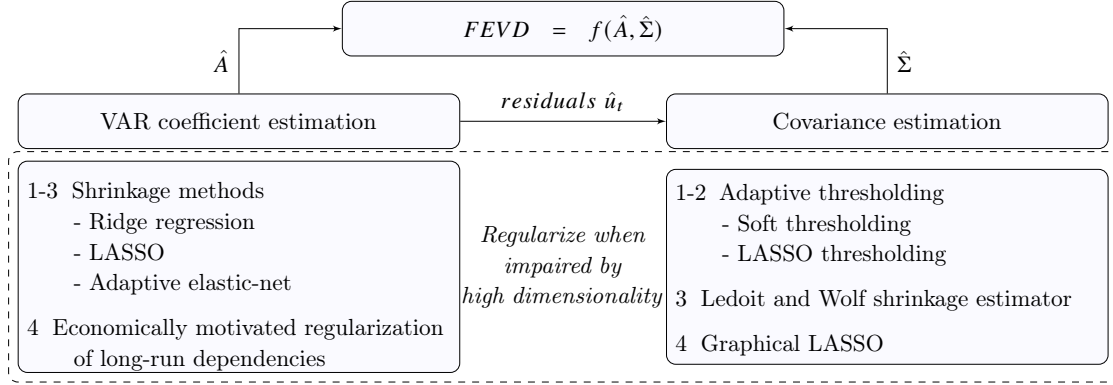


Figure 1.1: Estimation overview for high-dimensional forecast error variance decompositions (FEVDs). The estimates $\hat{A}$ and $\hat{\Sigma}$ are required for the estimation of FEVDs. The lists describe the considered regularization methods in this paper. Note that although the VAR residuals are fed into the estimation of the covariance matrix, properties such as the degree of shrinkage and sparsity structure are determined independently by the estimators.

of A and Σ in the FEVD, however, suggests that poor estimation of the innovation covariance matrix renders the overall estimate noisy. That is, the estimation of Σ suffers from a similar uncertainty induced by high dimensionality. Hence, we resort to covariance shrinkage estimators as an alternative to the sample covariance matrix. Our approach, therefore, is to combine regularization methods for the unknown regression coefficient and covariance matrix to achieve the best possible estimate of the FEVD with respect to the estimation error. We describe each method in detail in Section 1.2.2.

1.2.1 Generalized forecast error variance decompositions

Due to the stability assumption, the VAR process in (1.1) can be written in moving average (MA) representation

$$y_t = \mu + \sum_{k=0}^{\infty} \Phi_k u_{t-k}, \quad (1.2)$$

where the MA parameters are defined as $\Phi_k = A^k$, $\forall k \geq 0$ and μ represents the mean term. Note that the components of u_t are generally not orthogonal such that structural interpretations are economically meaningless. Koop et al. (1996) and Pesaran and Shin (1998) define the (unscaled) generalized impulse response (IR) function at horizon h to an impulse Δ_j on the j th entry of the reduced form innovation u_t . They do so by integrating out the effect of all of the remaining impulses in the innovation vector:

$$IR(h, \Delta_j, j) = E[y_{t+h} | u_{j,t} = \Delta_j] - E[y_{t+h} | u_{j,t} = 0]. \quad (1.3)$$

Under a Gaussian assumption, we can use

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

with σ_{ij} being the ij th entry in Σ and e_j as the j th column of the identity matrix. Substituting this expression in (1.3), the impulse response function can be rewritten as

$$\text{IR}(h, \Delta_j, j) = \Phi_h \Sigma e_j \sigma_{jj}^{-1} \Delta_j.$$

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

$$\Psi^g(h) = [\psi_{ij}^g(h)] = [\text{IR}(h, \sqrt{\sigma_{11}}, 1), \dots, \text{IR}(h, \sqrt{\sigma_{NN}}, N)] = \Phi_h \Sigma \text{diag}(\Sigma)^{-\frac{1}{2}}, \quad (1.4)$$

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

Analogously to standard impulse response analysis, we gain further insights rewriting y_t as a vector-valued impulse response function multiplied by an innovation vector. Let $P = \Sigma \text{diag}(\Sigma)^{-\frac{1}{2}}$ 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 $\Omega = \text{diag}(\Sigma)^{\frac{1}{2}} \Sigma^{-1} \text{diag}(\Sigma)^{\frac{1}{2}}$. Then it is possible to express (1.2) as

$$y_t = \mu + \sum_{k=0}^{\infty} \Psi^g(k) u_{t-k}^g. \quad (1.5)$$

Henceforth, we can interpret u_t^g as the innovation vector in the generalized impulse response analysis, where a single element receives a unity shock and all others remain at zero. Implicitly, we assume that u_t^g is an orthogonal and exogenous shock vector. However, due to the distribution of u_t^g with the inverse correlation matrix Ω , this step results in an approximation error. The approximation error can be linked to the partial correlation matrix of u_t , and it is generally bigger when more entries are partially correlated.³

Note that the h -period forecast error variance of variable i 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

$$\text{MSE}[y_{i,t+H-1} | \{u_{t+h}^g = e_j\}_{h=0}^{H-1}] = \sum_{h=0}^{H-1} (\psi_{ij}^g(h))^2. \quad (1.6)$$

The H -step-ahead forecast error variance contributions from all variables to i is just its mean squared error (MSE):

$$\text{MSE}[y_{i,t+H-1}] = \left(\sum_{h=0}^{H-1} (\Phi_h \Sigma \Phi_h') \right)_{ii}. \quad (1.7)$$

Pesaran and Shin (1998) divide (1.6) by (1.7) and get a table showing the contributions from innovations in variable j to the H -step-ahead forecast error variance of variable i .

³For a discussion of Ω see Ravelli (1985).

Like Diebold and Yilmaz (2014), we denote the H -step-ahead generalized FEVD as $D^{gH} = [d_{ij}^{gH}]$ with entries

$$d_{ij}^{gH} = \frac{MSE[y_{i,t+H-1} | \{u_{t+h}^g = e_j\}_{h=0}^{H-1}]}{MSE[y_{i,t+H-1}]}. \quad (1.8)$$

Note that the numerator implicitly shocks single entries of u_t^g and the denominator shocks single entries of u_t . In other words, the “generalized” FEVD approximates shocks with u_t^g and is only accurate if $\Omega = I_N$; that is, if Σ is diagonal. Optionally, Diebold and Yilmaz (2014) row-normalize these tables for a cleaner network interpretation. However, row normalization distorts the entries and further complicates estimation errors. Thus, if not explicitly stated, we do not perform this step.

At this point, it is worth mentioning that for clean calculations of variance decompositions, we require the model in (1.1) to be identified. The literature on structural VARs contains many identification approaches that are suitable for meeting this objective in various circumstances. Yet, most restrictions such as exclusion restrictions (see Sims (1980)) dissent from the motivation of detecting links between variables. Some other schemes, such as heteroskedasticity identification (see Rigobon and Sack (2003)), are less restrictive on the directionality of effects and also have been successfully applied to FEVDs (see Hipp (2020)). However, in the high-dimensional context, such identification schemes are practically infeasible. Thus, we opt for the generalized version with imperfectly orthogonalized shocks to avoid stricter assumptions about the process, allowing us to interpret results more neutrally. Nevertheless, we acknowledge the drawbacks of using “generalized identification”, and address the resulting imprecision by testing deviations from orthogonality in the empirical section.

1.2.2 Estimating large forecast error variance decompositions

Regularizing vector autoregressive coefficients

First, we follow the notation of Kascha and Trenkler (2015) for $p = 1$ lags and transform the VAR(1) such that the coefficient matrix A can be estimated in vector form. That is,

$$y = (Z' \otimes I_N) \beta + u, \quad (1.9)$$

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

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

Here $\|\cdot\|^2$ denotes the square of the Frobenius norm. Based on this objective function, we aim to regularize the coefficient matrix A . On this account, we consider elastic-net regularization, which comprises the extreme cases of LASSO and ridge regression. Additionally, we introduce a new regularization target that enforces sparsity on the long-run dependencies.

(Adaptive) elastic-net, LASSO and ridge regression We outline the most general concept following the adaptive elastic-net of Zou and Zhang (2009). 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. A comprising definition of the adaptive elastic-net estimator class is

$$\hat{\beta}_{AEnet} = \underset{\beta}{\operatorname{argmin}} \left[\|y - X\beta\|^2 + \lambda^{NET} \sum_{i=1}^{N^2+N} w_i \left(\alpha |\beta_i| + (1 - \alpha) \frac{1}{2} \beta_i^2 \right) \right], \quad (1.10)$$

where $w_i = |\hat{\beta}_{i,ini}|^{-\gamma}$ is an initial guess with $\gamma > 0$ and λ^{NET} is a tuning parameter that controls the strength of the elastic-net penalty, and must be chosen by the researcher. 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).⁴

The regression in (1.10) is an enhanced version of the penalty regression, and thus, generalizes a family of regularized estimators. For example, the elastic-net penalty with $\alpha \in (0, 1)$ combines the LASSO and the ridge estimator and inherits the desirable properties of both; for example, it removes the degeneracy of the LASSO estimator caused by extreme correlations while still performing variable selection. Moreover, the absolute penalty term automatically selects variables while the quadratic penalty shrinks entries and stabilizes the solution paths (see Zou and Zhang (2009)).

Now, choosing $w_i = 1$ gives the naive elastic-net,⁵ with the nested cases of ridge regression ($\alpha = 0$) and classical LASSO ($\alpha = 1$). By imposing different w_i , we obtain the adaptive elastic-net, with the special case of the adaptive LASSO ($\alpha = 1$) from Zou (2006). In Section 1.3, we compare the performances of LASSO, ridge regression and adaptive elastic-net in a simulation of FEVDs. If not stated differently, we standardize observations y and X and use $w_i = |\hat{\beta}_{i,OLS}|^{-1}$ for the regularization and hyperparameter selection.

⁴The `glmnet` routine is available for many programming languages. We employ the implementation in MATLAB by Qian et al. (2013).

⁵Zou and Zhang (2009) rescale the naive estimator by $(1 + (1 - \alpha)\lambda^{NET}/T)$. Similar to Friedman et al. (2010), we drop this distinction.

Regularizing long-run effects While most regularization methods assume sparsity in either the regression coefficients or the covariance matrix, economic setups sometimes motivate independence between variables at all time horizons. 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 novel method that models sparsity in the long-run dependency.

For example, take the moving average matrices of a VAR(1), $\Phi_k = A^k$. If economic theory implies that only a few variables 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 power of a matrix is a complex function of the coefficient matrix A .

To overcome this issue, we take the forecast error, that is, the response to a one-standard-deviation impulse. The long-run (lagged) response of an impulse is

$$\begin{aligned} FE(H) &= \sum_{h=0}^{H-1} \Phi_h = \sum_{h=0}^{H-1} A^h, \\ \lim_{H \rightarrow \infty} FE(H) &= \sum_{h=0}^{\infty} A^h = (I_N - A)^{-1}, \end{aligned} \quad (1.11)$$

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 — that is, the element-wise squared version — is also zero. Thus, to impose sparsity on the spillover network, this matrix is a potential regularization target. Take (1.11) and plug it into the model (1.1):

$$\begin{aligned} y_t &= \nu + (I_N - FE(\infty)^{-1})y_{t-1} + u_t, \\ \Delta y_t = y_t - y_{t-1} &= \nu - FE(\infty)^{-1}y_{t-1} + u_t, \\ y_{t-1} &= FE(\infty)\nu + FE(\infty)(-\Delta y_t) + FE(\infty)u_t \end{aligned} \quad (1.12)$$

Estimating (1.12) as a penalized regression permits us to regularize $FE(\infty)$ and with $A = I_N - FE(\infty)^{-1}$ we can back out the autoregressive coefficients. To the best of our knowledge, this representation is novel to the literature and we investigate its performance for sparse FEVDs in the succeeding simulation. Note that even though this representation is most likely not suited for general applications, it may prove valuable for specific analyses with long-run dependencies. Henceforth, we denote the LASSO regularization on this target as the geometric regularization.

Regularizing the covariance matrix

Recall that FEVDs are functions of the coefficient and covariance estimates. Since the covariance matrix estimate also suffers under high dimensionality, we regularize the covariance matrix to obtain better inputs for the FEVDs. A variety of methods originated from the literature on the regularization of covariance matrices; we introduce three of them. The following overview is formulated in terms of the mean zero random vector $\mathbf{u}$. Hence, before applying any of the following methods, it is advisable to demean the variables. In our simulation and application, observations of $\mathbf{u}$ will be the innovation series $\mathbf{u}_t$ corresponding to the regression residuals $\hat{\mathbf{u}}_t$ in the first stage.

Adaptive thresholding Thresholding methods are designed to mimic the ideas of shrinkage and selection of entries for sparse covariance matrices. Let $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_N)'$ be a N -variate random vector with covariance matrix $\Sigma = [\sigma_{ij}]_{N \times N}$, and assume an *i.i.d.* random sample $\{\mathbf{u}_1, \dots, \mathbf{u}_T\}$ from the distribution of $\mathbf{u}$. Our goal is to estimate the covariance matrix Σ with big N and small T . Start with the sample covariance matrix

$$\hat{\Sigma} = [\hat{\sigma}_{ij}]_{N \times N} := \frac{1}{T-1} \sum_{t=1}^T \mathbf{u}_t \mathbf{u}_t'.$$

Further, define the variance of the sample covariance's entries as

$$\theta_{ij} := \text{Var}(\mathbf{u}_i \mathbf{u}_j) = E \left[(\mathbf{u}_i \mathbf{u}_j - \sigma_{ij})^2 \right].$$

We can now interpret the sparse covariance estimation as a mean vector estimation. That is, an individual entry, $\hat{\sigma}_{ij}$, without the Bessel correction can be described as

$$\frac{1}{T} \sum_{t=1}^T u_{i,t} u_{j,t} = \sigma_{ij} + \sqrt{\frac{\theta_{ij}}{T}} z_{ij},$$

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 by its sample counterpart $\hat{\theta}_{ij} = T^{-1} \sum_{t=1}^T [u_{i,t} u_{j,t} - \hat{\sigma}_{ij}]^2$. Cai and Liu (2011) propose calculating entrywise thresholds by

$$\lambda_{ij} := \lambda_{ij}(\delta^{AT}) = \delta^{AT} \sqrt{\frac{\hat{\theta}_{ij} \log N}{T}}, \quad (1.13)$$

where δ^{AT} is a strictly positive regularization parameter that determines the degree of the penalization. This hyperparameter has to be selected by the econometrician, or as

the authors suggest, can be set to 2. Finally, the adaptive (entry dependent) threshold estimator

$$\hat{\Sigma}^{AT} = [\hat{\sigma}_{ij}^{AT}]_{N \times N} = [s_{\lambda_{ij}}(\hat{\sigma}_{ij})]_{N \times N} \quad (1.14)$$

allows for different threshold levels for the entries and incorporates the variability of the entries, where the function $s_{\lambda}(\cdot)$ describes the thresholding rule applied to the entries of the sample covariance.

Note that the researcher can choose the thresholding functions $s_{\lambda}(\cdot)$ to determine the type of the penalty. In particular, we consider the adaptive-thresholding rule $s_{\lambda}(\hat{\sigma}_{ij}) = \hat{\sigma}_{ij}(1 - |\lambda_{ij}/\hat{\sigma}_{ij}|^{\eta})_+$, with $(\cdot)_+ = \max\{\cdot, 0\}$.⁶ Unlike in the case of the adaptive elastic-net estimator, λ_{ij} is a function described by (1.13), and thus, is not freely selectable in the adaptive-threshold estimator. With the adaptive-thresholding rule, the estimator achieves optimal convergence and performs better than the universal thresholding estimator that uses the same threshold for all entries. More generally, any rule that satisfies the conditions described in Appendix 1.A.1 achieves optimal convergence.

Cai and Liu (2011) suggest a thresholding rule that applies to all entries in the covariance matrix. In simulations, we experienced zero entries on the diagonal of the threshold estimates. Since this behavior contradicts the idea of covariances, we regularize only the off-diagonals entries.

Ledoit and Wolf's 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, it is the optimal convex linear combination of the sample covariance and the identity. The Ledoit–Wolf shrinkage estimator optimizes

$$\hat{\Sigma}^{LW} = \delta^{LW} m I_N + (1 - \delta^{LW}) \hat{\Sigma}, \quad (1.15)$$

where $\hat{\Sigma}$ denotes the sample covariance matrix, $m = N^{-1} \text{tr}(\hat{\Sigma})$ is the average of the diagonal values of the sample covariance matrix, and δ^{LW} is a predefined shrinkage weight. In contrast to Ledoit and Wolf (2004), we stick to the scaled sample covariance matrix for $\hat{\Sigma}$; that is, we divide the sum of squared residuals by $T - 1$ instead of T . This scaling is negligible for sufficiently large T .

⁶This rule incorporates soft thresholding for $\eta = 1$: $s_{\lambda}(\hat{\sigma}_{ij}) = \text{sgn}(\hat{\sigma}_{ij})(|\hat{\sigma}_{ij}| - \lambda)_+$. Cai and Liu (2011) describe this rule for $\eta \geq 1$. We will relax this condition and include values between zero and one in later steps.

To achieve asymptotic optimality, the authors introduce a *bona fide* version, that is, one that does not require additional knowledge:

$$\begin{aligned}\delta^{LW} &= \frac{b^2}{d^2}, \\ d^2 &= \|\hat{\Sigma} - mI_N\|^2, \\ b^2 &= \min \left[T^{-2} \sum_{t=1}^T \|u_t u_t' - \hat{\Sigma}\|^2, d^2 \right].\end{aligned}$$

The estimator is a linear shrinkage estimator, which optimally mixes the “all-bias no-variance” estimator mI_N with the “all-variance no-bias” estimator $\hat{\Sigma}$. Similar to (1.13) in the adaptive threshold case, the term $\delta^{LW} = b^2/d^2$ automatically assigns more weight to mI_N if the sample’s second-order variance, measured by b^2 , is large.

GLASSO Friedman et al. (2008) propose estimating 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

$$\hat{\Theta}^{GLASSO} = \underset{\Theta \geq 0}{\operatorname{argmax}} \left\{ \log \det \Theta - \operatorname{tr}(\hat{\Sigma}\Theta) - \delta^{GL} \|\Theta\|_1 \right\} \quad (1.16)$$

over all nonnegative definite matrices Θ , where $\|\cdot\|_1$ denotes the L_1 norm, that is, 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 corresponding 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 assume sparsity in the partial correlations rather than in the overall correlation structure. In other words, it is easier to rule out direct than indirect effects. In order to have more homogeneous penalization weights, we adapt this version for standardized observations without penalizing the diagonal.

1.2.3 Data-driven choice of regularization parameters

Except for the *bona fide* Ledoit–Wolf shrinkage estimator, all of the above estimators require the choice of regularization parameters.⁷ In particular, techniques related to the elastic-net require the choice of λ^{NET} and α , and the regularized covariance estimators require the choice of δ^{AT} , η , δ^{LW} , and δ^{GL} . Each choice of regularization parameters then results in a different set of parameter estimates and, therefore, a different model. Except for Ledoit and Wolf (2004)’s optimized value for the shrinkage weight δ^{LW} , we need to manually determine optimal levels of regularization. To validate the goodness-of-fit of the fitted candidate models, we resort to statistical learning methods. Specifically, we use cross-validation (CV) in our application.

In general terms, model validation procedures divide the sample into training and test data.⁸ Given fixed hyperparameters, these methods estimate the remaining parameters from the training data and validate their performance on the test data. That is, the quality of the estimate from the training data set is assessed by its ability to explain the test data set. The model selected after validation is the version that best explains the test sample. A common practice for small data sets is K -fold CV, which partitions the sample into K equally-sized test samples.⁹ For each of these K samples, CV trains the estimator on the remaining training data and then validates the test sample performance through a metric such as the MSE or the negative log-likelihood. In the context of time-series data, it is advisable to use (nonrandom) block sampling, as described by Figure 1.A.8 in Appendix 1.A.2. The validity of this approach is shown in Bergmeir et al. (2018) and we consequently stick to K -fold CV with block sampling.

CV for the coefficients commonly uses the MSE to select the best fitting model. That is, the MSE serves both as the optimization loss in the training sample and as the validation metric in the test sample. In comparison, there exists no convention for evaluating covariance estimators because there is no observed target for the residual covariance matrix. Cai and Liu (2011) propose using the sample covariance estimate of the k -th test sample to validate the goodness-of-fit. By the motivation of our paper, however, this loss function is imprecise since the sample covariance is a bad estimate of the test sample. Thus, we prefer to validate the trained estimate $\widehat{\Sigma}_{\{1:T\}\setminus\kappa}(\delta)$ directly

⁷The literature often refers to such parameters as hyperparameters as they must be fixed before estimating the other parameters.

⁸The statistical learning literature uses the term “validation set” to denote the hold-out sample in each CV fold. In econometric applications, however, it is common to denote the subset of the data used for hyperparameter tuning as the “test data.” We stick to the latter terminology throughout our paper.

⁹Common choices for K range from 5 to 12, where larger values increase the computational burden and small values entail higher estimation uncertainty (Hastie et al., 2009, Chapter 7.12). The special case of leave-one-out CV sets K equal to the sample size, validating the estimator on a single observation.

on the squared observations $u_t u_t'$ in the k -th test sample κ ,

$$\ell_k(\delta) = \frac{1}{T_k} \sum_{t \in \kappa} \|\widehat{\Sigma}_{\{1:T\} \setminus \kappa}(\delta) - u_t u_t'\|^2, \quad \text{for } k = \{1, \dots, K\},$$

where the first term of the difference is the training data estimate of Σ with penalization parameter δ , and T_k denotes the sample size of κ . In contrast to the version of Cai and Liu (2011), this loss function shows the mean of the distance between the squared observations and the trained estimate.

For computational reasons, econometricians are limited to a small grid of candidate hyperparameters, as there are K estimations for each combination. Thus, we aim to produce a set of candidate hyperparameters in a neighborhood of the presumed minimum loss to then perform a grid search. Similar to Kascha and Trenkler (2015), we increase the candidate hyperparameters linearly on a log scale between data-driven minimum and maximum values.¹⁰

At this point, it is worth mentioning an obvious extension. Since CV selects penalty terms on predefined sample splits, we can also plug in different regularization techniques for the estimators. CV then selects the estimator with the best predictive power. Alternatively, information criteria such as the Akaike Information Criterion (AIC) provide another way to choose among substantially different methods. We apply the latter procedure to our empirical setting in Section 1.4.

1.3 Simulation study

1.3.1 Data Generating Processes

In the context of FEVDs, we believe that sparsity emerges 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. Although this simulation study is limited to VAR(1) models, it also extends to higher lag orders in companion form. Estimating models with additional lags will, however, aggravate the high-dimensionality problem. To resemble real-world observational data, the *DGPs* include sparse coefficients in the VAR matrix A and innovation covariance Σ .

DGP 1: Diagonal VAR coefficient (autocorrelation without spillovers)

$$A = 0.5I_N.$$

¹⁰Friedman et al. (2010) provide formulas to set the minimum and maximum hyperparameters for general elastic-nets.

$$\Sigma = I_N.$$

DGP 2: Diminishing-diagonal VAR coefficient (approximately sparse FEVD)

A is a banded diagonal matrix with entries $a_{ij} = 0.3^{|i-j|+1}$.

$$\Sigma = I_N.$$

DGP 3: Random graph (potentially dense FEVD)

A is a random sparse matrix. That is, it has entries with probability $P(a_{ij} \neq 0) = \tau$. $\tau = 1/\sqrt{N}$ denotes the degree of density. We set nonzero entries to one and rescale them uniformly such that A 's maximum eigenvalue is 0.5 in modulus.

$\Sigma = SD S'$. Here, $S = (I_N - B)^{-1}$ where B is an upper triangular matrix with entries $b_{ij} = \tau$ if $a_{ij} \neq 0$, $\forall i > j$. $D = \text{diag}(1, \dots, 2)$ is an $N \times N$ diagonal matrix with values spaced equally between 1 and 2. Put differently, the 0-period impact is a directed acyclical graph defined by the upper triangular matrix B .

DGP 4: Block-diagonal FEVD (sparse FEVD)

A has a block-diagonal structure with $1/\tau$ equal-sized quadratic blocks. $\tau = 1/\sqrt{N}$ denotes the degree of density. We set nonzero entries to one and rescale uniformly them such that A 's maximum eigenvalue is 0.5 in modulus.

$\Sigma = SD S'$. Here, S is the long-run dependency spanned by A , that is, $S = (I_N - A)^{-1}$. $D = \text{diag}(1, \dots, 2)$ is an $N \times N$ diagonal matrix with values increasing from 1 to 2.

Note that *DGP*s 1 and 4 always produce a sparse FEVD while *DGP*s 2 and 3 do not. Moreover, all but *DGP* 3 are deterministic and *DGP*s 1-2 have the identity as the covariance matrix. *DGP* 3 has 25 different random realizations in the simulation. We compare the methods' performances for FEVDs with forecast horizon 10.

1.3.2 Comparison of different regularization methods

This section analyzes the relative performance gain of using regularization methods to estimate FEVDs. That is, we compare the estimation errors of the various regularization methods to the case of OLS plus the sample covariance. In particular, we run simulations for the aforementioned *DGP*s and calculate the Frobenius norm of the regularized versus the nonregularized version: $\|\widehat{D}_{reg}^{gH} - D^{gH}\| / \|\widehat{D}_{OLS}^{gH} - D^{gH}\|$, where $H = 10$. We compare the performance gains over $N = \{50, 150, 250\}$ and $T = \{75, 175, 500\}$. Note that OLS breaks down for $N > T$ such that we are not able to calculate any value for these cases.¹¹ Since the estimation is a two-step procedure, we split the simulation into two parts. First, we regularize A paired with the sample covariance for Σ . We compare the ridge, LASSO, adaptive elastic-net, and geometric long-run regularization. The latter uses the LASSO penalty to perform variable selection. We choose the penalty parameters λ^{NET}

¹¹In general, regularization methods are not limited to $N < T$.

such that they minimize the respective norm $\lambda^* = \operatorname{argmin}_\lambda \|\widehat{D}_{reg}^{gH}(\lambda) - D^{gH}\|$. Thus, the values show the best possible performance gain.

<i>DGP1</i>		Ridge			LASSO			AENET			Geometric		
$N \backslash T$		75	175	500	75	175	500	75	175	500	75	175	500
50	21.5%	22.5%	22.6%	21.6%	22.7%	22.6%	21.5%	22.6%	22.6%	22.4%	23.1%	22.8%	
150		12.3%	13.8%		12.3%	13.8%		12.3%	13.8%		12.6%	13.9%	
250			10.8%			10.8%			10.7%			10.9%	
<i>DGP2</i>		Ridge			LASSO			AENET			Geometric		
$N \backslash T$		75	175	500	75	175	500	75	175	500	75	175	500
50	23.6%	23%	23.9%	23.3%	22.9%	25.0%	23.6%	23.0%	24.5%	29.8%	29.9%	35.4%	
150		13.7%	14.5%		13%	14.1%		13.6%	14.2%		17.2%	19.2%	
250			11.7%			11.1%			11.2%			15.2%	
<i>DGP3</i>		Ridge			LASSO			AENET			Geometric		
$N \backslash T$		75	175	500	75	175	500	75	175	500	75	175	500
50	27.9%	36%	49.1%	27.8%	35.8%	49.7%	27.6%	35.6%	48.9%	33.2%	45.5%	77.5%	
150		14.2%	18.6%		14%	18.4%		13.9%	18.5%		15.5%	22.7%	
250			12.7%			12.6%			12.6%			14.1%	
<i>DGP4</i>		Ridge			LASSO			AENET			Geometric		
$N \backslash T$		75	175	500	75	175	500	75	175	500	75	175	500
50	30.7%	39.2%	55.7%	30.5%	39.6%	57.8%	30.3%	39.3%	57.5%	43.3%	62.7%	111.5%	
150		16.7%	24.7%		16.4%	24.9%		16.4%	24.8%		24.5%	45.3%	
250			12.4%			12.3%			12.3%			15.1%	

Table 1.1: Simulation results for the regularization of A paired with the sample covariance matrix estimate of Σ . Values represent the relative Frobenius norm of the FEVD estimates to the true generalized connectedness tables, when compared to the OLS estimates as $\|\widehat{D}_{reg}^{gH} - D^{gH}\| / \|\widehat{D}_{OLS}^{gH} - D^{gH}\|$. This number states the relative sizes of the estimation errors when compared to the OLS estimate. For example, a value of 20% signifies that the estimator makes only 20% of the error of the OLS estimates. We present such values for various N s and T s with 500 Monte Carlo repetitions. *DGP 3* has 25 different random realizations of A and Σ . Ridge denotes the ridge regression and AENET denotes the adaptive elastic-net.

Table 1.1 contains the simulation results for the regularization of A for 500 Monte Carlo repetitions. First, it is evident that regularization achieves a large overall efficiency gain. The largest 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 best regularized estimators achieve reductions in the norms to the true values to 10.7 – 23.9% for *DGPs 1-2*, and to 12.3 – 55.7% for *DGPs 3-4*. Surprisingly, ridge, LASSO and adaptive elastic-net perform similarly well. There is no clear winner among the estimators since the differences in performance are marginal. If at all, the variable selection capabilities of LASSO and adaptive elastic-net provide a slight performance advantage over the pure shrinkage of ridge regression for simulations where T is close to N . Finally, the geometric regularization of long-run effects underperforms other regularization methods for all *DGPs*, and does worse than the unregularized estimation for $N = 50$ and $T = 500$.

As a second step, we compare regularization methods for Σ . That is, we calculate the residuals using the adaptive elastic-net estimate and construct the FEVD with the (regularized) estimate of Σ . We choose the adaptive elastic-net penalty parameter to minimize the Frobenius norm of the difference between the estimate matrix $\hat{A}$ and the true parameter matrix A : $\lambda^{NET*} = \text{argmin}_{\lambda} \|\hat{A}(\lambda) - A\|$. Additionally, we set the respective penalty parameter for the covariance regularization such that $\delta^* = \text{argmin}_{\delta} \|\hat{D}_{reg}^{gH}(\delta) - D^{gH}\|$.

<i>DGP1</i>												
$N \backslash T$	Sample-Cov			Threshold			Ledoit-Wolf			GLASSO		
	75	175	500	75	175	500	75	175	500	75	175	500
50	26.1%	18.7%	15.4%	5.2%	4.7%	0.5%	6.9%	4.9%	0.6%	5.3%	4.7%	13.2%
150		14.9%	10.8%		3.8%	1.0%		4.3%	1%		3.8%	1.0%
250			9.5%			1.1%			1.1%			1.1%
<i>DGP2</i>												
$N \backslash T$	Sample-Cov			Threshold			Ledoit-Wolf			GLASSO		
	75	175	500	75	175	500	75	175	500	75	175	500
50	24.3%	23.1%	27.4%	3.4%	8.4%	19.8%	3.6%	8.5%	19.9%	3.4%	8.6%	25.6%
150		14.4%	14.8%		2.8%	7.7%		2.8%	7.7%		2.8%	7.7%
250			11.5%			4.6%			4.6%			4.6%
<i>DGP3</i>												
$N \backslash T$	Sample-Cov			Threshold			Ledoit-Wolf			GLASSO		
	75	175	500	75	175	500	75	175	500	75	175	500
50	28.7%	39.2%	59.4%	14.6%	29.7%	59.2%	14.1%	29.9%	62.4%	14.5%	28.8%	59.8%
150		14.4%	20.4%		4.8%	13.4%		4.6%	13.9%		4.7%	13.5%
250			13.2%			6.1%			6.1%			6.1%
<i>DGP4</i>												
$N \backslash T$	Sample-Cov			Threshold			Ledoit-Wolf			GLASSO		
	75	175	500	75	175	500	75	175	500	75	175	500
50	32.3%	41.5%	61.9%	23.8%	38.4%	60.0%	25.5%	46.9%	84.6%	23.4%	38.4%	62.3%
150		18.1%	28.8%		12.5%	26.4%		14.5%	34.9%		12.1%	26.3%
250			12.8%			5.9%			6.2%			5.8%

Table 1.2: Simulation results for the regularization of Σ , paired with the best-performing adaptive elastic-net estimator for A . Values report the relative Frobenius norm of the FEVD estimates to the true connectedness tables, when compared to the OLS estimator with the sample covariance matrix as $\|\hat{D}_{reg}^{gH} - D^{gH}\| / \|\hat{D}_{OLS}^{gH} - D^{gH}\|$. This number states the relative sizes of the estimation errors when compared to the OLS and sample covariance estimate. For example, a value of 20% signifies that the estimates make only 20% of the error of the unregularized estimates. We present such values for various N s and T s with 500 Monte Carlo repetitions. *DGP 3* has 25 different random realizations of A and Σ .

Table 1.2 shows the simulation results for the different regularizations. The first estimator is the sample covariance matrix and sets the benchmark. Again, we measure the performance of the regularization methods with the norm of the estimated FEVD to the true values relative to the unregularized estimates. For *DGP1* with the identity as the data generating covariance matrix, the regularization methods provide a near perfect approximation of the FEVD as they favor diagonal entries. When *DGP 2* adds autocorrelation, all regularization methods perform similarly, with the exception of the GLASSO estimator for $N = 50$ and $T = 500$. In contrast, *DGPs 3* and *4* have nondiagonal covariance matrices. In *DGP 3*, the estimators perform similarly, with

each of them outperforming others for two combinations. We believe this result is due to chance, and thus we do not declare a winner. Finally, the GLASSO estimator outperforms the other regularization techniques in *DGP 4* by a small margin. The only exception is for $N = 50$ and $T = 500$, where only the adaptive-threshold estimator surpasses the sample covariance.

1.3.3 Bias, variance, and edge detection

Next, we investigate the mean of the entries and the norms for regularized and unregularized FEVDs. For this purpose, we simulate time-series with *DGP 4*, $N = 100$ and $T \in [100, 500]$. The analysis compares OLS and the sample covariance to the adaptive elastic-net in combination with the sample covariance matrix, GLASSO, and the adaptive threshold estimator, respectively. We select the hyperparameters based on the best performance (minimization of the Frobenius norm to the true FEVD). Note that this procedure requires knowledge about the true parameters, such that it is applicable only in a simulation exercise. The left-hand panel in Figure 1.2 shows the average mean distance, $N^{-2} \sum_{i=1}^N \sum_{j=1}^N \text{mean}(\widehat{D}_{ij}^{gH} - D_{ij}^{gH})$, which reflects the biases of the FEVD estimates. The right-hand panel of Figure 1.2 then shows the average variance $N^{-2} \sum_{i=1}^N \sum_{j=1}^N \text{var}(\widehat{D}_{ij}^{gH})$. Note that this analysis relates to the “bias-variance” trade-off for estimators as regularization methods generally sacrifice unbiasedness to achieve a lower variance and, therefore, a reduction in prediction errors overall.

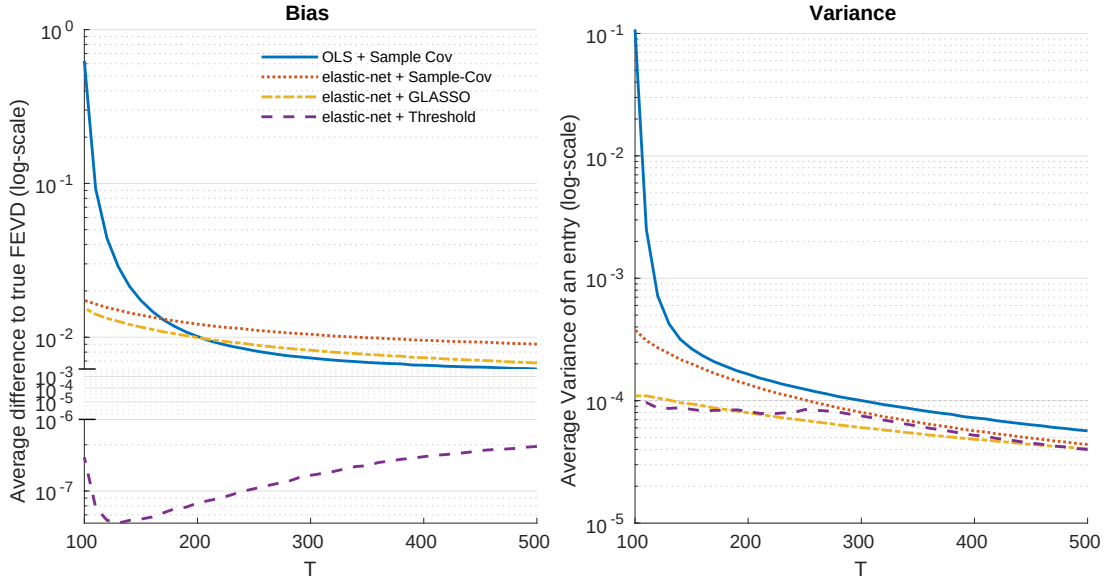


Figure 1.2: Simulation results for 500 Monte Carlo repetitions of *DGP 4* and $N = 100$. The left-hand panel shows the average mean difference of the estimates to the true values $N^{-2} \sum_{i=1}^N \sum_{j=1}^N \text{mean}(\widehat{D}_{ij}^{gH} - D_{ij}^{gH})$. The right-hand panel shows the respective average variance $N^{-2} \sum_{i=1}^N \sum_{j=1}^N \text{var}(\widehat{D}_{ij}^{gH})$. The sample size T is on the x-axis.

The left-hand panel depicts the magnitude of the bias for small T (approaching $N = 100$ from the right). We see that on average all estimators are positively biased, with the nonregularized estimator (solid/blue line) profoundly overestimating entries in the FEVDs. That is, this estimator faces a strong positive bias the closer T is to N . Perhaps surprisingly, the adaptive elastic-net (red dotted curve) already diminishes the bias for small T by a margin (almost by a factor of 100 for $T = 100$). While we expected that regularization methods trade-off bias for variance in FEVD tables, adaptive elastic-net also improves with respect to the bias for $T < 200$. Similarly, when using regularizations on the covariance matrix, we see a consistent improvement with respect to the bias. The threshold estimate improves strongly and the GLASSO estimate improves slightly over its sample covariance counterpart, even for larger T .

The right-hand panel plots each estimator's variance, indicating the precision of the estimation. The nonregularized version not only faces heavy inaccuracies following from its bias but also shows an extremely large variance for $T < 150$. Its variance lessens with increasing T but still underperforms compared to the regularized versions. For all regularization methods, we see gains in the variance, likely stemming from the “bias-variance” trade-off. The combination of coefficient and covariance regularization therefore dominates for all T s. Pairing this finding with the findings of the left-hand panel, there appears to be no trade-off for $T < 200$ but an overall improvement in bias and variance. Summing up, it is advisable to combine the regularizations for the coefficient and the covariance matrices as this combination provides a lower variance, and for most sample sizes, also comes with a lower bias.

Next, in the context of networks, we explicitly care about the diagnostic ability of the estimator. That is, we are interested in how well the classification into zero and nonzero entries performs in the network matrix. For that purpose, we summarize the performance in terms of the correct detection of zero and nonzero entries in the three panels of Figure 1.3. A value is considered true positive (TP) in the case of a hit and false positive (FP) in the case of a false alarm (Type I error). Likewise, a true negative (TN) is given if the FEVD is correctly estimated to be sparse at a given edge, and there is a false-negative (FN) when an existing edge is not found (Type II error). First, the probability of correct classification is summarized by the accuracy metric. Accuracy is defined as the fraction of correct predictions: $accuracy = (TP + TN)/N^2$. We compare this metric of the estimators for increasing T in Figure 1.3. The left panel shows that using the adaptive elastic-net estimator instead of OLS unlocks a big improvement in small samples. A small but additional gain can be achieved through the usage of GLASSO instead of the sample covariance matrix, while the adaptive threshold estimator does not show any improvements.

The receiver operating characteristic (ROC) plots the false-positive rate $FPR = FP/(FP + TN)$ against the true-positive rate $TPR = TP/(TP + FN)$ while varying the discrimination threshold of setting values to zero. More precisely, the threshold varies from the lowest to the highest entry in the FEVD, and thus, sets increasingly more values to zero. For each threshold, the ROC plots the respective FPR and TPR in a diagram ranging from 0 to 1. A perfect estimator — that is, one that correctly classifies all edges no matter the threshold — would result in a line starting at $(x,y)=(0,1)$ and ending at $(x,y)=(1,1)$. Conversely, a completely random guess would be shown as a 45° diagonal line. We plot the ROC curve for $T = 150$ in the center Panel of Figure 1.3. Again, the regularized estimators clearly improve the performance when it comes to classification, with the combination of adaptive elastic-net and GLASSO providing the best diagnostic capabilities.

Finally, the right panel plots the area under the ROC curve (AUROC). The values range from 0.5 to 1, where 0.5 is a completely random guess and 1 is the perfect classification. This metric summarizes an estimate's performance in detecting edges in a single number and lets us compare estimators for different T s. The regularized estimators display high confidence in their classifications and achieve better performance, in particular for T close to N . While the gains over the unregularized version mainly originate from the regression stage of the estimation, the results improve even further after regularizing the covariance matrix with GLASSO.

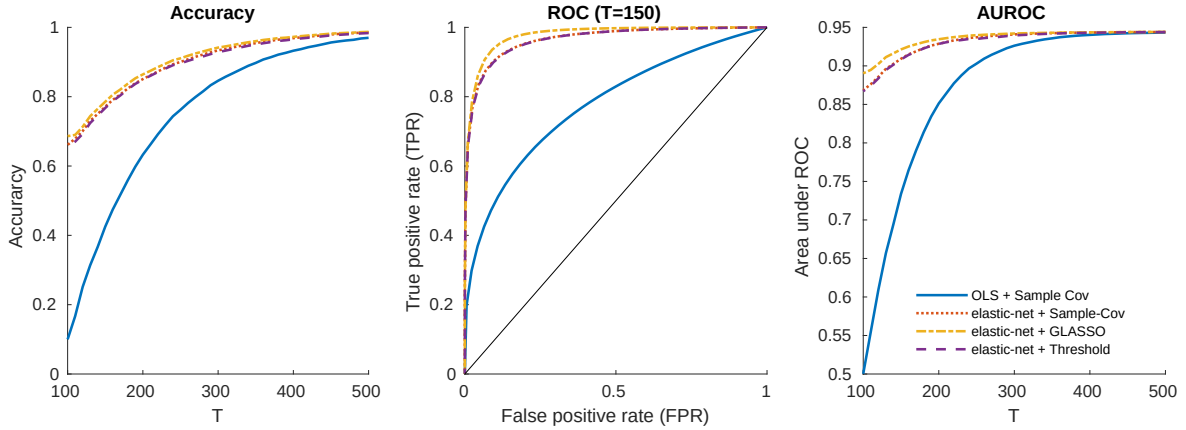


Figure 1.3: Simulation results for 500 Monte Carlo repetitions of *DGP 4* and $N = 100$. The first panel shows the accuracy for increasing sample size T on the x-axis. The center panel shows the receiver operating characteristic (ROC) for $T = 200$, which has FPR on the x-axis and TPR on the y-axis. The diagonal thin black line is the equivalent of a random estimate. The right-hand panel shows the area under the ROC (AUROC) for increasing sample size T .

To sum up, our simulations demonstrate that the regularization of both estimation steps leads to a substantial improvement in the estimation of FEVDs. This is evident when comparing estimation errors as well as when using entrywise classification metrics.

For large N and small T , our results indicate that it is critical to regularize not only the coefficient but also the covariance matrix. Perhaps surprisingly, there are no trade-offs when N approaches T with performance gains in all metrics. However, the use of regularization also shows improvements when estimating large-dimensional networks from *large- T* data sets and, therefore, we advise researchers to apply regularization techniques when faced with large-dimensional estimation problems. Comparing the covariance estimators, the adaptive threshold estimator appears to best reduce the estimation bias, while GLASSO mainly excels in its diagnostic ability. Thus, we advise researchers to pick the method based on the desired properties for the specific research question. Alternatively, model selection techniques such as out-of-sample CV or in-sample information criteria may help to validate the methods' performances on specific data.

1.4 Empirical application: Production volatility spillovers and the Great Moderation

The period known as the Great Moderation, starting in the mid-1980s, is characterized as a period with reduced fluctuations in many macroeconomic time-series, such as real growth rates, industrial production (IP), and unemployment. The question whether the Great Moderation happened due to good luck, better monetary policy, or any other structural change is vital for policymakers as they need to understand the impact of their actions on macroeconomic volatility.¹² The large volatility of IP is particularly puzzling since the aggregated IP index sums over many weighted sector-level shocks. Thus, Foerster, Sarte, et al. (2011) investigate this shift by decomposing IP into sectoral and common shocks and conclude that a decline in the volatility of common shocks induced most of the break in IP's volatility. Conversely, Carvalho and Gabaix (2013) provide evidence that the shift in macroeconomic volatility could have originated from idiosyncratic microeconomic shocks on the sectoral level, offering a contrasting explanation that leaves the question of the origins of the Great Moderation all but conclusively answered.

The estimation of sectoral connectedness tables allows us to gain a new perspective on the source of the decline in the volatility of aggregate IP. While the findings in Foerster, Sarte, et al. (2011) are in accordance with much research that attributes the Great Moderation to declines in the shock volatility of common exogenous factors such as monetary policy (for example, Leduc and Sill (2007), Justiniano and Primiceri

¹²For early discussions of the potential drivers of the Great Moderation see Stock and Watson (2003) and Bernanke (2004).

(2008)), factor productivity (for example, Arias et al. (2007)) or oil supply (for example, Nakov and Pescatori (2010)), an alternative explanation has emerged from the granular hypothesis of Gabaix (2011). The latter paper proposes that a small number of constituents can explain aggregate shocks if the index weight distribution is fat-tailed, an idea the findings of Foerster, Sarte, et al. (2011) refute. In our view, one needs to consider not only sectoral weights but also directional dependencies to analyze the influence of a particular industry on the aggregate. That is, we entertain the possibility that correlations between sectors originate from a strong spillover network of sector-level innovations rather than common shocks as in Foerster, Sarte, et al. (2011). Acemoglu et al. (2012) establish this idea based on input–output linkages and show that the diversification argument does not apply in the presence of strong network structures. More recently, Foerster, Hornstein, et al. (2022) give additional leeway to this hypothesis by attributing much of trend GDP growth rates to sector-specific factors. In a nutshell, we investigate if structural change of such a network has been an important driver of the Great Moderation, which is in line with supply-chain and inventory-based explanations presented in Kahn et al. (2002), Summers et al. (2005) and Davis and Kahn (2008).

For insights into sectoral interconnectedness, we obtain estimates of the directional dependencies. In contrast to a fast-expanding literature on production networks that exploits input–output tables as network proxies (for a review see Carvalho and Tahbaz-Salehi (2019)), our methodology directly sheds light on the network implied by sectoral IP correlations. Our approach is in line with the finding of Carvalho (2014) showing that pairwise correlations between two sectors are also bigger for stronger input–output links. An advantage of inferring the network directly from the output data is the ability to capture not only the supply–demand relationships of economic sectors but also alternative channels of transmission. For example, the propagation of production volatility between two sectors goes beyond existing input–output relations if they compete for the same resources, if their outputs are substitutes or complements, or if poor inventory management amplifies demand shocks and their propagation up the supply chain. Thus, by estimating FEVD tables, our application is agnostic about the channel of propagation and adds to the understanding of sectoral spillovers and their contribution to aggregate fluctuations.

We examine pre- and post-1984 periods in detail to give a comparison of how spillovers have changed. To address the possibility of exogenous macroeconomic shocks, we analyze the spillover networks while factoring out macroeconomic variation. Our findings provide novel empirical evidence and offer a unifying view on the otherwise opposing narratives of Foerster, Sarte, et al. (2011), Gabaix (2011) and Acemoglu et al.

(2012). We conjecture that sizeable spillovers from a handful of sectors initially generated strong sectoral comovements, which appear in other analysis as large aggregate shocks. With the Great Moderation, the structure of the spillover network changed such that sectoral shocks are less likely to amplify into aggregate volatility. Hence, our findings raise an alternative explanation in which structural change largely contributed to the Great Moderation. Finally, an analysis of the increase in production volatility around the Great Recession does not reveal a reversal of the initial structural change, but rather attributes a growing contribution share to macroeconomic factors.

1.4.1 Data

Similar to Foerster, Sarte, et al. (2011), we use sectoral data on IP throughout the period 1972–2020. We mainly analyze the three-digit industry classification of the North American Industry Classification System (NAICS) with $N = 88$ sectors, whereas the data spans up to $N = 138$ sectors corresponding to the five-digit industry classification. 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 four equally-sized subsamples of $T = 144$ months. The subsamples span from 03/1972 to 02/1984, from 03/1984 to 02/1996, from 03/1996 to 02/2008, and from 03/2008 to 02/2020. In our analysis, the boundary between the first and second samples marks the onset of the Great Moderation. For simplicity, we label these samples as 1972–1983, 1984–1995, 1996–2007, and 2008–2019, respectively.

Let $IP_{i,t}$ denote the value of IP 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 weights $w_{i,t}$. Our first subsample, from 1972 to 1983, coincides with the pre-Great Moderation period and the other three subsamples are post-Great Moderation. The subsample IP volatilities of 11.67%, 5.85%, 6.36%, and 8.79% illustrate that the average monthly volatility of aggregate IP diminished with the Great Moderation and stayed fairly constant thereafter with the exception of the Great Recession of 2007–2009.¹³

To account for the possibility of exogenous factors affecting all sectors simultaneously and thereby inducing the illusion of a strong spillover network, we condition on three contemporaneous and lagged factors that the literature has brought forward as possible explanations for the Great Moderation. First, we include the monthly monetary policy shock series g_t^{MP} of Wieland and Yang (2020) based on Romer and Romer (2004)

¹³See Figure 1.A.9 in Appendix 1.A.4 for a visual display of the growth rates of IP on an aggregate level.

for the first three subsamples and a similar version of Bauer and Swanson (2022) for the most recent subsample. Second, we consider the possibility of exogenous productivity shocks through the inclusion of the updated interpolated utilization-adjusted TFP series g_t^{TFP} of Fernald (2014). And third, we account for commodity supply shocks by adding percentage changes of the S&P GSCI Commodity Index g_t^{COM} to our data.

1.4.2 Estimation

We interpret the spillover constituent in the data as a VAR(1) model and infer connectedness from monthly cross-autocorrelations of $g_{i,t}$. The higher frequency allows setting the forecast horizon to 3 months, which corresponds to the (undirected) covariance matrix of the quarterly data. This connection between monthly and quarterly frequencies provides insights into the contagion within a quarter. In particular, the average pairwise correlations and aggregate shocks in Foerster, Sarte, et al. (2011) may be better understood if we break up quarterly volatilities into three serially correlated monthly volatilities. Hence, the central regression specification is as follows:

$$\begin{aligned} y_t &= \mu + Ay_{t-1} + \sum_{l=0}^{12} B_l x_{t-l} + u_t, \quad \forall t = 1, \dots, T, \\ y_t &= [g_{1,t}, \dots, g_{N,t}]', \\ x_t &= [g_t^{MP}, g_t^{TFP}, g_t^{COM}]', \\ u_t &\sim \mathcal{N}(0, \Sigma). \end{aligned}$$

While Foerster, Sarte, et al. (2011) test hypotheses about correlations through common factors, our framework aims to see correlations through the lens of intersectoral spillovers in A and Σ . Nevertheless, we include the factors x_t in our analysis to account for the original authors' results concerning common variation. As control variables, they serve as a means to rule out that the estimated connectedness tables falsely ground on common innovations that the literature has investigated as explanatory for the Great Moderation.¹⁴

We regularize A , $\{B_l\}_{l=0}^{12}$ and Σ with the techniques mentioned in the previous sections. Since the simulations did not point toward a consistent winner throughout all settings, we validate all regularization methods on the data. Namely, we run a 12-fold CV as in Section 1.2.3 to select the hyperparameters α and λ^{NET} in the regularized regressions, and δ and η in the covariance estimation. Here, the best estimate of A from the regression stage serves to obtain the residuals for the estimation of Σ . Note

¹⁴Note that confirmatory factor analysis with observable factors aids us in answering our main research question, whereas latent factor techniques like principal components would not allow to distinguish between factor narratives and the granular network story.

	MSE				AIC
	1972- 1983	1984- 1995	1996- 2007	2008- 2019	1972- 2019
<i>Panel A: Model selection for regression</i>					
OLS	100%	100%	100%	100%	Inf
Elastic-Net	7.87%	7.40%	7.51%	6.74%	4.863
Adaptive Elastic-Net (OLS init.)	7.71%	7.34%	7.36%	6.58%	4.956
Adaptive Elastic-Net (ENet init.)	6.13%	6.12%	6.22%	5.33%	4.649
<i>Panel B: Model selection for covariance matrix</i>					
Sample Covariance	100%	100%	100%	100%	4.649
GLASSO	99.70%	99.49%	99.56%	99.55%	4.592
Ledoit-Wolf	99.81%	99.85%	99.85%	99.80%	4.812
Adaptive Threshold	99.73%	99.51%	99.57%	99.58%	4.613

Table 1.3: Summary of the model selection and with cross-validated hyperparameters. MSE is the out-of-sample CV loss, denoted as a percentage of the respective unregularized candidate model (OLS and sample covariance). AIC is the Akaike information criterion for the dynamic model fit with scale 10^5 . Panel A reports values for candidate models of the coefficient matrix and using the sample covariance matrix as the innovation covariance matrix. OLS overfits, such that the log-determinant of the residual covariance matrix is 0, yielding infinity in the AIC column. Panel B contains the same comparison for the residual covariance matrix estimators.

that although the regression residuals are fed into the estimation of the covariance matrix, the two estimation steps are independent, such that sparsity can emerge at different positions in the coefficient and covariance matrix. Then, similar to the model comparisons of Sims and Zha (2006), we choose among the optimized candidate models based on Akaike’s information criterion (AIC) over the full sample.¹⁵

Table 1.3 shows the out-of-sample MSEs and in-sample AICs for all optimized candidate estimators. First, Panel A reports the results for the regression stage of the estimation. Unsurprisingly, the adaptive elastic-net with initial elastic-net weights dominates all other estimation schemes. In our *large-N-small-T* setting, the OLS estimator even overfits the data to an extent that the log-likelihood is not defined with the residual covariance matrix being noninvertible. Second, Panel B reports the same statistics for the estimation of the residual covariance matrix after the first-step model selection. Here, regularized estimation methods again consistently outperform the sample covariance matrix in terms of MSE, while GLASSO provides the best model fit in terms of AIC throughout. For all subsequent analysis, we focus only on the selected

¹⁵As pointed out in Hurvich and Tsai (1989), AIC needs a second-order correction for small samples, that is, the *large-N-small-T* setting we operate in. We explored robustness with respect to outcomes of a corrected AIC as in Bedrick and Tsai (1994). More precisely, the correction term in the corrected AIC is bigger for dense models such that the model selection criterion more clearly favors sparse models.

AIC		A	
		static	dynamic
Σ	static	4.6900	4.6417
	dynamic	4.6444	4.5920

Table 1.4: Comparison between static and dynamic model choices evaluated for the Akaike information criterion (AIC). Dynamic refers to the sample being split into four equally-sized subsamples 1972–1983, 1984–1995, 1996–2007 and 2008–2019, each of size $T = 144$. Static refers to the full sample 1972–2019 with size $T = 576$. All modeling choices are 12-fold cross-validated and the best predictor model is selected according to the methods described in the main text. Values have scale 10^5 .

estimator, that is, the one that achieves the lowest AIC.¹⁶

Similar to Sims and Zha (2006), we additionally investigate whether the break in the sample is due to the coefficient matrix A or to the contemporaneous matrix Σ in Table 1.4. That is, we compare specifications that use static matrices estimated for the whole sample from 1972 to 2019 instead of dynamically allowing them to change per subsample. Naturally, dynamic specifications achieve better goodness-of-fit, but AIC penalizes them more strongly because of the higher use of degrees of freedom. Nevertheless, dynamic specifications for both A and Σ indeed improve the model quality. We see this finding as support for a change in the coefficient as well as the covariance matrix with the onset of the Great Moderation.

As structural identification of two-way causality in u_t proves difficult in the presence of high dimensionality, we rely on “generalized identification” to offer neutral insight into the connectedness of variables. With the inclusion of all production sectors in y_t and potential sources of common variation in x_t , the covariance of the innovation vector u_t becomes more sparse. Thus, the generalized approach gets more precise at approximating exogenous variability such that the generalized-*shock* covariance matrix Ω is indeed close to the identity. To support the hypothesis that we are not missing out on significant sources of exogenous variations, we will additionally conduct a formal test for Ω being statistically distinguishable from the identity matrix.

Finally, taking the estimates for A and Σ as inputs, we calculate the FEVDs with the forecast horizon $H = 3$. We row-normalize D^{gH} in (1.8) to show the percentage contribution to the variance. Additionally, we present key figures related to the network literature. In particular, we use the same measures as Diebold and Yilmaz (2014): in-, out-, and average connectedness. These measures are defined as the row sum, column

¹⁶Stone (1977) establishes the asymptotic equivalence of CV and AIC in terms of model choice.

sum, and the average row sum without the diagonal entries, respectively.¹⁷

$$C_{i \leftarrow \bullet} (D^{gH}) = \sum_{j \neq i} d_{ij}^{gH}, \quad (\text{in-connectedness to } i)$$

$$C_{\bullet \leftarrow j} (D^{gH}) = \sum_{i \neq j} d_{ij}^{gH}, \quad (\text{out-connectedness from } j)$$

$$C (D^{gH}) = \frac{1}{N} \sum_i \sum_{j \neq i} d_{ij}^{gH}. \quad (\text{average connectedness})$$

The former two measures are sector-specific measures, while the latter summarizes the overall explanatory power of connectedness as the average amount of sectoral variation that our model links to spillovers. Note that we are mainly interested in the distribution of the outgoing spillovers of the sectors measured by the out-connectedness. Precisely, if a handful of large sectors had a high out-connectedness, the volatility of those sectors' IPs would not average out in the aggregate IP index.

1.4.3 Results

Table 1.5 shows selected estimation statistics for our main specification alongside other levels of disaggregation for robustness. Two results on sectoral connectedness are evident in this table. First, the nonregularized version has a significantly higher average connectedness and stays remarkably constant over the subsamples. That is, it does not detect any change with the onset of the Great Moderation. In contrast, the regularized versions capture a clear difference between the pre- and post-Great Moderation subsamples with average spillovers of 46.4%, 28.0%, 25.8%, and 25.4% at the three-digit level, respectively. Second, the average results of the regularized versions are robust over different levels of disaggregation. The nonregularized estimators, however, have higher average connectedness when we increase the dimensions. This observation emphasizes the need of regularization in this context and exemplifies the bias that can occur in FEVDs (see Section 1.3).

Moreover, we want to emphasize the decreased usage of the degrees of freedom of the regularized estimators in Table 1.5. That is, the regularized estimators optimally choose to include only between 9.9% and 15.6% of the free parameters to fit the data. We read this as a clear sign that standard methods overfit, failing to provide viable insights. In particular, most of the drop in df usage stems from the coefficient matrix A . Moreover, standard methods lead to a clear rejection of the null hypothesis $H_0 : \Omega = I$, whereas regularized estimates cannot. That is, regularization methods are able to recover the sparsity emerging from the large number of variables, such

¹⁷To facilitate intuition, we slightly deviate from the original authors' terminology and use the terms *in-* and *out-*connectedness for what they call *from-* and *to-*connectedness.

	standard				regularized			
	1972- 1983	1984- 1995	1996- 2007	2008- 2019	1972- 1983	1984- 1995	1996- 2007	2008- 2019
<i>Three-digit (88 sectors)</i>								
$C(D^{gH})$	93.9%	94.3%	93.6%	93.3%	46.4%	28.0%	25.8%	25.4%
$df(A)$	11264	11264	11264	11264	1520	848	840	1185
$df(\Sigma)$	3916	3916	3916	3916	849	668	802	840
LW test (p-value)	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
Partial R^2 (Factors)	93.6%	93.6%	94.9%	95.4%	8.6%	6.0%	6.6%	12.4%
Partial R^2 (Spillovers)	95.6%	95.3%	96.0%	96.3%	35.3%	23.8%	23.3%	28.6%
<i>Two-digit (26 sectors)</i>								
$C(D^{gH})$	61.5%	55.3%	52.3%	54.5%	41.8%	36.4%	21.4%	29.5%
$df(A)$	1716	1716	1716	1716	302	221	210	346
$df(\Sigma)$	351	351	351	351	155	227	124	178
LW test (p-value)	0.00	0.00	0.00	0.00	0.70	0.50	1.00	1.00
Partial R^2 (Factors)	43.8%	40.3%	46.7%	51.5%	8.6%	9.8%	8.8%	16.9%
Partial R^2 (Spillovers)	51.3%	42.3%	44.7%	54.1%	25.0%	15.6%	16.9%	22.2%
<i>Four-digit (117 sectors)</i>								
$C(D^{gH})$	99.1%	99.1%	99.1%	99.4%	41.4%	31.9%	28.9%	27.0%
$df(A)$	18369	18369	18369	18369	2089	1128	1395	1724
$df(\Sigma)$	6903	6903	6903	6903	1059	1160	1342	1178
LW test (p-value)	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
Partial R^2 (Factors)	100%	100%	100%	100%	7.4%	5.0%	6.9%	12.2%
Partial R^2 (Spillovers)	100%	100%	100%	100%	38.0%	25.8%	29.7%	31.2%
<i>Five-digit (138 sectors)</i>								
$C(D^{gH})$	99.3%	99.4%	99.2%	99.2%	43.3%	33.5%	30.9%	27.6%
$df(A)$	24564	24564	24564	24564	2555	1638	1927	2313
$df(\Sigma)$	9591	9591	9591	9591	3883	2587	2962	2623
LW test (p-value)	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
Partial R^2 (Factors)	100%	100%	100%	100%	6.9%	5.6%	7.2%	12.4%
Partial R^2 (Spillovers)	100%	100%	100%	100%	40.3%	29.6%	32.9%	35.8%

Table 1.5: Summary of the estimation results for different levels of sectoral disaggregation. The rows labeled $C(D^{gH})$ show the estimated average connectedness. Rows denoted Partial R^2 report the partial coefficients of determination for the factor loadings and the spillovers weighted by sectoral weight. The row labeled df additionally indicates the nonzero values in the autoregression coefficient matrix A and the covariance matrix Σ together, corresponding to the total degrees of freedom. The row labeled LW test displays the p-value of testing the null hypothesis of the generalized innovation covariance matrix being equal to the identity $H_0 : \Omega = I$ using the test statistic of Ledoit and Wolf (2002, see Appendix 1.A.3).

that the generalized-shock covariance matrix Ω is statistically not distinguishable from the identity. Finally, we report partial R^2 coefficients for the competing explanations. In our regularized specifications, the factors are able to explain 5% to 16.9% of the variation with almost no change between the first and second subsamples, but a visible increase in the most recent episode. Complementary, spillovers capture 25% to 40.3% of the nonfactor variation in the first subsample with a decrease of approximately 10 percentage points post-1984. This result is robust at all levels of disaggregation. For the remainder of our analysis, we focus again on the three-digit sectoral disaggregation data with 88 sectors.

To give a holistic view of the network, we summarize the estimated row-normalized connectedness tables in Figure 1.4. From eye-balling, it is evident that the network significantly changed after 1984. Whereas the pre-Great Moderation subsample shows a closely connected graph with a handful of powerful nodes in the center, the three consecutive subsamples have more widespread graphs with less concentration and lower average connectedness $C(D^{gH})$.¹⁸ For deeper insights, we investigate the sectoral in- and out-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 and decreased contributions afterwards. As a first observation from Figure 1.4, large values for out-connectedness (lighter/yellow nodes) become less frequent after 1984.

Although more spillovers do not necessarily result in higher aggregate volatility, the root may be in the concentration of outgoing spillovers. In that regard, out-connectedness measures how much a single sector's volatility explains the volatility of all other sectors. If a handful of sectors have high levels of out-connectedness, then their volatility determines the volatility of other sectors to a large extent. Consequently, the volatility of the aggregate IP index is also indirectly affected by innovations to those sectors and their shocks do not average out.

Figure 1.5 displays the complementary cumulative distribution functions (CCDFs) of the out-connectedness measure for the four subsamples. Notably, the distribution in the pre-Great Moderation subsample from 1972 to 1983 shows a characteristic bump for approximately half of the sectors with high levels of out-connectedness and expands much further to the right than in subsequent subsamples. The pronounced right tail for this subsample is an indication of a heightened network concentration, supporting the hypothesis of network-induced underdiversification of the aggregate IP index. Roughly 10% of the sectors have an out-connectedness higher than 1, meaning that each explains

¹⁸In Appendix 1.A.4, the corresponding nonregularized estimates (Figure 1.A.12) fail to provide similar insight.

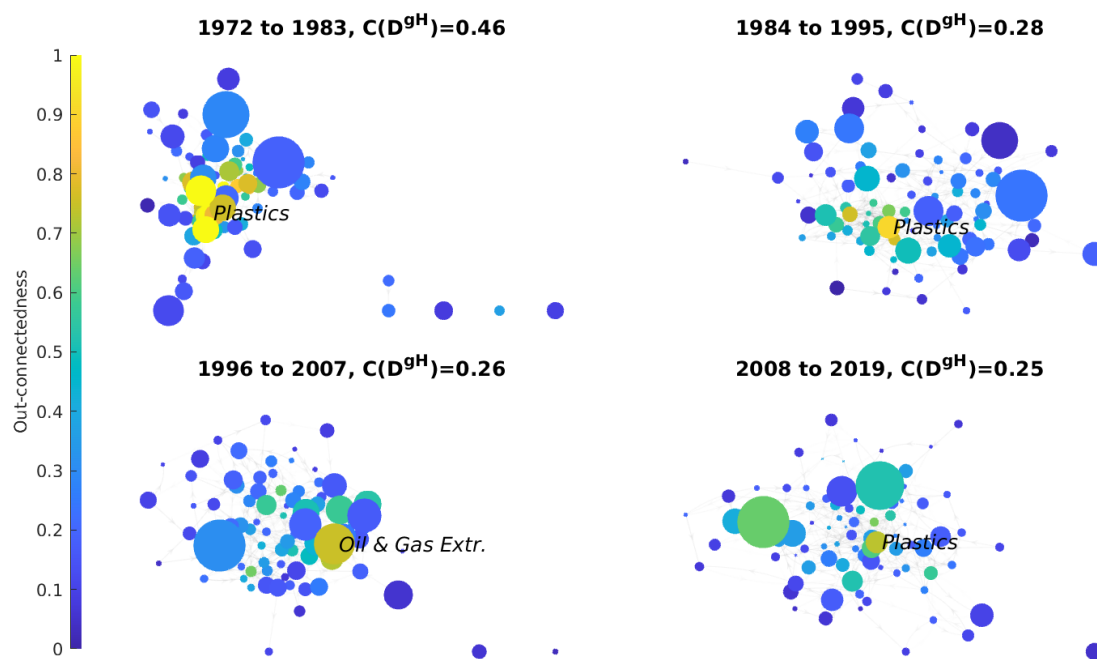


Figure 1.4: Connectedness networks for the respective periods. The force-directed graph drawing algorithm arranges the nodes. That is, two nodes appear closer in the graph if they have stronger connections to each other. Although we initialize the subgraphs on the same scale, the algorithm cannot guarantee that the graphs are comparable in size. The size of the node relates to the respective average weight $\bar{w}_{i,t}$ of the sector in the IP index. The colors depict the out-connectedness. For the sake of the visualization, we cap the color scale at 1. The sectors with the highest out-connectedness are labeled.

more than 100% of the volatility of the other sectors in total.¹⁹ Therefore, their shock volatilities would show up more than twofold in an aggregate index with equal weights and variances. Finally, the fat tail mostly disappears with the Great Moderation, that is, the subsamples after 1983 have more diversified production networks.

Figure 1.6 tracks the dynamics of the out-connectedness for individual sectors. Sectors above the 45° line have decreased their effect on other sectors and vice versa. The shift in distribution is immediately evident as most sectors appear in the top left of the graph. However, this graph clearly shows a strong decrease for a number of sectors (above the diagonal/blue), whereas for the majority of sectors the out-connectedness stays relatively constant (close to the diagonal/yellow). Remarkably, it seems that only a few sectors increased in out-connectedness (below the diagonal/red). Overall, we observe a shift of impactful sectors to less systemically relevant positions in the network. This finding supports the hypothesis of structural change as a contributor to the Great Moderation. Focusing on the sectors with the biggest drops in out-connectedness, we see mainly intermediate goods sectors that use outputs from and provide inputs to

¹⁹Supporting histograms of the in- and out-connectedness measures (Figures 1.A.13 and 1.A.15) are available in Appendix 1.A.4. Here, the 25% strongest sectors by pre-Great Moderation out-connectedness increasingly blend into the distribution in subsequent subsamples.

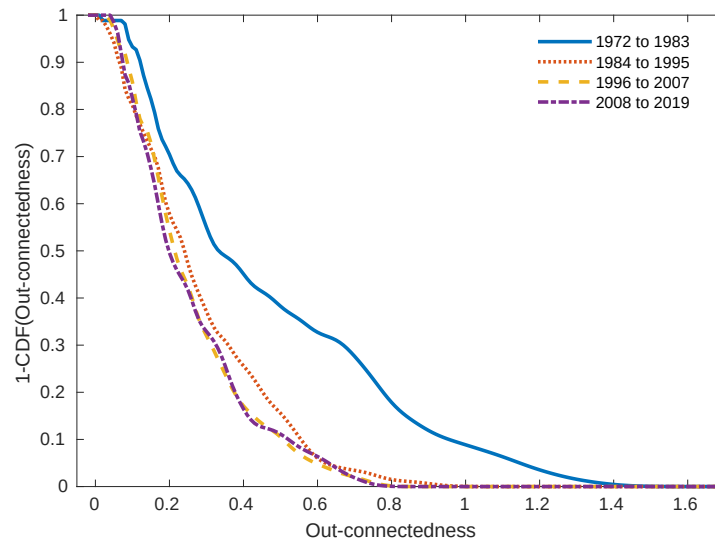


Figure 1.5: Estimated complementary cumulative distribution functions (CCDFs) of the out-connectedness of the 88 three-digit-level sectors. This function is the equivalent to survival functions in the time domain.

other sectors. As sectoral input–output relations have unlikely ceased, this observation gives rise to the idea that sectors were less exposed to disruptions in down or upstream sectors.

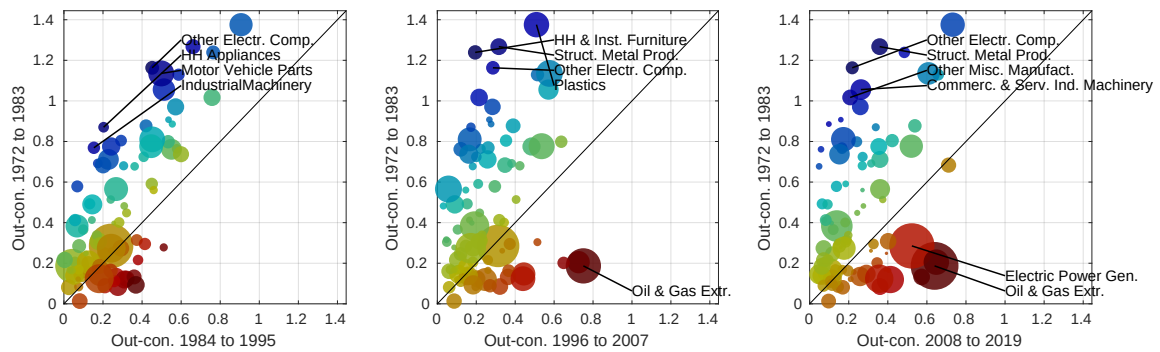


Figure 1.6: Out-connectedness scatter plot of the 88 three-digit-level sectors pre- versus post-Great Moderation. Each dot in the plot corresponds to a sector, with the position along the y- and x-axes according to their out-connectedness before and after the Great Moderation, respectively. The sizes of the markers correspond to the weight of the sectors in the IP index. The colors show how far away the sectors are from the diagonal line, that is, how much their out-connectedness changed relative to the pre-Great Moderation subsample. The sectors with the largest decreases and increases in out-connectedness are labeled.

These findings connect to a strand of literature on inventory-based explanations of the Great Moderation (see Kahn et al. (2002), Summers et al. (2005), Davis and Kahn (2008)). In this literature, the adoption of improved inventory management techniques and information technology has led to a reduction in supply chain disruptions, effectively leading to a decrease in output volatility. Davis and Kahn (2008) argue that, before the Great Moderation, firms carry higher stocks of inventory to avoid foregone profits in times of uncertain demand. With the onset of the Great Moderation, the

average inventories-to-shipments ratio has dropped significantly. To understand how this affected output volatility, consider, for example, an adverse and persistent demand shock. With higher inventories, firms not only are stuck with excess inventories, but also face reduced sales forecasts due to the persistence of the shock. Thus, large inventory holdings amplify production volatility as firms reduce their production by more than the size of the demand shock — a phenomenon the operations research literature knows as the “bullwhip effect” (see Forrester (1961), Lee et al. (1997)). The amplification of demand shocks may contribute to excessively low demand in input-supplying sectors, and thus, intensify input–output dependencies.

Empirically, Chen et al. (2005) document a strong decline in inventories of US manufacturing companies from 1981, and Bray and Mendelson (2012) provide evidence that the amplifying effect of supply chains has faded between 1974 and 2008. To validate the connection of our findings to sectoral inventories, we run a regression attempting to explain the changes in out-connectedness with percentage changes in inventories-to-shipments ratios. A stronger adaptation to new inventory management standards should coincide with a stronger change in out-connectedness. In Appendix 1.A.4, Figure 1.A.16 shows the relation between the change in inventories-to-shipments ratio to the change in out-connectedness. We observe a clear positive relationship with a p-value around 1%. Paired with the preceding analysis, this observation suggests that previously influential sectors profited more from improved inventory management.

In addition to sectoral out-connectedness, the distribution of in-connectedness may have played a critical role in the structural change that emerged during the Great Moderation. However, we find its role in amplifying volatility to be ambiguous because of two channels. First, big values of in-connectedness make it more likely that a small sector’s volatility spills over to sectors with bigger weights. That is, if we see a handful of sectors with large weights having very strong in-connectedness, the spillovers might skew the weighting scheme of the aggregate index toward smaller and more volatile sectors. Second, higher values of in-connectedness signify a higher degree of diversification, which in return would wash out the intrinsic volatility contribution. Paired with an unfortunate distribution of weights — for example, sectors with big weights might have received more spillovers after 1984 — this could lead to more diversification after the Great Moderation. Both channels are in line with the argument of Gabaix (2011) that a small number of constituents with large index weights can explain aggregate shocks. We analyze both conjectures in Figures 1.A.13 and 1.A.14 in Appendix 1.A.4, respectively, but do not find support for any of the aforementioned.

Next, we want to understand how the full network structure affected aggregate index volatility and how individual sectors contributed to the aggregate. From the

aggregate index representation $g_t = \sum_{i=1}^N w_{i,t} g_{i,t}$, one estimate of the IP index variance is $\hat{\sigma}_{IP}^2 = \bar{w}' \hat{\Sigma}_g \bar{w}$, where $\bar{w}$ is a vector of average weights in the respective subsample, and $\hat{\Sigma}_g$ is a covariance estimate of the index constituents. Similar to the generalized impulse response function, we investigate the response of $g_{i,t}$ to innovations in other sectors $u_{j,t-k}$ for $k \geq 0$. This way we can trace back variations of the weighted index to its originating sector. We use the $H = 12$ period sum of all sectors' covariances to the innovation vector, $IRV(H) = \sum_{h=0}^{H-1} \Phi_h \Sigma = \sum_{h=0}^{H-1} cov(y_t, u_{t-h})$. We denote this term as the innovation response variance (IRV), and define the vector $v_{\bullet \rightarrow IP}^{gran}$ as the sectoral variance contributions to the model-implied aggregate IP index:

$$v_{\bullet \rightarrow IP}^{gran} = \bar{w}^{\cdot 2'} IRV(H), \quad \text{and } v^{gran} := v_{\bullet \rightarrow IP}^H \mathbf{1}, \quad (1.17)$$

where $\bar{w}^{\cdot 2}$ is a vector of squared average weights in the respective period, that is, its direct weight in the aggregate index variance. This representation is particularly appealing, as, for example, entry i, j in the IRV matrix shows the response of sector i to innovations happening in sector j . By multiplying the IRV with the squared weights, we see the effect of a sector on the aggregate index. In our framework, the IRV is complemented by the factor response variance $FRV(H) = \sum_{h=0}^{H-1} \Phi_h B \Sigma_f B' = \sum_{h=0}^{H-1} cov(y_t, B f_{t-h})$. Analogously, the vector of sectoral contributions to the index variance induced by reactions to factor innovations is $v^f = \bar{w}^{\cdot 2'} FRV(H) \mathbf{1}$.

For additional insights into the channels through which granular innovations and factor innovations contributed to the index, we further decompose the IRV representation into intrinsic and extrinsic components, where the latter comprises the entirety of spillover contributions. Similarly, we separate the contemporaneous and lagged channels for the IRV, and the direct and indirect components for the FRV as potential drivers of the Great Moderation in our analysis.

Table 1.6 documents the contributions of multiple components to variations of the aggregate index. First, we confirm that the aggregate variations implied by our model, v^{agg} , follows the evolution of the data variance, v^{IP} . While the estimated factor contributions v^f show little impact on the overall responses to innovations, the granular contributions v^{gran} explain the major part of the aggregate variance in our model. That is, our model attributes most variance contributions to granular innovations. A further decomposition into an intrinsic component, v^{IC} , and an extrinsic component, v^{EC} , reveals the weakening of the spillover network as the main driver of the decline in granular contributions. Similarly, strong changes happen in the lagged transmission channel as opposed to contemporaneous reactions, refuting the narrative of aggregate shocks as the main explanation for the Great Moderation. On the other hand, returning to the narratives of Kahn et al. (2002), Summers et al. (2005) and Davis and Kahn (2008), the decline in extrinsic and lagged contributions

		1972- 1983	1984- 1995	1996- 2007	2008- 2019
Data variance	$v^{IP} = \text{var}(g_t)$	148.1	34.7	36.8	87.4
Aggregate response	$v^{agg} = v^f + v^{gran}$	149.1	23.5	25.9	85.0
Factor Contributions	$v^f = \bar{w} \cdot 2' FRV \mathbf{1}$	11.7	1.1	-0.1	17.8
of which direct	$v^{fDir} = \bar{w} \cdot 2' B \Sigma_f B' \mathbf{1}$	2.7	1.0	-0.1	10.9
of which indirect	$v^{fInd} = v^f - v^{fDir}$	9.0	0.1	0.0	7.0
Granular Contributions	$v^{gran} = \bar{w} \cdot 2' IRV \mathbf{1}$	137.4	22.4	26.0	67.1
of which intrinsic	$v^{IC} = \bar{w} \cdot 2' \text{diag}(IRV) \mathbf{1}$	16.7	9.4	12.1	17.1
of which extrinsic	$v^{EC} = v^{gran} - v^{IC}$	120.7	13.0	13.8	50.0
of which contempor.	$v^{cont.} = \bar{w} \cdot 2' \Sigma \mathbf{1}$	54.5	19.7	21.0	33.7
of which lagged	$v^{lag} = v^{gran} - v^{cont.}$	82.9	2.6	5.0	33.4
Extrinsic Share	v^{EC} / v^{agg}	0.81	0.55	0.53	0.59
Lagged Share	v^{lag} / v^{agg}	0.56	0.11	0.19	0.39

Table 1.6: Decomposition of variance contributions by channels for $H = 12$. $\mathbf{1}$ denotes the one vector of respective size. Aggregate responses provide an approximation of the variance of the aggregate IP index via innovation response variances. Factor innovations and granular innovations have direct/intrinsic effects through their own variance and indirect/extrinsic effects through intersectoral linkages on index constituents. The contemporaneous spillover contribution comprises all variations within the same period, and the lagged component consists of all responses happening via the autoregressive coefficients.

is consistent with the idea that advances in transportation, supply and distribution networks, inventory management, and sales forecasting related to new information technologies have streamlined intersectoral dependencies, and thus, have reduced the volatility of individual sectors. Overall, there is a considerable share of the decline in aggregate volatility that our model attributes to the weakening of the spillover network, remarkably after adjusting for common factor-based explanations. Hence, our findings suggest that structural change of the network contributed to the Great Moderation.

Lastly, we care about single-sector contributions via the IRV in (1.17). To do so, Figure 1.7 tracks the sectors' contributions to aggregate variations over the four subsamples. We place the sectors with the highest contributions at the bottom and vice versa. In brackets, we show the percentage of the sectors' contributions via the extrinsic channel as in Table 1.6:

$$v_{\bullet \rightarrow IP}^{gran} = \underbrace{\bar{w} \cdot 2' \text{diag}(IRV(H))}_{\text{intrinsic component}} + \underbrace{\bar{w} \cdot 2' [IRV(H) - \text{diag}(IRV(H))]}_{\text{extrinsic component}}. \quad (1.18)$$

For example, a value of 73% signifies that 73% of the estimated contributions come via estimated spillovers; or in other words, the network noticeably amplifies the sector's

structural change in the form of improvements to inventory management, supply chains, and information technology offers a plausible narrative for the less volatile industrial output. Despite the partial increase in spillover contributions in the 2008 to 2019 subsample (compare Figure 1.A.18 in Appendix 1.A.4), there is no strong indication for a reversal of the Great Moderation’s structural change. High intrinsic and common-factor contributions surrounding the Great Recession tell a more convincing story than a return of the strong spillovers.

Throughout the application, two principal findings stand out and repeat in various analyses. First, we document a sharp decline in network connectedness after 1984 in Table 1.5, which also shows in many visualizations such as Figures 1.4, 1.6 and 1.7. As illustrated in Figure 1.5, our evidence suggests that outgoing links have weakened considerably. A decomposition of aggregate variance in Table 1.6 shows that the importance of spillover effects has particularly faded with the Great Moderation. Second, we find that sectors demanding inputs from other sectors are among those that reduced their outgoing links the most, which is visible in Figures 1.6 and 1.7. We see this result as indication that supply chain improvements have played an important role with the onset of the Great Moderation. More precisely, improvements to inventory management offer a potential explanation as they deter demand shocks from propagating upstream (see Figure 1.A.16). Finally, we find no substantial evidence that a revival of sectoral spillovers drove the temporary increase in production volatility around the Great Recession.

Our results provide novel insights into the changes observable during the Great Moderation. Foerster, Sarte, et al. (2011) emphasize that the decrease in aggregate variance is not due to some sectors but is rooted instead in the change of aggregate shocks; that is, shocks to multiple sectors at once. In contrast, our analysis reveals that changes in the propagation of sector-specific shocks may have contributed largely to the decline in aggregate variance. Put differently, a sector contributes to the aggregate variance not only via its index weight, but also via the spillover effects it has on other constituents. Such a reduction in spillovers supports the granular hypothesis of Gabaix (2011) and its network refinement in Acemoglu et al. (2012). Further, we conjecture that improved inventory management structurally changed the spillover network such that its amplifying effect subsided. Thus, our findings offer a unifying perspective to the opposing explanations of Foerster, Sarte, et al. (2011) and Gabaix (2011) by connecting their narratives to the inventory-based explanations of Kahn et al. (2002), Summers et al. (2005) and Davis and Kahn (2008).

1.5 Conclusions

In this paper, we investigate the estimation of high-dimensional connectedness tables via vector autoregressive models. In a simulation study, we compare different regularization methods for the coefficient and the covariance matrix. We evaluate their performance in the estimation of FEVDs and find that the regularization of both matrices improves bias and variance. Since there is no one-fits-all estimator in our simulations, we suggest validating the estimators through cross-validation in practice.

In an application to US industrial production, we are able to uncover changes in the structure of the intersectoral spillover network around the Great Moderation. Our results support the notion that receding network effects have largely aided the decline in aggregate variance by decreasing the influence of granular innovations. Specifically, the outgoing spillovers of downstream sectors before the Great Moderation was unmatched thereafter, suggesting that supply chains and inventory management may be important drivers of the structural change behind the Great Moderation.

Finally, we would like to attempt to extrapolate our findings to the current macroeconomic environment. With the COVID-19 pandemic, disruptions in the production process due to supply chain bottlenecks and rapid changes in the consumer product demand have rendered lean inventory management less effective, such that many firms adapted safeguard inventories. Although it is too early to draw conclusions due to the lack of sufficient data, we conjecture that persistent supply chain issues and high inventory levels may again result in a strengthening of the sectoral spillover network, heralding a period of heightened macroeconomic volatility. Therefore, we encourage future research to pick up on the narrative of supply shortages and changes in inventories to explain the volatile macroeconomic environment since the onset of the pandemic.

1.A Appendices

1.A.1 Conditions for the thresholding rule

- (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-thresholding $s_\lambda(z) = \text{sgn}(z)(|z| - \lambda)_+$ and the adaptive LASSO rule $s_\lambda(z) = z(1 - |\lambda/z|^\eta)_+$ with $\eta \geq 1$, are included.

1.A.2 Illustration of cross-validation procedure

	Partition data into 12 equally sized folds											
Fold 1	Test											
Fold 2		Test										
Fold 3			Test									
Fold 4				Test								
Fold 5					Test							
Fold 6						Test						
Fold 7							Test					
Fold 8								Test				
Fold 9									Test			
Fold 10										Test		
Fold 11											Test	
Fold 12												Test

Figure 1.A.8: Illustration of 12-fold CV. First, we split the sample into 12 folds of equal length. Then we use 11 folds for training and validate the trained model's performance on the hold-out test fold.

1.A.3 Test statistic for equality of covariance matrices

We want to test the hypothesis if the generalized shock covariance matrix Ω is different from the identity matrix, i.e.,

$$H_0 : \Omega = I_N.$$

Let N be the number of time series in our panel and T be the number of time periods in the respective sample. Ledoit and Wolf (2002) develop a χ^2 test statistic to test H_0 specifically when N is large and T is small:

$$W = \frac{NT}{2} \left\{ \frac{1}{N} \text{tr}((\Omega - I_N)^2) - \frac{N}{T} \left[\frac{1}{N} \text{tr}(\Omega) \right]^2 + \frac{N}{T} \right\}.$$

Under the null hypothesis H_0 , W is distributed as $\chi^2 \left[\frac{N(N+1)}{2} \right]$. Note that the degrees of freedom parameter for the χ^2 -distribution corresponds to the number of distinct parameters in Ω . Finally, we reject H_0 at significance level α if W is greater than $\chi^2 \left[\alpha, \frac{N(N+1)}{2} \right]$. A corresponding p -value follows as $p = 1 - \text{CDF}_{\chi^2_{N(N+1)/2}}(W)$.

1.A.4 Complementary graphs

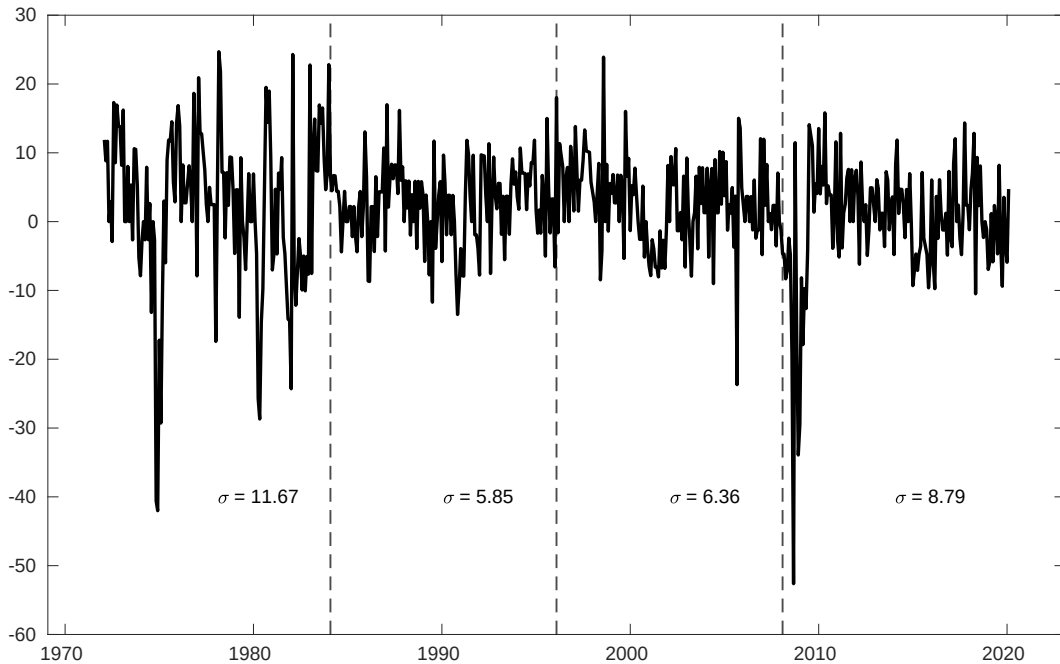


Figure 1.A.9: Annualized growth rates of monthly aggregate industrial production (IP) in percentage points. σ denotes the sample volatility in the respective subsample.

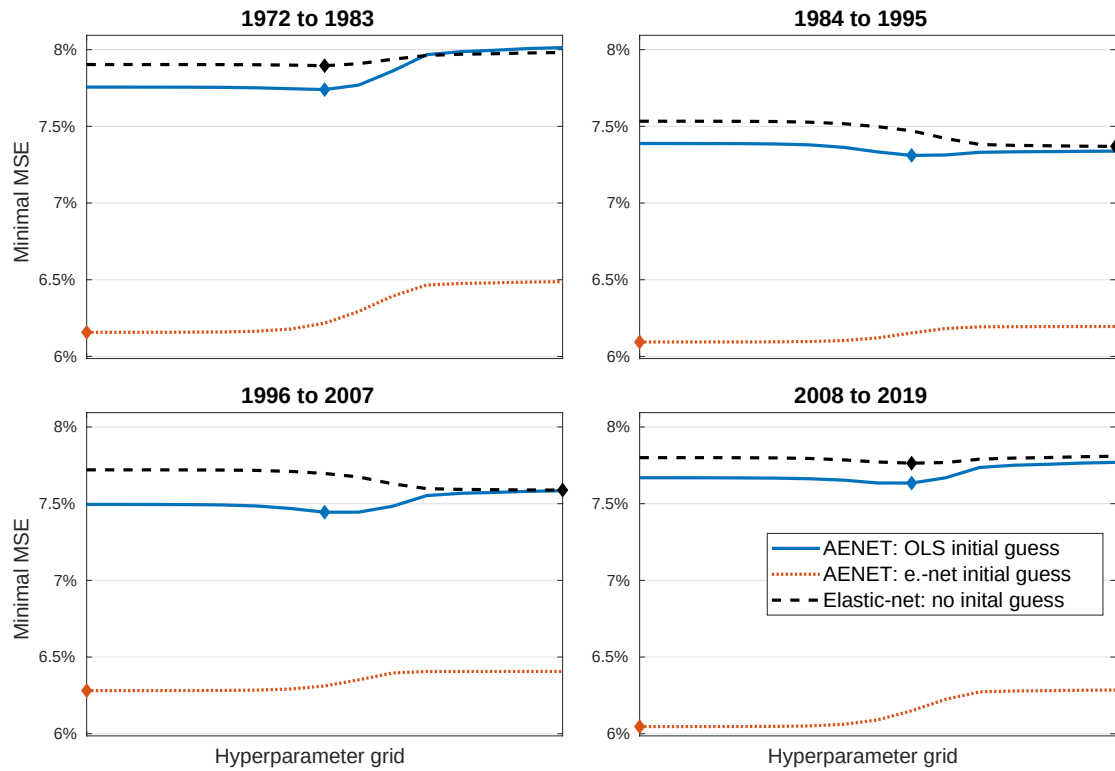


Figure 1.A.10: 12-fold CV results for the tuning parameter α in the adaptive elastic-net. The scale shows the percentage improvement over the OLS case, e.g., 33% refers to a forecast error being only a third that of the OLS estimate. The blue solid curves show the minimal MSE with $\hat{A}_{OLS}$ as the initial estimate in the weights of the individual penalties for different α s, following the approach of Demirer et al. (2017). The red dotted curves show the same for $\hat{A}_{ENET}$ as an initial guess (selected by 12-fold CV, $\alpha = 0.5$, and $w_i = 1$), the black dashed lines correspond to the original (non-adaptive) elastic-net as in Zou and Hastie (2005).

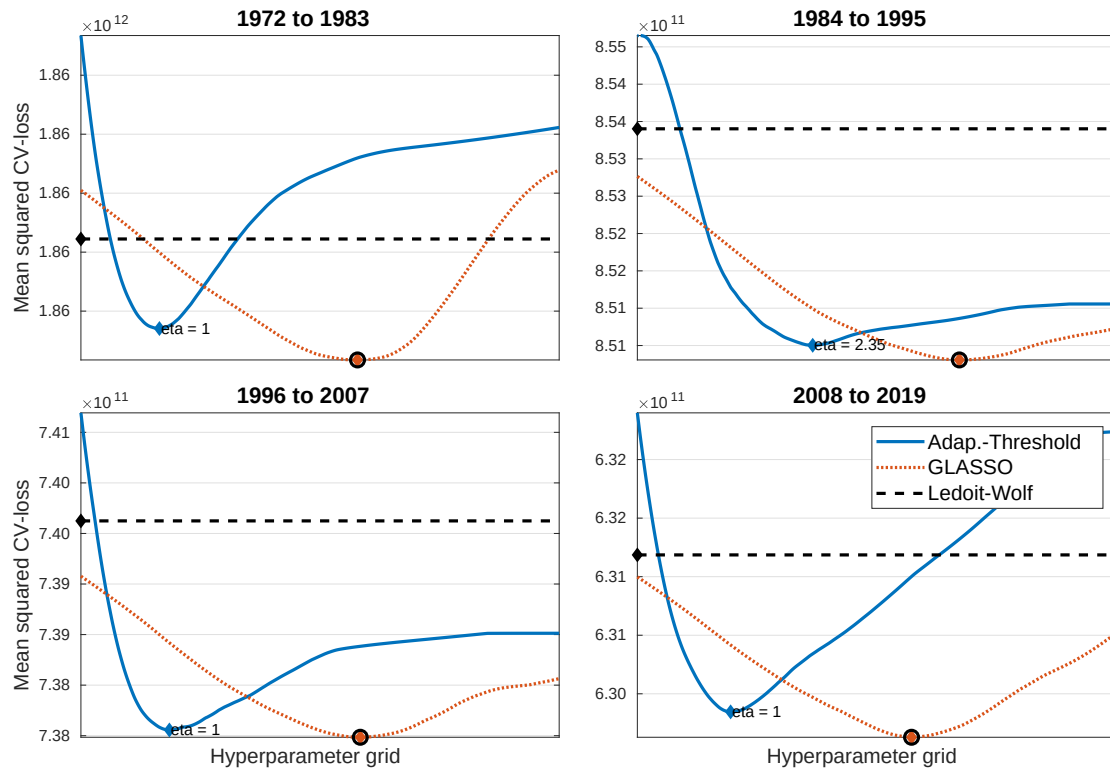


Figure 1.A.11: The 12-fold CV results for the tuning parameter of different covariance regularization methods. Values of the tuning parameters are on the x-axis and the MSEs are shown on the y-axis. The two thresholding estimators tune δ^{AT} in (1.13), GLASSO tunes δ^{GL} in (1.16), and the manual Ledoit-Wolf tunes the otherwise automatic shrinkage parameter δ^{LW} . The x-axis is normalized for each regularization method such that the convexity of all of the MSE curves is visible.

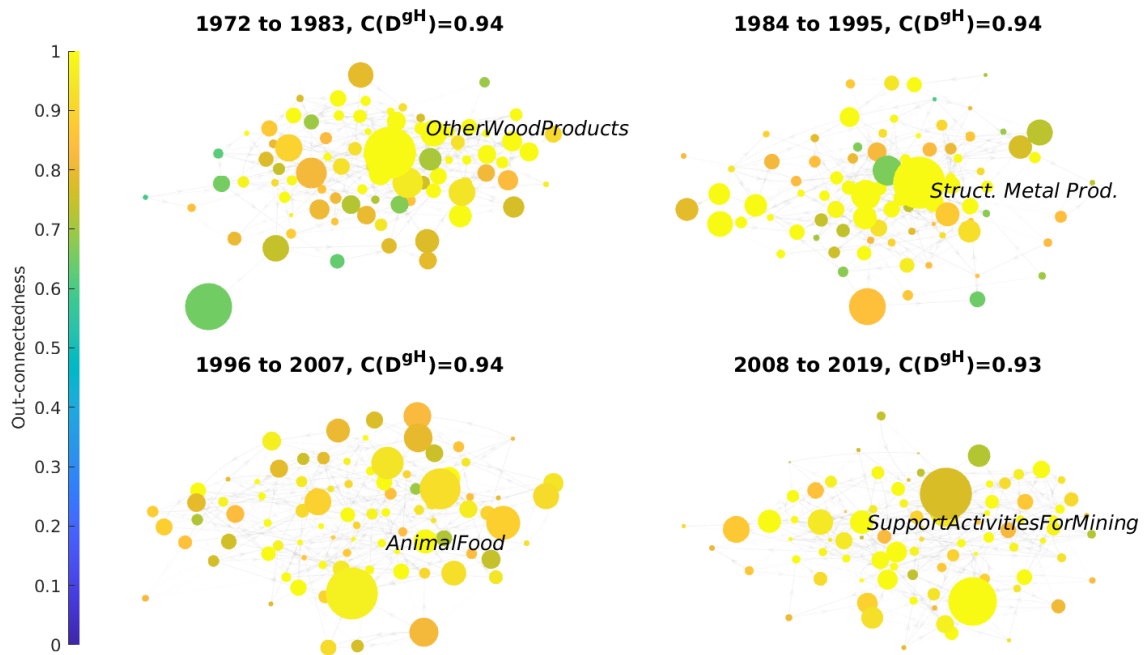


Figure 1.A.12: Connectedness networks for the respective subsamples using the non-regularized estimates. The force-directed graph drawing algorithm arranges the nodes. That is, two nodes appear closer in the graph if they have stronger connections to each other. Although we initialize the subgraphs on the same scale, the algorithm cannot guarantee that the graphs are comparable in size. The size of the node relates to the respective average weight $\bar{w}_{i,t^*}$ of the sector in the IP index. The colors depict the out-connectedness. For the sake of the visualization, we cap the color scale at 1. The sectors with the highest out-connectedness are labeled.

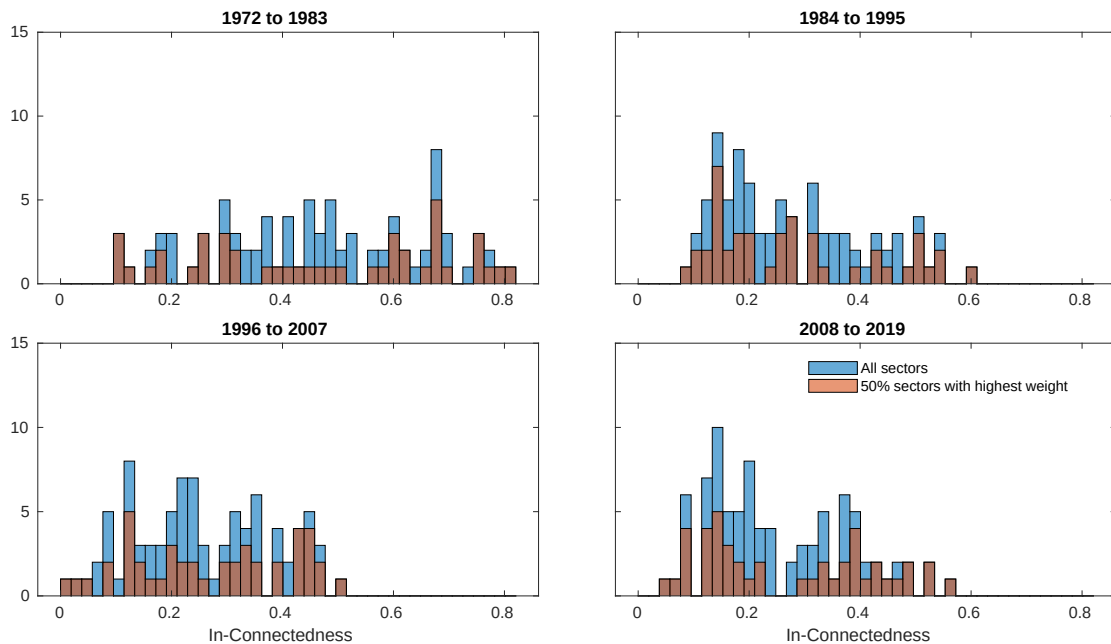


Figure 1.A.13: Histogram of the in-connectedness of the 88 three-digit level sectors. The 50% of the sectors with the biggest weights are highlighted in red. The in-connectedness relates to incoming spillovers from other sectors.

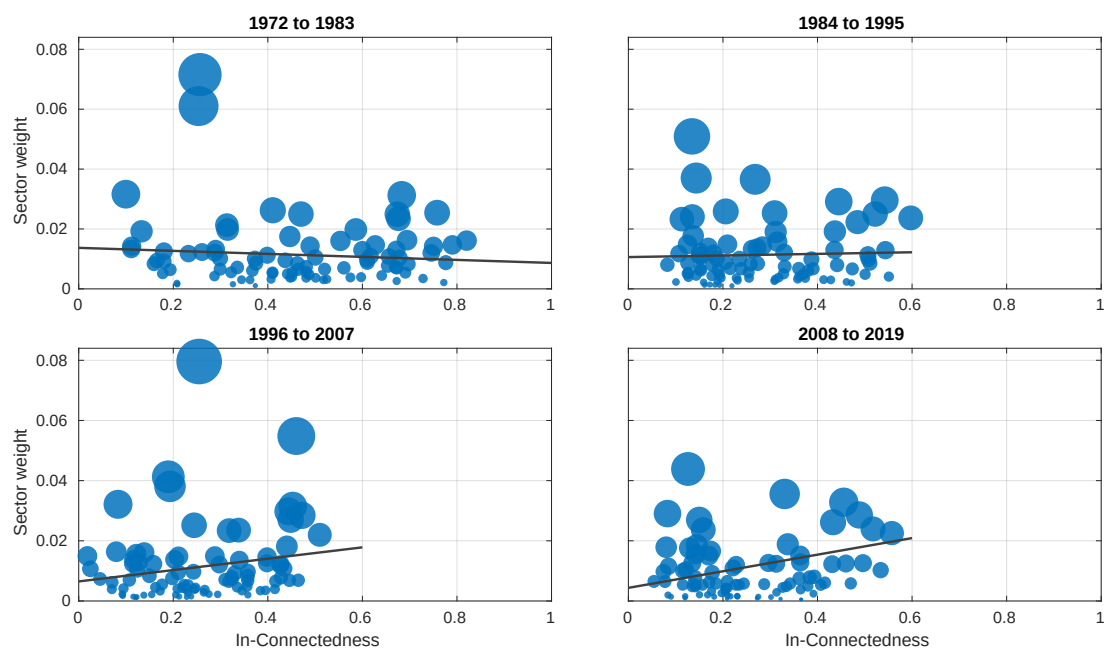


Figure 1.A.14: Scatterplots of the in-connectedness of the 88 three-digit-level sectors against the sectoral weights in the aggregate IP index. The in-connectedness relates to incoming spillovers from other sectors. We plot the regression line per sample.

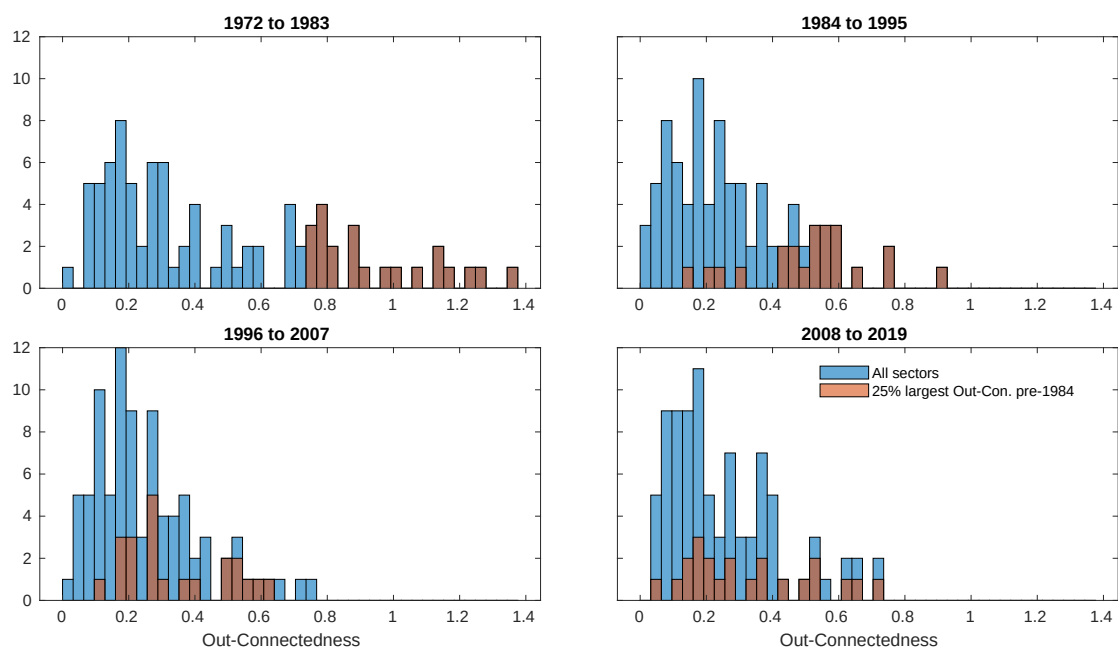


Figure 1.A.15: Histogram of the out-connectedness of the 88 three-digit-level sectors. The 25% of the sectors with the biggest weights in the first subsample are highlighted in red. The out-connectedness relates to outgoing spillovers to other sectors.

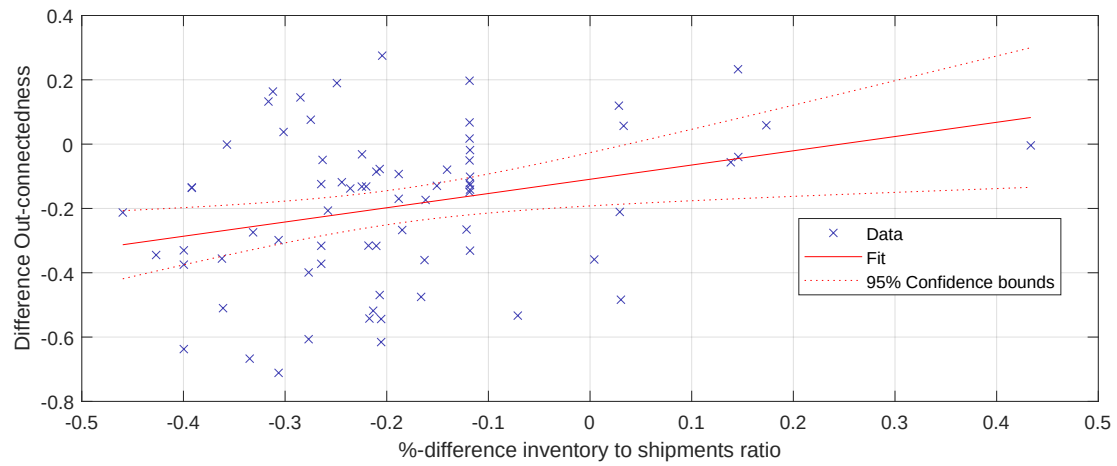


Figure 1.A.16: Regression plot of sectoral changes in out-connectedness on percentage changes in inventories-to-shipments ratio. Changes in out-connectedness for sector i $\Delta OutCon_i$ are differences in the estimates from the pre-Great Moderation subsample (1972 to 1983) to the next subsample (1984 to 1995). Inventories-to-shipments ratios are obtained from the Manufacturers' Shipments, Inventories, and Orders (M3) survey and percentage changes $\Delta I2S_i$ follow from pre- and post-1984 means. 73 SIC categories are matched to NAICS categories via the official [dispersement table](#) of the US Census Bureau. Note that some sectors' series are unpublished and the matching is imperfect due to new sector definitions. The fitted regression equation reads $\Delta OutCon_i = -0.1093 + 0.4429 \Delta I2S_i$. t -statistics (p-values) for the intercept and coefficient are $-2.6295(0.0105)$ and $2.6088(0.0111)$, respectively.

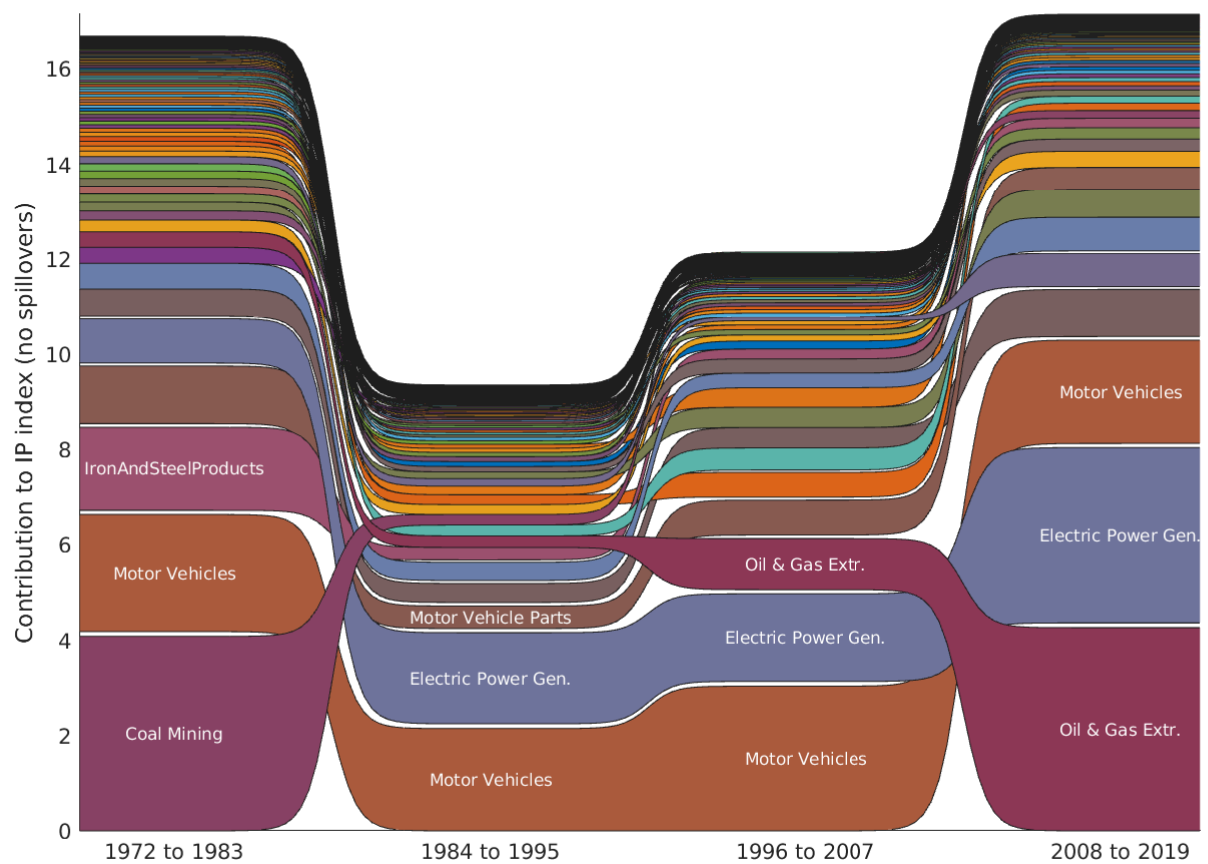


Figure 1.A.17: Bump chart of sectoral contributions to the aggregate variance of the IP index without spillovers. Contributions are calculated as in (1.17). Data includes IP indices for 88 three-digit-level sectors. The top three contributors per subsample are labeled.

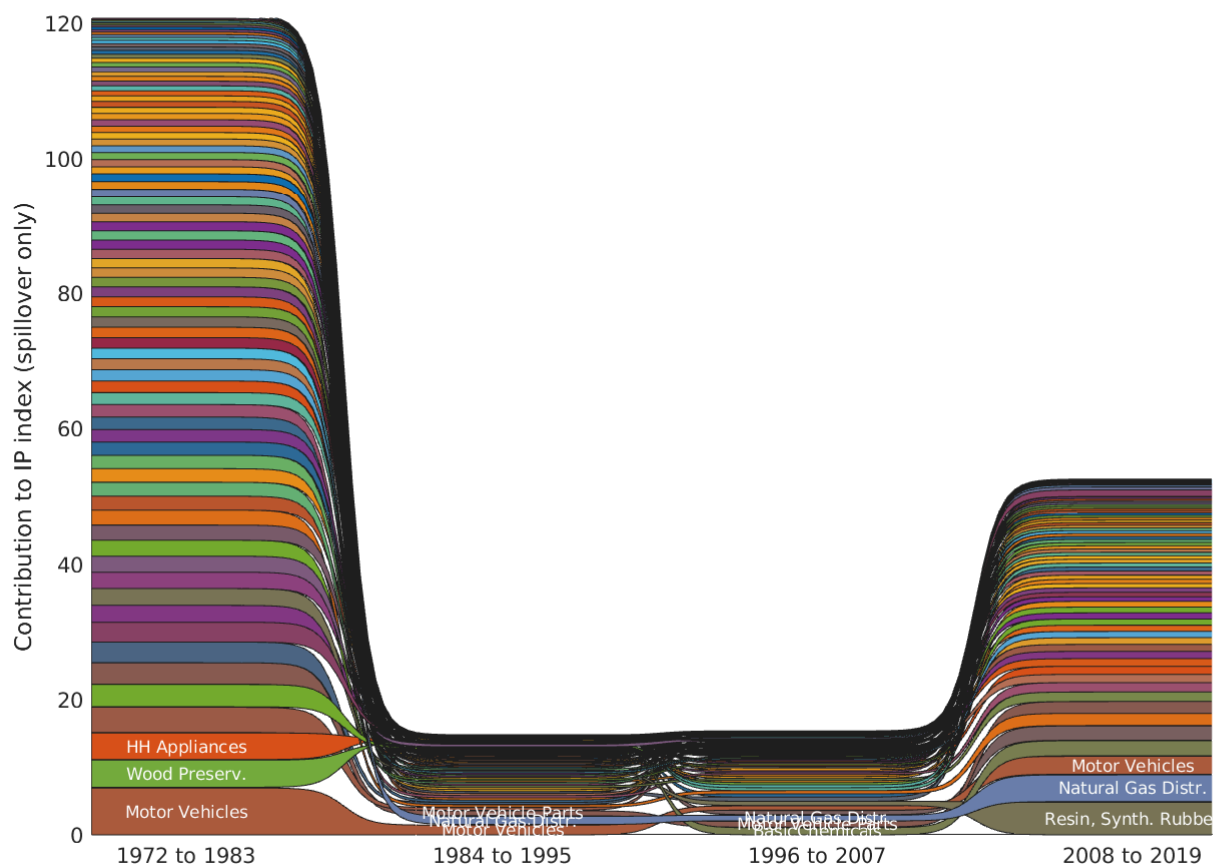


Figure 1.A.18: Bump chart of sectoral contributions to the aggregate variance of the IP index with spillovers only. Contributions are calculated as in (1.17). Data includes IP indices for 88 three-digit-level sectors. The top three contributors per subsample are labeled.

1.B Bibliography

- Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). “The network origins of aggregate fluctuations.” In: *Econometrica* 80.5, pp. 1977–2016.
- Arias, A., G. D. Hansen, and L. E. Ohanian (2007). “Why have business cycle fluctuations become less volatile?” In: *Economic theory* 32.1, pp. 43–58.
- Barigozzi, M. and C. Brownlees (2013). “Nets: Network estimation for time series.” In: *Journal of Applied Econometrics*.
- Bauer, M. D. and E. T. Swanson (2022). *A reassessment of monetary policy surprises and high-frequency identification*. Tech. rep. National Bureau of Economic Research.
- Bedrick, E. J. and C.-L. Tsai (1994). “Model selection for multivariate regression in small samples.” In: *Biometrics*, pp. 226–231.
- Bergmeir, C., R. J. Hyndman, and B. Koo (2018). “A note on the validity of cross-validation for evaluating autoregressive time series prediction.” In: *Computational Statistics & Data Analysis* 120, pp. 70–83.
- Bernanke, B. (2004). “The great moderation.” In: *Washington, DC*.
- Bickel, P. J. and E. Levina (2008). “Regularized estimation of large covariance matrices.” In: *The Annals of Statistics*, pp. 199–227.
- Bray, R. L. and H. Mendelson (2012). “Information transmission and the bullwhip effect: An empirical investigation.” In: *Management Science* 58.5, pp. 860–875.
- Brunner, F. and R. Hipp (2023). “Estimating large-dimensional connectedness tables: The Great Moderation through the lens of sectoral spillovers.” In: *Quantitative Economics* 14.3, pp. 1021–1058.
- Cai, T. and W. Liu (2011). “Adaptive thresholding for sparse covariance matrix estimation.” In: *Journal of the American Statistical Association* 106.494, pp. 672–684.

- Carvalho, V. and X. Gabaix (2013). “The great diversification and its undoing.” In: *American Economic Review* 103.5, pp. 1697–1727.
- Carvalho, V. M. (2014). “From micro to macro via production networks.” In: *Journal of Economic Perspectives* 28.4, pp. 23–48.
- Carvalho, V. M. and A. Tahbaz-Salehi (2019). “Production networks: A primer.” In: *Annual Review of Economics* 11, pp. 635–663.
- Chen, H., M. Z. Frank, and O. Q. Wu (2005). “What actually happened to the inventories of American companies between 1981 and 2000?” In: *Management science* 51.7, pp. 1015–1031.
- Davis, S. J. and J. A. Kahn (2008). “Interpreting the great moderation: Changes in the volatility of economic activity at the macro and micro levels.” In: *Journal of Economic perspectives* 22.4, pp. 155–80.
- Demirer, M., F. X. Diebold, L. Liu, and K. Yilmaz (2017). “Estimating global bank network connectedness.” In: *Journal of Applied Econometrics*.
- Diebold, F. X. and K. Yilmaz (2014). “On the network topology of variance decompositions: Measuring the connectedness of financial firms.” In: *Journal of Econometrics* 182.1, pp. 119–134.
- Fernald, J. (2014). *A quarterly, utilization-adjusted series on total factor productivity*. Federal Reserve Bank of San Francisco Working Paper 2012-19. Federal Reserve Bank of San Francisco.
- Foerster, A., A. Hornstein, P.-D. Sarte, and M. Watson (2022). “Aggregate Implications of Changing Sectoral Trends.” In: *Journal of Political Economy* forthcoming.
- Foerster, A. T., P.-D. G. Sarte, and M. W. Watson (2011). “Sectoral versus aggregate shocks: A structural factor analysis of industrial production.” In: *Journal of Political Economy* 119.1, pp. 1–38.
- Forrester, J. W. (1961). *Industrial Dynamics*. MIT Press.
- Friedman, J., T. Hastie, and R. Tibshirani (2010). “Regularization paths for generalized linear models via coordinate descent.” In: *Journal of statistical software* 33.1, p. 1.
- Friedman, J., T. Hastie, and R. Tibshirani (2008). “Sparse inverse covariance estimation with the graphical lasso.” In: *Biostatistics* 9.3, pp. 432–441.

- Gabaix, X. (2011). “The granular origins of aggregate fluctuations.” In: *Econometrica* 79.3, pp. 733–772.
- Galí, J. and L. Gambetti (2009). “On the sources of the great moderation.” In: *American Economic Journal: Macroeconomics* 1.1, pp. 26–57.
- Hastie, T., R. Tibshirani, J. H. Friedman, and J. H. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer.
- Hipp, R. (2020). *On Causal Networks of Financial Firms: Structural Identification via Non-parametric Heteroskedasticity*. Staff Working Papers 20-42. Bank of Canada.
- Hoerl, A. E. and R. W. Kennard (1970). “Ridge regression: Biased estimation for nonorthogonal problems.” In: *Technometrics* 12.1, pp. 55–67.
- Hurvich, C. M. and C.-L. Tsai (1989). “Regression and time series model selection in small samples.” In: *Biometrika* 76.2, pp. 297–307.
- Justiniano, A. and G. E. Primiceri (2008). “The time-varying volatility of macroeconomic fluctuations.” In: *American Economic Review* 98.3, pp. 604–41.
- Kahn, J. A., M. M. McConnell, and G. Perez-Quiros (2002). “On the Causes of the Increased Stability of the US Economy.” In: *Economic Policy Review* 8.1.
- Kappen, H. J. and V. Gómez (2014). “The variational garrote.” In: *Machine Learning* 96.3, pp. 269–294.
- Kascha, C. and C. Trenkler (2015). *Forecasting VARs, model selection, and shrinkage*. Tech. rep. Working Paper Series, Department of Economics, University of Mannheim.
- Koop, G., M. H. Pesaran, and S. M. Potter (1996). “Impulse response analysis in nonlinear multivariate models.” In: *Journal of Econometrics* 74.1, pp. 119–147.
- Lam, C. et al. (2016). “Nonparametric eigenvalue-regularized precision or covariance matrix estimator.” In: *The Annals of Statistics* 44.3, pp. 928–953.
- Ledoit, O. and M. Wolf (2002). “Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size.” In: *The Annals of Statistics* 30.4, pp. 1081–1102.
- Ledoit, O. and M. Wolf (2004). “A well-conditioned estimator for large-dimensional covariance matrices.” In: *Journal of Multivariate Analysis* 88.2, pp. 365–411.

- Leduc, S. and K. Sill (2007). “Monetary policy, oil shocks, and TFP: accounting for the decline in US volatility.” In: *Review of Economic Dynamics* 10.4, pp. 595–614.
- Lee, H. L., V. Padmanabhan, and S. Whang (1997). “Information distortion in a supply chain: The bullwhip effect.” In: *Management science* 43.4, pp. 546–558.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.
- Nakov, A. and A. Pescatori (2010). “Oil and the great moderation.” In: *The Economic Journal* 120.543, pp. 131–156.
- Pesaran, H. H. and Y. Shin (1998). “Generalized impulse response analysis in linear multivariate models.” In: *Economics letters* 58.1, pp. 17–29.
- Qian, J. et al. (2013). *Glmnet for Matlab*. http://www.stanford.edu/~hastie/glmnet_matlab/.
- Raveh, A. (1985). “On the use of the inverse of the correlation matrix in multivariate data analysis.” In: *The American Statistician* 39.1, pp. 39–42.
- Rigobon, R. and B. Sack (2003). “Measuring the reaction of monetary policy to the stock market.” In: *The quarterly journal of Economics* 118.2, pp. 639–669.
- Romer, C. D. and D. H. Romer (2004). “A new measure of monetary shocks: Derivation and implications.” In: *American Economic Review* 94.4, pp. 1055–1084.
- Rothman, A. J., E. Levina, and J. Zhu (2009). “Generalized thresholding of large covariance matrices.” In: *Journal of the American Statistical Association* 104.485, pp. 177–186.
- Sims, C. A. (1980). “Macroeconomics and reality.” In: *Econometrica: journal of the Econometric Society*, pp. 1–48.
- Sims, C. A. and T. Zha (2006). “Were there regime switches in US monetary policy?” In: *American Economic Review* 96.1, pp. 54–81.
- Stock, J. H. and M. W. Watson (2003). “Has the business cycle changed? Evidence and explanations.” In: *Monetary policy and uncertainty: adapting to a changing economy*, pp. 9–56.
- Stone, M. (1977). “An asymptotic equivalence of choice of model by cross-validation and Akaike’s criterion.” In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1, pp. 44–47.

- Summers, P. M. et al. (2005). “What caused the Great Moderation? Some cross-country evidence.” In: *Economic Review-Federal Reserve Bank of Kansas City* 90.3, p. 5.
- Tibshirani, R. (1996). “Regression shrinkage and selection via the lasso.” In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288.
- Wieland, J. F. and M.-J. Yang (2020). “Financial dampening.” In: *Journal of Money, Credit and Banking* 52.1, pp. 79–113.
- Zou, H. (2006). “The adaptive lasso and its oracle properties.” In: *Journal of the American Statistical Association* 101.476, pp. 1418–1429.
- Zou, H. and T. Hastie (2005). “Regularization and variable selection via the elastic net.” In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2, pp. 301–320.
- Zou, H. and H. H. Zhang (2009). “On the adaptive elastic-net with a diverging number of parameters.” In: *The Annals of Statistics* 37.4, p. 1733.

Chapter 2

Echo-nomics: Does the stock market resonate with granular innovations?

with Ruben Hipp

Abstract

This paper investigates the interplay between asset-level innovations and aggregate market variation through the lens of a directed spillover network. Utilizing high-dimensional samples with daily data of 100 US stocks, we probe the “granular hypothesis” in a network setting and explore two key aspects. First, we establish the presence and dynamics of a directed spillover network that leads to informational contagion in asset markets, demonstrating the potential for granular innovations to propagate in the cross-section of stocks. We find that spillover intensity and assets’ relative importance are fast-changing and usually short-lived. Second, we test whether such a spillover network, combined with a concentrated size distribution, enable granular shocks to significantly influence aggregate market outcomes. The results indicate a positive significant effect from micro innovations to the aggregate market index. Thus, both findings bring into focus a potential missing link in financial market models — the inter-asset spillover network – which helps to connect micro-economic narratives to macro-financial outcomes.

2.1 Introduction

In empirical asset pricing, researchers often resort to broad market indices to explain individual asset returns. This practice is a reflection of the prevailing belief that aggregate factor movements result from exogenous macroeconomic shocks. Yet, surprisingly little is known about the driving forces of the market return. As a composite of many individual stocks, the aggregate market only responds to a shock if it affects a substantial portion of assets simultaneously. While it is true that aggregate shocks fit this criterion, we challenge the paradigm that such shocks are the only drivers of aggregate movement. More precisely, we also consider granular effects, that is, effects of micro-economic events on the aggregate. While common frameworks rule out granular effects under the assumptions of low concentration and independence of asset-level shocks, these conditions are unlikely met in reality. That is, diversification arguments break down under concentrated index weights or a dense spillover network between assets, rendering granular innovations a potential driver of market returns.

In this paper, we investigate the hypothesis that granular asset-level innovations spread through a spillover network to significant portions of the equity market, resulting in aggregate variation. That is, we test a network variant of the “granular hypothesis”, following the seminal contributions of Gabaix (2011) and Acemoglu, Carvalho, et al. (2012). We break our analysis of the research question into two parts. First, we seek to uncover a directed spillover network in asset markets. Provided such a network structure exists, we ask if it has, in the past, promoted the spread of granular innovations in the cross-section. More specifically, we want to understand the extent to which spillovers manifested and when they mattered the most. Second, we investigate the question if the spillover network paired with a concentrated size distribution provides a channel for granular innovations to affect the market aggregate. In other words, can we trace aggregate variation back to asset-level origins?

Answering this question has important implications for the understanding of financial markets. Should the aggregate market indeed respond to asset-level innovations, many additional influences come into consideration when tracing macro-financial variables. Essentially, this establishes a tangible link between the micro-level narratives of the corporate finance literature and the macro-economic perspectives of asset pricing. Then, with aggregate market movement resulting from the realizations of many smaller sources of risk, the market risk premium is itself truly a composite of individual risk premiums on a collection of aggregate and granular innovations. Put simply, the directed spillover network could be a neglected channel in common models of financial markets. Therefore, gaining clarity on what shocks drive market returns is key to understand what risks get renumerated by the market risk premium.

We leverage recent advances in econometrics to estimate directed networks of asset connectedness using daily realized volatility data. Our approach aims to capture the dynamic transmission of pricing-relevant information among assets, consistent with the idea of informational contagion. By analyzing the estimated networks, we identify prominent assets that play a significant role in disseminating information throughout the system. We link spillovers to the aggregate market through the affected assets' index weights and quantify the arrival of systemically important information at the asset level. Building on the obtained granular importances, we construct two residual indices that accumulate exaggerated impacts of granular innovations due to unequal weight distributions in the aggregate. In an empirical analysis, we then explore the effect of these residual indices on the overall market index, shedding light on the relationship between granular variation and aggregate market movement.

Addressing our research question is challenging as it requires inference on a sufficiently large part of the spillover network of inter-asset linkages. Given the vast cross-section of listed equities, the empirical estimation of such networks is a large-dimensional problem and requires the use of regularization techniques. Building on the results of Brunner and Hipp (2023), we employ the Adaptive Elastic Net (AENET) of Zou and Zhang (2009) and the graphical LASSO (GLASSO) of Friedman et al. (2008) to estimate comprehensive volatility spillover networks among the 100 most important assets in the style of Forecast Error Variance Decompositions (FEVDs) as in Diebold and Yilmaz (2014). Additionally, the direct assessment of granular effects on the aggregate market ideally necessitates shock identification at the granular level, which poses an severe empirical challenge. Hence, instead of engaging into the losing game of identifying asset-level shocks in the presence of spillovers, we adapt the concept of the “granular residual” as in Gabaix (2011) to capture common price movements due to relative differences in the systemic impact between assets.

We find significant positive effects of granular innovations that spread through the spillover network to the aggregate market. That is, our results indicate a causal relationship of asset-level innovations on the aggregate market index in line with the conceptual frameworks in Acemoglu, Ozdaglar, et al. (2015a), Acemoglu, Ozdaglar, et al. (2015b), and Acemoglu, Ozdaglar, et al. (2017). We also test if we find a similar effect for a granular residual based on the concentration of market capitalization. Contrary to the original granular hypothesis in Gabaix (2011), this granular residual is negatively related to the market aggregate, calling the role of the unequal distribution of index weights for granular effects into question.

To counter the possibility of omitted variables and reverse causality, we conduct two types of robustness checks: On the one hand, we condition on potentially important

aggregates, and on the other hand, we utilize pricing errors as instrumental variables for the granular innovation residuals, an approach loosely related to the recently introduced granular instrumental variables of Gabaix and Koijen (2020). While the significance of the positive coefficient on granular effects due to network spillovers remains robust in these specifications, the negative coefficient on the residual driven by index weights does not withstand all robustness checks.

Moreover, our descriptive analyses of the inter-asset spillover network and its historic evolution are novel to the literature. Our estimates reveal sizeable network effects in asset volatilities over rolling samples spanning from 1973 to 2022, consistent with the observed firm volatility co-movement discussed in Herskovic, Kelly, et al. (2016) and Herskovic, Kelly, et al. (2020). We document varying degrees of connectedness and network concentration, with peaks around the Black Monday market crash and the Great Recession. Different from previous finance literature that analyzed network effects in relation to asset prices on the basis of sectoral input–output tables such as Ahern (2013) and Herskovic (2018), we perform a more granular analysis on the asset-level. This feat allows us to look into the specific assets that spill new information into the market through the network, and link temporary increases in importance to capital markets and corporate financial events. We show that these assets change at a rapid rate, indicating that the reasons for outsized impact are short-lived. Hence, the negligence of possible asset-to-asset linkages in the canonical asset pricing framework is likely linked to their instability, and in consequence, the inability to uncover them by means with insufficient temporal and cross-sectional resolution.

We understand the documented spillover networks as a consequence of informational contagion in financial markets. As news about one asset becomes public, investors also readjust their valuations of other assets based on expectations formed with the new information. Therefore, under this umbrella, many economic mechanisms proposed in the literature can lead assets to inherit other asset’s volatility. For example, financial contagion between banks due to credit risk appears in Acemoglu, Ozdaglar, et al. (2015b) and Demirer et al. (2018), Hasler and Ornathanalai (2018) and Andrei et al. (2023) consider investor attention around announcement dates, Ahern (2017) documents information networks of individuals, and a large literature studies fundamental links in input–output networks (for instance Ahern (2013) and Carvalho (2014)) or customer–supplier networks (see Herskovic, Kelly, et al. (2020)). Our estimates capture asset-level connections into a single table without taking strong assumptions about the mechanisms that evoke the respective information spillover.

Finally, our work is connected to a large literature that studies connectedness in financial markets, contagion and systemic risk. Such research is often confined to

the banking sector (compare, for example, Billio et al. (2012), Acharya et al. (2017), Baruník and Křehlík (2018), Demirer et al. (2018)) but examples in the context of more general cross-firm linkages exist (see, for example, Bianchi et al. (2015)). In parallel to the insights in this branch of literature, our results suggest investors should pay attention to firm-level events with market-wide effects. A similar connection of micro-level uncertainty and market-wide outcomes has recently been drawn in Herskovic, Kind, et al. (2023). Further, temporary granular effects from potentially any asset challenge state-of-the-art asset pricing frameworks that assume a low-dimensional factor structure of aggregate shocks. Consequently, our results bear implications for the study of the market risk premium and for the interpretation of asset prices in reduced systems such as linear factor models.

The remainder of this research paper is structured as follows: Section 2.2 presents the methodology used to approximate the spillover networks. Section 2.3 exhibits the estimated networks and presents descriptive evidence. Finally, we empirically investigate the predictive power of the spillover network and discuss our findings in the light of classical asset pricing models in Section 2.4, before Section 2.5 concludes.

2.2 Empirical setting

2.2.1 Data and preprocessing

In this paper, we are interested in tracing the sources of variations in aggregate investor wealth. In contrast to most approaches in the finance literature, we focus on individual assets rather than common factors as the potential drivers of price movement. For that purpose, we collect financial data on the stock level from WRDS. More precisely, we gather daily prices and market capitalizations of all firms in the CRSP universe on a given date. Additionally, we obtain SIC codes from Compustat and follow the industry definitions of Fama and French (1997) to assign each asset to one of 12 sectors.¹

Intraday volatility data

We aim at quantifying directed spillovers of volatility in the cross-sectional and inter-temporal network of security returns. That is, we understand spillovers as a reaction of the investors of one asset to news about another, giving rise to lagged dependencies and correlations. While previous research has shown that there is little autocorrelation in daily returns, there is substantial evidence for the persistence of return volatilities,

¹The classification of each SIC code can be obtained from [French's data library](#).

giving rise to their forecastability.² Hence, we use asset-level volatility data at the daily frequency to infer granular connectedness tables while providing sufficient data for high-dimensional estimations.

Analogous to Barigozzi and Brownlees (2019), we opt for the extreme value method of Parkinson (1980) to measure intraday realized volatility:³

$$\tilde{\sigma}_{i,t} = \sqrt{0.3607} (p_{i,t}^{high} - p_{i,t}^{low}), \quad (2.1)$$

where $p_{i,t}^{high/low}$ is the natural logarithm of the intraday high and low price, respectively. To avoid zero values in our dataset for cases where the intraday high and low are identical, that is, days without any trading, we use the same range-based measure on bid and ask prices to proxy for disagreement as a lower bound of volatility.

In line with Andersen et al. (2003), we take the natural logarithms of the intraday realized volatilities to approximate a Gaussian distribution. Henceforth, let the resulting daily variable be denoted as:

$$s_{i,t} := \ln(\tilde{\sigma}_{i,t}). \quad (2.2)$$

To account for common-factor movement, we construct a common volatility index from daily log-volatilities. Due to the challenge of large weights in the common factor, we want to minimize the effect of a single asset on the resulting index. Hence, we obtain intraday volatilities of all available stocks in CRSP on a given day and standardize each time series by its rolling 1-year mean and standard deviation lagged by one month. Further, we apply equal weighting to all available stocks with positive intraday volatility to counteract the effect of concentration. Formally, the resulting daily index reads⁴

$$f_t = \frac{1}{|CRSP|} \sum_{i \in CRSP} \frac{s_{i,t} - \bar{s}_{i,t-13M:t-1M}}{\sigma(s_{i,t-13M:t-1M})}. \quad (2.3)$$

Sampling

Due to limitations with high-dimensional estimations, we constrain our samples to a subset of the full cross-section in the CRSP universe. Nevertheless, we aim at providing a representative analysis by proceeding as follows.

²Compare, for example, the literature on ARCH type processes. The effective log-volatility specification of our model is similar to the exponential GARCH model of Nelson (1991), which has proven to be particularly successful in forecasting daily volatility.

³Earlier works that employ similar data include Diebold and Yilmaz (2009), who study spillovers between international stock markets, Diebold and Yilmaz (2012), who build a spillover table of asset classes, and Barigozzi and Brownlees (2019), who employ intraday volatility data to study firm-level interconnectedness between 2004 and 2015.

⁴The resulting index is correlated with the VIX at 55%, but reaches back farther. Figure 2.9 in Appendix 2.A.1 shows its evolution over the sample.

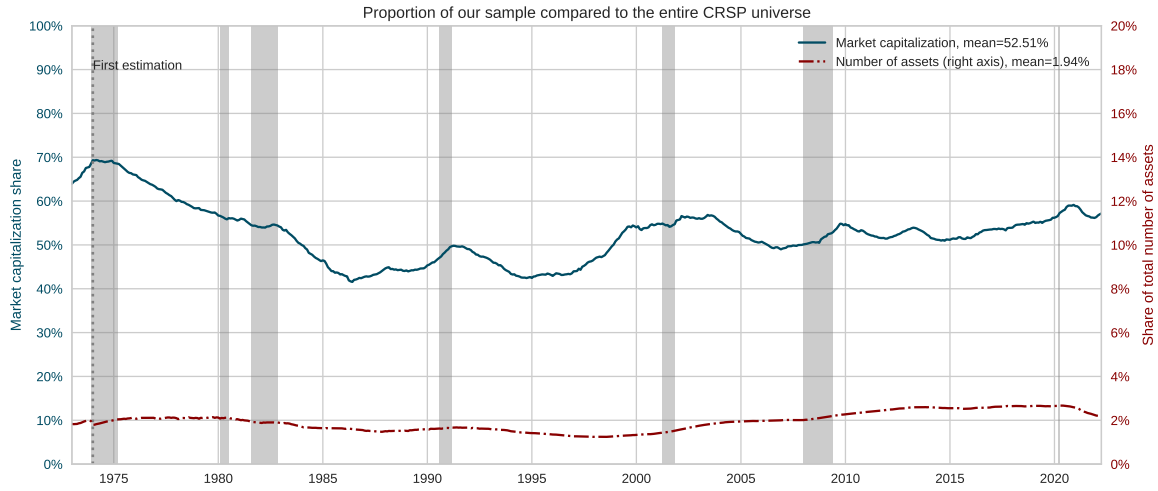


Figure 2.1: Proportion of the CRSP dataset contained in our sample. The total number of assets in CRSP ranges from 3795 in April 2020 to 8174 in June 1998. At the end of each month, we sample the 100 assets with the largest average market capitalization over the course of the previous 12 months. The shaded intervals indicate NBER recessions.

At the end of each calendar month, we sample the 100 equity tickers with the largest market capitalization over the previous 12 months from CRSP.⁵ That is, since we are ultimately interested in effects in investor wealth, the most relevant assets are the ones contributing most to the aggregate market portfolio. Figure 2.8 in Appendix 2.A.1 illustrates our sampling procedure. Note that there is a sampling bias towards well-performing assets for the estimation year up to the sampling date, whereas the subsequent evaluation period provides a valid sample for investors that conduct asset selection on the sampling date.

In summary, each data sample consists of the daily log-volatilities of the 100 equities that contribute the most variation to investor wealth (*ceteris paribus*, that is, ignoring inter-asset correlations) in a window of 12 months preceding a given sampling date. Given the unequal distribution of market value among firms in the cross-section of stocks, our samples capture a large part of the total market value in the CRSP universe. Figure 2.1 illustrates the relation of the number of assets to the total market share and volatility of our sample. After the inclusion of the NASDAQ into the CRSP dataset in 1972, although our sample contained only an average of 1.94% of the total number of assets, it made up 52.59% of the total market value on average.

⁵We consider securities with exchange code $\in \{1, 2, 3\}$ and share code $\in \{11, 12\}$. Also, we require assets to have an intraday high price and low price available for at least 99% of the trading days in the last 12 months. We exclude assets that are delisted on the sampling day and therefore do not have subsequent price information.

2.2.2 Estimation

The application of FEVD networks entails parameter estimation in high-dimensional space even with a relatively small slice of the full cross-section of equity returns. For example, estimating a network of 100 entities with intercepts and factor loadings with respect to a single common factor requires the estimation of VAR coefficients in $100 + 100 + 100 \times 100 = 10,200$ dimensions, plus the residual covariance matrix with $(100 \times 101)/2 = 5,050$ values. In the setting of a time-varying DGP, the amount of available time series observations is limited, further aggravating the inference problem. Put simply, classical estimation of 15,250 parameters in windows of $\sim 250 \times 100 = 25,000$ individual observations is inaccurate. In the context of FEVDs, this problematic has been discussed in Demirer et al. (2018) and further investigated in Chapter 1. The latter provides detailed guidance on how to improve the estimation of FEVD tables. In a nutshell, penalization of the parameter usage improves the accuracy of the estimation, and thus the overall confidence in the estimates. That is, the use of regularization of coefficient and covariance estimations improves bias, variance, and edge-detection for FEVDs. However, regularization comes at the cost of having to fix the hyperparameters used to penalize parameter usage ex-ante. Thus, we cross-validate the selection of hyperparameters to find the estimator with the best out-of-sample performance within the estimation period.

In the following, we discuss all steps of the estimation approach in detail. Additionally, Appendix 2.A.3 illustrates the complete procedure on the basis of the most recent sample period ending 31 March 2022 as an example and provides many auxiliary plots of intermediary outputs. Finally, we apply the described steps to a rolling time interval at the monthly frequency.

Large-dimensional vector auto-regressions

Given the pre-processed panel of intraday valuation volatilities, we estimate a VARX(1) model controlling for the equal-weighted variance index $f_t = \sum_{i \in CRSP} \frac{s_{i,t} - \bar{s}_i}{\sigma(s_i)}$ as a confounding factor that exogenously affects all assets as in classical linear factor models.

$$s_{i,t} = \mu_i + \beta_i f_t + \sum_{j=1}^N \gamma_{i,j} s_{j,t-1} + u_{i,t}. \quad (2.4)$$

We apply the Adaptive Elastic Net (AENET) estimator of Zou and Zhang (2009) to the problem, due to its desirable properties shown in Chapter 1. That is, we first standardize the data by subtracting the time-series means and dividing by the standard

deviations. Then the estimation equation of the AENET with hyperparameters λ, κ reads

$$(\hat{\mu}, \hat{\beta}, \hat{\gamma}) = \underset{(\mu, \beta, \gamma)}{\operatorname{argmin}} \left[\frac{1}{2NT} \sum_{i=1}^N \sum_{t=1}^T \left(s_{i,t} - \left(\mu_i + \sum_{k=1}^K \beta_{i,k} f_{k,t} + \sum_{j=1}^N \gamma_{i,j} s_{j,t-1} \right) \right)^2 + \lambda \sum_{i=1}^N \sum_{j=1}^N \hat{m}_{i,j} \left(\kappa |\gamma_{i,j}| + (1 - \kappa) \frac{1}{2} \gamma_{i,j}^2 \right) \right] \quad (2.5)$$

with $\hat{m}_{i,j} = |\hat{\gamma}_{i,j}^{ENet} + \frac{1}{NT}|^{-1}$ being the absolute inverse of the estimates of (2.4) by an Elastic Net with hyperparameters $\kappa = 0.0001$ and cross-validated λ . We add a small constant to the first-step estimates to avoid division by zero.⁶ The absolute (L1) and square (L2) penalty terms in (2.5) effectively result in variable selection and coefficient shrinkage to avoid overfitting and improve forecasting.⁷ Finally, we recover the coefficients in terms of the original data through reversion of the standardization step as detailed in Appendix 2.A.2.

Due to their direct relation with the coefficient estimates μ, β, γ in (2.4), it is crucial that the choice of the hyperparameters λ and κ takes into account the distribution of the respective dataset. Thus, we resort to 12-fold cross-validation for the hyperparameter selection before running the final estimation. More precisely, we split the panel data 12 ways along the time series dimension, corresponding to the number of months in the respective sample window. Then, always holding out one sub-sample for model validation, we estimate the model over the remaining 11 folds for each (λ, κ) -pair in a pre-specified grid. We validate the predictive performance of the resulting model on the hold-out fold using the mean squared error as the validation metric. Averaging over the 12 out-of-sample metrics, we compare estimators and pick the best out-of-sample performing pair of hyperparameters. Hence, as a first input to the construction of the FEVD table, we obtain an estimate of the VAR(1) coefficient matrix $\hat{A} = [\gamma_{i,j}]$.

Large-dimensional covariance matrices

As a second ingredient to the construction of our forecast error variance table, we require an estimate of the residual covariance matrix of the error terms $u_{i,t}$ in (2.4).

⁶We optimise (2.5) using the `glmnet` routine of Friedman et al. (2010). The python API to the algorithm is available at Balakumar et al. (2016).

⁷Note that in the original work of Zou and Zhang (2009), the estimation equation of the adaptive elastic net has slightly different form, with $\hat{m}_{i,j}$ only multiplying the L1 penalty term. However, earlier applications such as Rapach et al. (2013) and Demirer et al. (2018) optimize the version in (2.5) to make use of the computational efficiency of the `glmnet` algorithm. We adhere to that practice.

That is, we assume a multivariate Gaussian distribution of the time t residual vector u_t :

$$u_t \sim \mathcal{N}(0, \Sigma), \quad \forall t. \quad (2.6)$$

Similarly to the coefficient estimation, the residual covariance matrix estimation faces the challenge of high dimensionality. Guided by the comparative simulation results in Chapter 1, we apply the Graphical LASSO Estimator (GLASSO) of Banerjee et al. (2008) and Friedman et al. (2008). We denote the sample estimate of the residual covariance matrix Σ as $\hat{\Sigma}$. Now, the precision matrix, that is, the inverse covariance matrix $\Theta_u = \Sigma^{-1}$ with entries $\theta_{i,j}$, is related to conditional (partial) covariances (see Barigozzi and Brownlees (2019)). The GLASSO estimator seeks to induce sparsity in the precision matrix by adding an absolute (L1) penalty term to the log-likelihood optimization:

$$\hat{\Theta}_u = \underset{\Theta_u > 0 \in \mathbb{R}^{N \times N}}{\operatorname{argmax}} \left[\ln \det \Theta_u - \operatorname{tr} \left(\left(\frac{1}{T-2} \sum_{t=2}^T u_t u_t' \right) \Theta_u \right) - \rho \sum_{i=1}^N \sum_{j=1}^N |\theta_{i,j}| \right] \quad (2.7)$$

The expression optimises over the set of all nonnegative definite matrices, $\Theta_u > 0$, in $\mathbb{R}^{N \times N}$. As a result of ruling out direct effects in the precision matrix, the undirected graph of contemporaneous relations is estimated to be sparse. We believe that the sparsity in partial correlations, for example, the exclusion of some direct contemporaneous effects, squares well with the story of large dimensional asset networks. Note, however, that the corresponding covariance matrix is generally dense as it captures all linkages, including indirect linkages that arise through chains of direct linkages.

Similar to the hyperparameters of Section 2.2.2 λ and κ , we need to set the hyperparameter ρ in Equation (2.7). Again, we use 12-fold cross-validation to choose its value over a pre-defined grid of candidate values. The metric we apply to evaluate the estimators' performance is the average L2-norm of squared realizations to the covariance matrix estimate and has the following form for a sample of T regression residual vectors:

$$\ell = \frac{2}{TN(N-1)} \sum_{t=1}^T \sum_{i=1}^N \sum_{j=i}^N \left([\Sigma_u]_{i,j} - [u_t u_t']_{i,j} \right)^2. \quad (2.8)$$

We then use the best performing value for the hyperparameter ρ , that is, the value that minimizes Equation (2.8), to estimate Equation (2.7), and obtain the covariance matrix estimate as $\hat{\Sigma}_{glasso} = \hat{\Theta}_u^{-1}$.

Rolling window

Due to its simplicity and coherence, we apply a rolling-window approach to account for the dynamic nature of our dataset. Specifically, we opt for a window bandwidth of one

year to balance the data requirements of the regularized estimators with the objective of allowing time-variation in the estimates. While it is arguable that this window cannot fully capture swift changes in connectedness, it still allows us to frequently adapt to changes in the importance of the sampled tickers. Our main period of analysis is from December 1972 to March 2022; that is, all available periods after the inclusion of NASDAQ into the CRSP universe. We sweep through the sample in monthly steps using a backward-looking one-year window. In each estimation window, we re-estimate all parameters, including the hyperparameters λ, κ, ρ through cross-validation.

Overall, when monitoring several statistics related to the rolling window, we observe that the estimation is fairly stable throughout the sample periods. That is, hyperparameters vary only slightly resulting in low variation in parameter usage, and, hence, degrees of freedom. Sparsity is generally stronger in the residual covariance matrix at a level above 60%, while the VAR coefficient matrix possesses usually around 30% zero values. Additionally, our estimators usually shrink the included parameters by around 70% to 80% when compared to unregularized estimation, illustrating the need for regularization in this context to reduce estimation uncertainty. Finally, our regularized VARX(1) model ends up explaining around 30% to 70% of the total time-series variation, with a increasing tendency from around 2000. It is interesting to note that much of the recent increase stems from increasing explanatory power of the common factor rather than spillovers, where the latter has a very stable partial R^2 of around 30% over the common factor. We include detailed illustrations of the evolution of various estimation and regularization statistics in Appendix 2.A.4.

2.2.3 Network methodology: Generalized Forecast Error Variance Decompositions

For the construction of FEVDs we follow the idea of generalized FEVDs as in Pesaran and Shin (1998), which Diebold and Yilmaz (2014) popularized. We briefly touch on the calculations of the FEVDs and refer to Chapter 1 for a more detailed derivation. First, rewrite Equation (2.4) into vector notation:

$$s_t = \mu + \beta f_t + \Gamma s_{t-1} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma), \quad (2.9)$$

where μ is a vector of intercepts, β is the vector of asset-specific factor sensitivities and $\Gamma = [\gamma_{ij}]$ is the VAR coefficient of the first lag as in Equation (2.4). Then, the generalized impulse response, $IR(s_{i,t+h}|u_{j,t} = \sigma_{jj})$, is the response of i to a one standard deviation innovation at j . Its square defines the forecast error variance (FEV) at

forecast horizon H :

$$FEV_{ij}(H) = \sum_{h=0}^{H-1} IR(s_{i,t+h}|u_{j,t} = \sigma_{jj})^2 \quad (2.10)$$

Finally, the forecast error variance decomposition (FEVD) reads

$$FEVD_{j \rightarrow i}(H) = \frac{FEV_{ij}(H)}{\sum_{k=1}^N FEV_{ik}(H)} \quad (2.11)$$

The resulting table $D^H = FEVD_{j \rightarrow i}(H)$ is a network whose entries answer the question of what percentage of variation of i can be explained by variations in j . Pesaran and Shin (1998) implement an approximation of the zero period impulse response, by ordering each variable first in a Cholesky decomposition of Σ . However, as pointed out in Chapter 1, the decomposition is imperfect. Thus, we implement a decomposition of the residual covariance matrix Σ that has more desirable properties in the context of FEVDs. More precisely, we use the square root decomposition, that is, the square root of the eigendecomposition $\Sigma = V\Lambda V' \Rightarrow \Sigma^{0.5} = V\Lambda^{0.5}V'$. In simulations, this approach has shown advantageous in settings of large-dimensionality.

Further, we want to emphasize that spillovers from second moments only serve as a tool to make linkages among assets visible. Our analysis of spillovers is based on the assumption that the structure of the network is similar at the price level. In a nutshell, the FEVD networks proxy for information spillovers between equities to draw conclusions about the influence of each asset on the full cross-section of the stock market, which in turn proxies for investor wealth when aggregated. Finally, we use the estimation counterparts of Γ and Σ to approximate Equation (2.11) empirically.

2.3 The volatility spillover network

2.3.1 Defining asset-level connectedness

Using the estimates of the VAR coefficient and residual covariance matrices from Section 2.2.2 and 2.2.2, respectively, we construct the FEVD networks as detailed in Section 2.2.3. We interpret the FEVD tables as adjacency matrices to analyze interconnectedness through a network perspective, where connection strength follows from the relation of daily volatility time series. We set the forecast horizon to $H = 21$, corresponding to the contributions of one month's business days. Heuristically, the edges in the constructed networks quantify the contributions of an asset's intraday volatility to changes in other assets' intraday volatilities, consistent with the idea of information spillovers. Assets are deemed influential if they notably contribute to other assets' volatilities or have a large number of *outgoing* links. Conversely, assets

FEVD Network (2022-03-31)

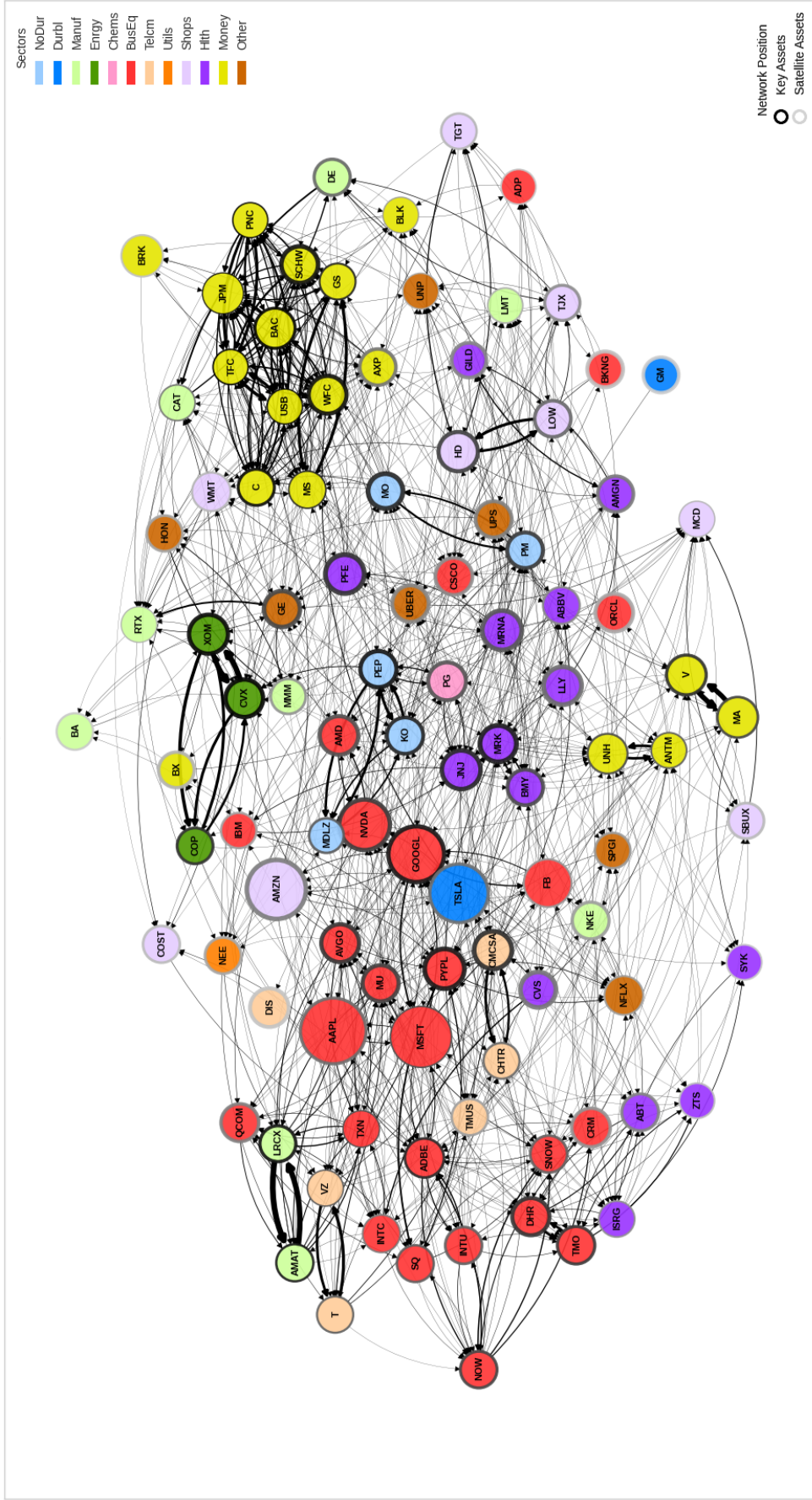


Figure 2.2: Network graph of the FEVD adjacency matrices in the one-year sample ending on 31-03-2022. The visualization is created with the help of the Fruchterman-Reingold force-directed algorithm (see Fruchterman and Reingold (1991)). Tickers are matched to one of 12 sectors according to their SIC codes in the Fama/French 12-sector industry classification. Edge thickness is determined by the strength of the respective directed link. We depict the 10% strongest connections, while node positions are determined by the full network. The nodes with the largest out-connectedness (key assets) are circled in black, and the nodes with the smallest out-connectedness (satellites) are circled in light grey. The thickness of the circle boundaries is determined by the net value of out-connectedness minus in-connectedness. A list of ticker-company name pairs is provided in Appendix 2.A.3.

are considered heavily influenced if they either have strong or numerous *incoming* connections.

The literature offers a multitude of narratives that could explain the relationships observed in the networks. For instance, financial contagion in banking networks has been subject to extensive study in the literature (see Billio et al. (2012), Acemoglu, Ozdaglar, et al. (2015b), Bianchi et al. (2015), Demirer et al. (2018)). Similarly, Hasler and Ornathanalai (2018) suggest that fluctuating investor attention can lead to the contagion of return and volatility, for example following earnings announcements as in Savor and Wilson (2016) and Andrei et al. (2023). Further, numerous works have focused on input–output linkages on the industry level (see Ahern (2013), Ahern and Harford (2014), Carvalho (2014)) or customer–supplier networks on the firm level directly (see Herskovic, Kelly, et al. (2020)) as the foundation for inter-dependencies. Other literature covers information networks of individuals (see Ozsoylev and Walden (2011), Ahern (2017)), knowledge spillovers (see Chuluun et al. (2017)), product and input market competition, and common exposure to regulatory and technology shocks (see Tambe and Hitt (2014)) as potential transmission channels between firms.

While all those mechanisms lead to inter-dependencies, the edges in our networks are most fittingly described by informational contagion. That is, as information becomes public in the context of a given asset, investors also readjust their expectations about the cash flows and risks of other assets. In other words, news about one asset reveals valuation-relevant information about other assets, triggering the contagion of information which we observe in the form of a volatility response. Hence, our analysis aims at measuring the strength of informational transmission channels, while remaining agnostic of the specific mechanisms that produce edges in the network across varying economic climates. The observed connectedness levels are consistent with a multitude of arguments in the literature for why assets are connected, and we do not advocate for one narrative over another. Instead, we focus solely on the ramifications of the spillovers' existence.

Figure 2.2 shows a graphical representation of the network estimated at the end of March 2022 based on an entire year of daily observations. A representation of the network before row-normalization is presented in Figure 2.21 in Appendix 2.A.3. Nodes are labeled by their respective tickers and color-coded according to the 12 Fama/French industry groups. It is observable that stocks belonging to the same industry loosely cluster together. The financial sector (Money) stands out with its pronounced within-sector interconnectedness. Yet, there is no clear separation between other industry groups, hinting at a dense interconnectedness structure in equity markets overall. A consequence of such interconnections is that granular events can rapidly amplify and

transcend alleged industry boundaries, rendering such events systemic.

2.3.2 Network position and influence

In order to illuminate the relation of firm-level and total variation, we want to identify the “key assets” in the spillover networks, following arguments in Ballester et al. (2006) and in Zenou (2016). The goal is to understand which assets contribute most significantly to the variability of other assets in the network. In other words, if we were to remove just one asset from the network, the removal of which asset would result in the most substantial reduction in the volatilities of other assets? To answer that question, the network literature presents a range of centrality measures, each serving to characterize the position of individual nodes. As we are working with directed networks such centrality measures can always be calculated in two directions. For example, to trace the origins of price volatility, one needs to follow the incoming links of a given node to find influential assets, that is, nodes with high *outgoing* centrality. Conversely, it can also be useful to follow the propagation of innovations along the direction of the network to identify the most exposed assets, that is, nodes with high *incoming* centrality. Hence, we employ the suffixes *in* and *out* to distinguish this dichotomy in the centrality measures.

Degree-centrality is defined by the magnitude of the edges connected to a particular node. Thus, the strength of the connections of a node is a reasonable measure of influence by itself. For example, suppose the strength of an edge proxies for the information flow between assets over multiple periods, it follows that a strongly *out*-connected asset disseminates more valuation-relevant information. Particularly, given the estimated adjacency matrix $\tilde{D} = [\tilde{d}_{ij}]$, let asset i 's *out*-degree deg_i^{out} be the column-wise or row-wise sum of off-diagonal connections:

$$deg_i^{out} = \sum_{j \neq i} \tilde{d}_{ji}. \quad (2.12)$$

Out-degree will consequently aid us to understand features of the estimated networks as a whole, and the cross-section of assets within the network. Analogously, let *in*-connectedness be defined as:

$$deg_i^{in} = \sum_{j \neq i} \tilde{d}_{ij}. \quad (2.13)$$

Note that there are more refined measures of centrality that take the cascading concept of networks further,⁸ but since our estimated FEVDs already aggregate spillovers over multiple periods in a single table, *out*-degree fits our context the best.

⁸For example, eigenvector centrality as suggested in Bonacich (1987) has been widely adapted in the economic networks literature, while the PageRank centrality algorithm of Page et al. (1999) paved the way for modern web search engines.

To understand the main influences in the FEVD networks, we look at the *out*-degree of individual tickers directly. For that purpose, we henceforth denote the assets with the largest *out*-connectedness as “key assets” and the assets with the smallest *out*-connectedness as “satellites”. Coming back to Figure 2.2 of the most recent period ending on 31 March 2022, we circle the assets with the strongest *out*-connections in darker tones and assets with the weakest *out*-connections in lighter tones. Contrary to the belief that large firms must have a strong influence on smaller firms, we find a more diverse set of asset to be driving the network during that period. Among the key assets are many firms of the financial and energy sector, but also particular firms of many sectors that take prominent positions within the respective sector. Examples of such are Pfizer (PFE), Merck (MRK), and Johnson & Johnson (JNJ) in the health sector, or Alphabet (GOOGL) and Paypal (PYPL) in tech, or Pepsi (PEP) and Coca-Cola (KO) in nondurable goods. On the other hand, satellites include conglomerates and holdings such as Berkshire Hathaway (BRK), Blackstone (BX), 3M Company (MMM), or Honeywell (HON), but also firms more dependent on the business cycle like Starbucks (SBUX), Costco (COST), Boeing (BA), General Motors (GM), or Booking Holdings (BKNG).

Next, it is worthwhile to track the influence of individual firms over time. This allows us to analyse extremes in the amount of volatility individual assets inject into the network. Figure 2.3 shows a lineplot of the *out*- and *in*-connectedness of all assets included in our samples. First, note that a strict factor structure as in most canonical asset pricing frameworks would imply all assets to have a connectedness of exactly zero at all times, which clearly does not hold for our estimates. Further, we see constituents of specific industries rise to disproportionate influence at different periods in time. For example, consistent with our knowledge of these times, tickers from the energy sector (Enrgy) were particularly important in the early 1980s, while financials (Money) became especially important around the Global Financial Crisis. In recent times, we see a fortified influence by firms of the tech sector (BusEq).

Additionally, there are frequent spikes in influence of individual firms that were involved in merger activities or went public at the time. Examples of this include Anheuser Busch (BUD) in 2008, or Reynolds American (RAI) in 2017. This observation is consistent with IPOs or M&A activity leading to reactions of uninvolved firms, especially when the companies involved are substantial constituents of widely tracked indices. For example, as exchange traded funds (ETFs) passively replicate these indices, a significant change in listings can force ETFs into strong rebalancing. This rebalancing could induce buying or selling pressure on the respective stocks, leading to spillover effects on other constituents, in line with arguments in Petajisto (2009).

Network Connectedness Following Diebold and Yilmaz (2014), we are interested in the average strength of the inter-asset network, that is, how much volatility is transmitted between assets through the network’s edges. Therefore, we obtain the average degree deg^{avg} by calculating the mean over all off-diagonal elements of the adjacency matrix as a measure of connectedness:

$$deg^{avg} = \frac{1}{N} \sum_i \sum_{j \neq i} \tilde{d}_{ij} = \frac{1}{N} \sum_i deg_i^{in} = \frac{1}{N} \sum_i deg_i^{out}. \quad (2.14)$$

This measure addresses the question of “what percentage of variation is explained by other stocks variations?”.

Network concentration While the overall network connectedness provides an intuitive interpretation, it is possibly washed out in the large- N setting. Thus, we include a second measure that focuses on the relative connection strength of individual nodes, summarizing the distribution of influence within the network. Note that economic networks commonly exhibit power law distributions in degree and centrality of its nodes.⁹ That is, there are usually many nodes with small scores and few nodes with extremely high values, resulting in high levels of concentration. Hence, it is interesting to study the level of concentration within the network. We do so by calculating the Herfindahl-Index over the *out*-degrees in the estimated networks as

$$conc^{out} = \sum_{i=1}^N \left(\frac{deg_i^{out}}{\sum_{i=1}^N deg_i^{out}} \right)^2. \quad (2.15)$$

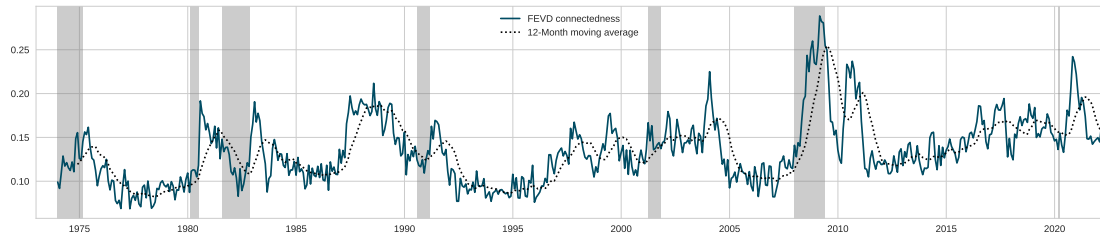
Figure 2.4 presents the evolution of key network metrics in our study: the average connectedness, 12-month rolling daily volatility and the network concentration of asset-level *out*-connectedness. Starting with average connectedness in Figure 2.4a, our analysis reveals some noteworthy peaks: Connectedness crosses the 0.2 line for the first time following October 19, 1987 - the date of the Black Monday market crash, Next, connectedness is high around the events of the burst of the dotcom bubble in the early 2000s, ending in a distinct peak. Most notably, however, is the repeated surge in connectedness during the Global Financial Crisis in 2008 and during the sovereign debt crisis in 2010, as has been documented in studies that focus on the financial sector (see Diebold and Yilmaz (2014), Demirer et al. (2018)). Finally, connectedness peaks again following the Covid-19 pandemic in 2020.

Unlike rolling volatility, the average connectedness does not mirror the width of the data window (compare Figure 2.4b). Instead, it appears to capture granular shocks

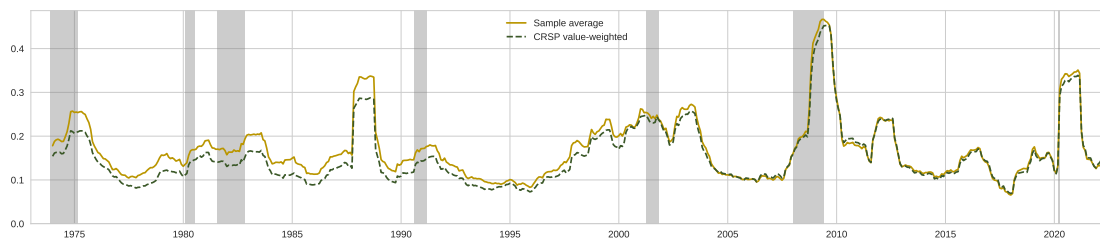
⁹See Gabaix (2016) for an empirical survey of power laws in economics, or Newman (2018), chapter 11.3 for the phenomenon of power laws in network degree distributions more generally.

that spread in the network and lead to correlations on top of propagated aggregate shocks. In 2010, for example, connectedness rises although there is no observable rise in aggregate volatility. Additionally, there appears to be an upward trend in average connectedness since then. Although the metric seems to be mean-reverting around 0.1 in earlier periods, it reaches a notably higher level recently. While this observation is compelling, we caution that it may be premature to establish reasons for this shift and warrant further research to unpack the intricacies of the recent trend.

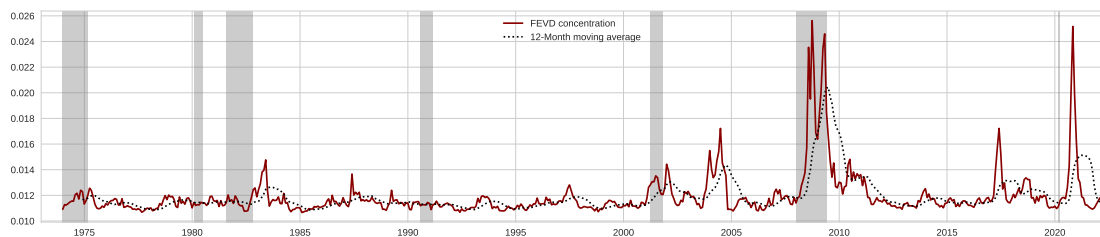
In Figure 2.4c, we document the average concentration. In contrast to average connectedness, this metric peaks more notably at dates where single industries experienced stress. That is, we see the highest peak early into the Global Financial Crisis in 2008 and the second highest peak after the onset of the Pandemic in 2020. These events relate to strong spillovers due to the banking sector meltdown and the tech sector hype, respectively. Moreover, it is worth considering the ramifications of these concentrated spillovers. Specifically, when a small group of assets is responsible for significant spillovers, they are more likely to exert an aggregate impact on the market portfolio. This brings us to the heart of our investigation, where we aim to scrutinize the implications of such concentrated spillovers.



(a) Average connectedness of the FEVD networks.



(b) 12-month rolling daily volatility.



(c) Concentration of asset-level *out*-connectedness in the FEVD networks.

Figure 2.4: Rolling network statistics for the FEVD networks. The shaded intervals indicate NBER recessions.

2.4 Network granularity and systemic effects

2.4.1 Systemically important assets

So far, we have examined the positions of assets within the network, while treating all connections the same, irrespective of the assets' market capitalization. However, to understand the impact of granular innovations on the market portfolio, it is essential to take into account not only the extent to which innovations affect other assets, but also the market weight of those affected assets. That is, the significance of a granular innovation to the wider economy depends not only on the degree with which it propagates to other assets, but also on the fraction of the total market capitalization tied up in the affected assets. Hence, when drawing wider conclusions about the influence of granular innovations on asset markets, it is crucial to account for the relative weights of assets in the market portfolio. Following Acemoglu, Ozdaglar, et al. (2015a), our next objective is to pinpoint assets that are systemically important — that is, assets with a disproportionately high impact on the stock market.

For that purpose, we left-multiply our estimated networks with a matrix that contains value-weights on its diagonal. Consequently, edge strength in the resulting weighted network Ω represents the portion of the market that is impacted by the corresponding link in the network:

$$\Omega = \text{diag}(w) D, \quad (2.16)$$

where w is a vector of weighting factors $w_i = \frac{Nv_i}{\sum_{i=1}^N v_i}$ that scale connections in the network proportionally to the market capitalization of the receiving asset v_i . Note that our samples of 100 assets are no complete representation of the full cross-section of equities, let alone the complete wealth portfolio. Nevertheless, as indicated in Figure 2.1 our selection represents a substantial share of investor wealth, allowing us to draw reasonable conclusions about aggregate effects.

To assess the relative impact granular shocks have on the aggregate, we sum over the weighted *outgoing* connections of each asset as in (2.12). Formally, this step relates to summing over the i th column in Ω ; $\text{deg}^{\text{out},\Omega}_i = \sum_{j \neq i} \omega_{ji}$, excluding elements on the diagonal. The metric thus quantifies the total spillover effects of nonsystematic innovations on the aggregate market portfolio. Intuitively, market impact $\text{deg}^{\text{out},\Omega}_i$ is an aggregated metric of the effect a granular innovation has through network spillovers weighted by the impacted asset's sizes. Consequently, the metric measures network effects, as outlined in Acemoglu, Carvalho, et al. (2012), while extending the idea of concentration effects as in Gabaix (2011) to impacted assets. To separate relative importances in each period from variation in overall network connectedness, we finally normalize network impacts to sum to one.

This weighted *out*-connectedness measure offers a new perspective on the potential for systemic risk caused by individual firms. Unlike conventional systemic risk measures that largely focus on the size of a firm, our measure also takes into account the directed spillovers from an asset, and more specifically, to larger assets in the system. As such, it presents a more nuanced understanding of the asset's systemic importance in the network. For example, traditional measures such as CoVaR by Adrian and Brunnermeier (2016) or SRISK proposed by Brownlees and Engle (2017) are concerned with bilateral effects of firms to the market and vice versa, often with size effects swaying the results. However, our measure suggests that it is not just size, but also the interconnectedness that should be factored into the systemic risk assessment. It aligns more with recent network-based measures such as the DebtRank proposed by Battiston et al. (2012), emphasizing that systemic risk emanates from the interconnections within the financial system. Consequently, our measure provides a tool to capture both the "too big to fail" and the "too connected to fail" aspect of systemic risk, presenting a more comprehensive view of an asset's potential impact on the market.

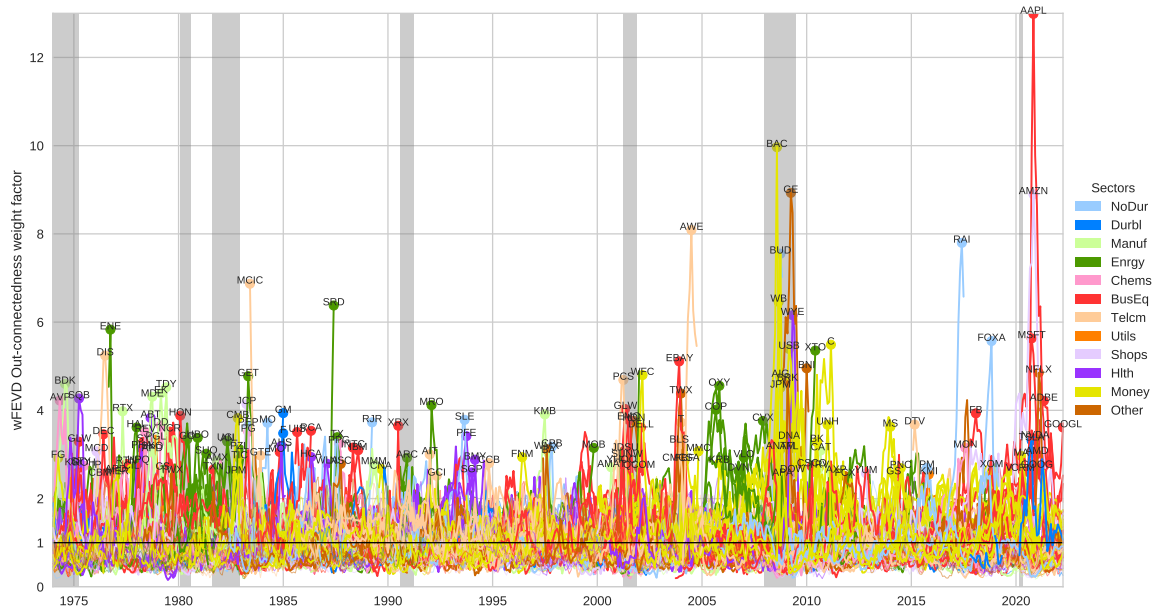
Figure 2.5 displays the market impact $deg_i^{out,\Omega}$ of individual assets over time. Given the fact that large firms tend to be influenced by other large firms in our networks, we observe a tendency for larger firms to have market impact. However, the assets with market impact radically differ from the ones with over-proportional market capitalization as argued in Gabaix (2011).¹⁰ Exaggerated market impact appears to only survive for a short period at a time, indicating a fast-changing nature of the spillover network. For example, there is a group of spikes of tickers from the energy sector in the early 1980s, while financial firms expectedly rise to prominence in the late 2010s. Further, we observe that a group of stocks in the technology sector spikes in recent times, potentially indicating systemically relevant market concentration in the later part of the sample. Such sectoral concentration of market impact is also visible in Figure 2.5b, which sums over firm-level impacts.

2.4.2 Characteristics of systemically important assets

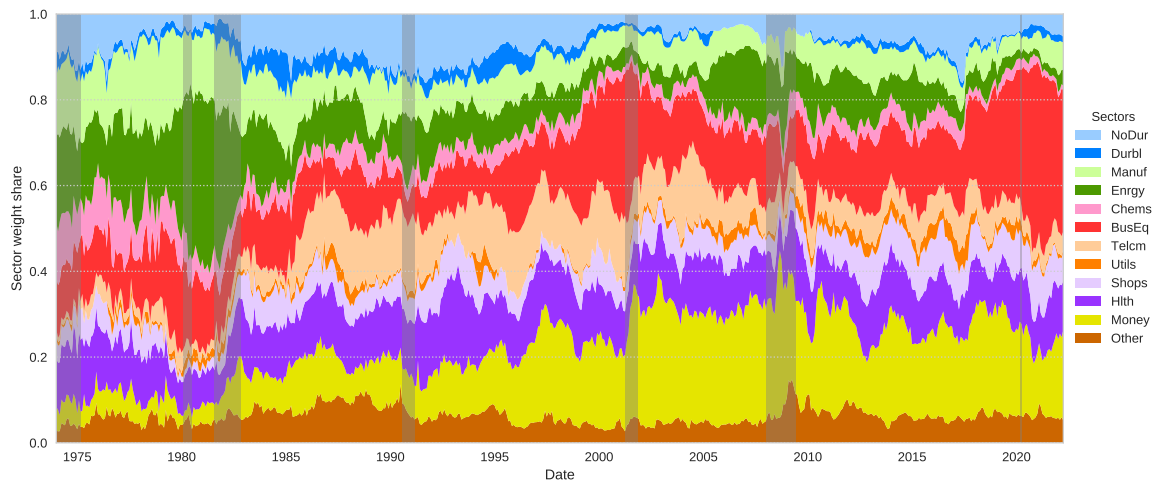
Next, we investigate asset characteristics by degree of influence. Table 2.1 contains average values for five groups of assets sorted by market impact $deg_i^{out,\Omega}$. Additionally, we report the difference between key assets and the satellite group together with t-statistics for the respective difference.

Panel A summarizes the asset's network position and composition of links. First, key assets have a noticeably stronger *out*-degree than the other groups. Somewhat counter-intuitively, influential assets tend to also have a more incoming connections,

¹⁰See Figure 2.24 in Appendix 2.A.1 for a corresponding plot.



(a) Individual asset-level weights.



(b) Aggregated sector-level weights.

Figure 2.5: Rolling spillover impact factors of individual assets. Spillover impact is calculated as the off-diagonal column sums in the value-weighted spillover network Ω , divided by the sum of all connections in Ω , that is, summing the market impacts of all *outgoing* connections. In the absence of spillovers, after rescaling all assets obtain an impact score equal to one. Hence, a score of two implies an impulse to the respective asset has twice the impact compared to the no-spillover case. Table 2.7 in Appendix 2.A.5 lists asset-level peaks and the corresponding company names for easier interpretation. The shaded intervals indicate NBER recessions.

additionally acting as amplifiers for other assets' variation. Further, it is sometimes helpful to characterize the influence a node has in the network by examining the sum of all edges that connect to it. That is, to calculate a given node's *net-degree* $deg_i^{net} = deg_i^{out} - deg_i^{in}$, we subtract all *incoming* links from the sum of all *outgoing* links. Heuristically, this metric is complementary to degree-centrality in that it captures the balance of *in-flow* and *out-flow* of volatility, such that the residual can be understood

	Market impact sorts (Ω -out-connectedness)					Difference	
	Satellite Assets	2	3	4	Key Assets	KMS	t-statistic
<i>Panel A: Network measures</i>							
out-degree	0.0802	0.1046	0.1244	0.1496	0.2169	0.1367***	19.1195
in-degree	0.1023	0.1166	0.1298	0.1472	0.1797	0.0774***	22.5218
net-degree	-0.0222	-0.0121	-0.0054	0.0024	0.0372	0.0593***	12.5646
in-concentration	0.0418	0.0426	0.0434	0.0445	0.0495	0.0076***	6.0826
out-concentration	0.0454	0.0449	0.0449	0.0456	0.0460	0.0006*	1.8993
<i>Panel B: Data</i>							
realized volatility s_i	-5.2929	-5.2994	-5.2975	-5.2908	-5.2591	0.0337***	2.7930
volatility-of-volatility $\sigma(s_i)$	0.4543	0.4609	0.4641	0.4669	0.4775	0.0232***	9.5006
market capitalization $\log(v_i)$	16.9387	17.0141	17.0796	17.1789	17.4009	0.4622***	11.9952
<i>Panel C: VARX estimates</i>							
intercept μ_i	-4.6096	-4.5569	-4.5561	-4.5479	-4.4077	0.2019***	5.3769
factor loading β_i	0.7263	0.7105	0.7047	0.6975	0.6783	-0.0480***	-5.2071
average spillover loading $\frac{1}{N-1} \sum_{j \neq i} \gamma_{ij}$	0.0131	0.0134	0.0136	0.0138	0.0144	0.0013***	10.7083
residual variance	0.1218	0.1268	0.1287	0.1293	0.1319	0.0100***	9.1399
average residual correlation	0.0431	0.0531	0.0605	0.0679	0.0765	0.0334***	21.4302
coefficient of determination R^2	0.4029	0.3977	0.3969	0.4007	0.4115	0.0086***	2.8702
partial R^2 : factors	0.2397	0.2291	0.2254	0.2244	0.2177	-0.0220***	-6.2493
partial R^2 : spillovers	0.2406	0.2445	0.2482	0.2563	0.2775	0.0368***	10.8332
<i>Panel D: Estimation year performance</i>							
annual return r^e	0.0971	0.1156	0.1159	0.1227	0.1309	0.0339**	2.4085
annualized volatility $\sigma(r^e)$	0.2921	0.2955	0.2991	0.3056	0.3259	0.0338***	4.7568
CAPM β_M	1.0093	1.0293	1.0563	1.0974	1.1756	0.1662***	7.1262
CAPM α (annual)	0.0242	0.0395	0.0397	0.0450	0.0432	0.0190**	2.2044
CAPM $\sigma(\varepsilon)$	0.2399	0.2404	0.2403	0.2427	0.2572	0.0174***	3.2474
CAPM R^2	0.3052	0.3161	0.3287	0.3431	0.3577	0.0525***	9.1517
<i>Panel E: Subsequent year performance</i>							
annual return r^e	0.0861	0.0806	0.0797	0.0801	0.0816	-0.0044	-0.4437
annualized volatility $\sigma(r^e)$	0.2970	0.3028	0.3071	0.3168	0.3310	0.0340***	3.8893
CAPM β_M	0.9982	1.0173	1.0445	1.0801	1.1482	0.1501***	6.3022
CAPM α (annual)	0.0062	0.0053	0.0063	0.0103	0.0139	0.0078	0.8635
CAPM $\sigma(\varepsilon)$	0.2459	0.2491	0.2502	0.2560	0.2644	0.0184**	2.3958
CAPM R^2	0.2996	0.3093	0.3207	0.3333	0.3486	0.0489***	8.5749

Table 2.1: Portfolio sorts on the market impact of individual assets. Market impact is calculated as the column sum in the value-weighted spillover network Ω , excluding diagonal elements. Assets are sorted into five portfolios for each of the 580 estimations between 1973-12-31 and 2022-03-31. We average statistics within each portfolio of 20 assets and form a long-short portfolio “key assets minus satellites”. We use Newey-West standard errors with $T^{1/4}$ lags to calculate t-statistics for the respective time series mean $\bar{x}$ against the null hypothesis $H_0 : \bar{x} = 0$. Returns, return volatilities and alphas are annual. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

as a weight adjustment to an node’s volatility in the overall network. Again, key assets have larger values for this metric, indicating an over-proportional representation in the total volatility. Finally, we calculate Herfindahl indices over all *incoming* and *outgoing* connections of node i as *in-concentration* $conc_i^{out} = \sum_{j \neq i} \left(\frac{\tilde{d}_{ji}}{\sum_{k=1}^N \tilde{d}_{ki}} \right)^2$ and *out-concentration* $conc_i^{in} = \sum_{j \neq i} \left(\frac{\tilde{d}_{ij}}{\sum_{k=1}^N \tilde{d}_{ik}} \right)^2$. In line with Herskovic, Kelly, et al. (2020) who base their analysis on the concentration of a firm’s customer network, we find that

more strongly connected firms also tend to have more concentrated *incoming* links. The evidence on the concentration of *outgoing* links is not conclusive.

In Panel B, the data summary statistics show higher average levels of realized volatility for more impactful assets, that is, key assets tend to be more volatile. Moreover, influential firms have a higher market capitalization on average, suggesting that large firms are more likely to have significant influence on other large firms. Turning to Panel C, which presents VARX and covariance estimates, that is, μ , β , Γ and Σ in (2.9), influential assets have larger values for the intercept and the residual variance, on average. However, the loading for the common factor is larger for satellite assets, consistent with the idea that key assets have a stringer idiosyncratic component. Consistent with stronger network linkages both contemporaneously and lagged, the mean spillover loading and average residual correlation with other assets is also higher for key assets. Interestingly, the common factor is also the dominant explanation for the variation of satellite assets, while key assets tend to have higher partial R^2 values for spillovers. This finding suggests that a factor explanation is more appropriate for satellite assets, while network spillovers are crucial for key assets, disputing the use of standard factor models for such assets.

Panel D and E offer a comparison of in-sample returns during the estimation year and out-of-sample returns in the subsequent year after estimation, respectively. We do this by calculating the total return and return volatility of each asset in the respective period. Additionally, we run CAPM regressions and obtain alphas, market betas, residual volatilities and goodness of fit statistics for both periods. Given the sampling bias we have in the estimation year, it is not surprising that the included assets earn above average returns and alphas during that period, with a significant outperformance for the key assets. Interestingly, this finding suggests that the volatility that positions an asset central in the network is often the result of positive news, rather than price drops. Subsequently, key assets experience insignificantly lower returns than satellite assets. Throughout, key assets experience higher total and nonsystematic return volatilities combined with higher market betas and CAPM R^2 values. The fact that the sign reverses for the beta loadings in second moments in Panel C and first moments in Panels D and E suggests that market beta could be a consequence of an assets own variation turning systemic.

2.4.3 From the granular residual to the network residual

Our goal is to investigate if and to what extent nonsystematic innovations on the firm level produce fluctuations in the aggregate asset market through the spillover network. This endeavour is challenging because firm-level innovations are only identified

under strict model assumptions such as a linear factor structure, which contradicts the existence of direct asset-to-asset spillovers. Without identification of exogenous firm-level shocks, however, causal arguments are subject to the criticism of Manski (1993); that is, granular fluctuations could reflect aggregate innovations rather than effectuate such. Instead of drawing a sophisticated scheme to still identify asset-level shocks, we follow the approach of Gabaix (2011) to construct two residual indices that are sensitive to granular innovations with potentially systemic impacts. As systemic effects of asset-level innovations could arise either due to large index weights or due to network spillovers, it is useful to separate the network amplification effect from the weight effect and construct separate residual indices for the component driven by the distribution of value-weights, and a network-driven component. More precisely, we adopt the concept of the “granular residual” that captures market movements due to the unequal size distribution of firms. Additionally, we introduce a second residual that reacts to innovations in systemically important assets based on the spillover network. We will then attempt to mitigate the “reflection problem” by employing econometric procedures in Section 2.4.4.

As a first gauge of granular effects, the *granular residual* captures market movements that stem from movements in assets with excess weights. Formally, it is given by:

$$\eta_t^g = \sum_{i=1}^N \left(\frac{v_{i,t}}{\sum_{j=1}^N v_{j,t}} - \frac{1}{N} \right) r_{it}. \quad (2.17)$$

That is, the granular residual is equivalent to a zero-investment portfolio with weights that increase in market capitalization, where the size of the total position depends on the size-distribution.¹¹ Intuitively, the granular residual captures asset-level innovations that do not average out in the aggregate portfolio due to the skewed distribution of market capitalization. The subtraction of equally-weighted returns approximately controls for aggregate shocks that affect all assets similarly, while the resulting residual weights capture the additional contributions of granular innovations to assets with large weights. Note that such granular residuals have been employed in other contexts to explain aggregate quantities such as GDP and productivity growth in Gabaix (2011), asset prices in Malevergne et al. (2009), banking distress in Blank et al. (2009), and industrial production in Foerster et al. (2011).

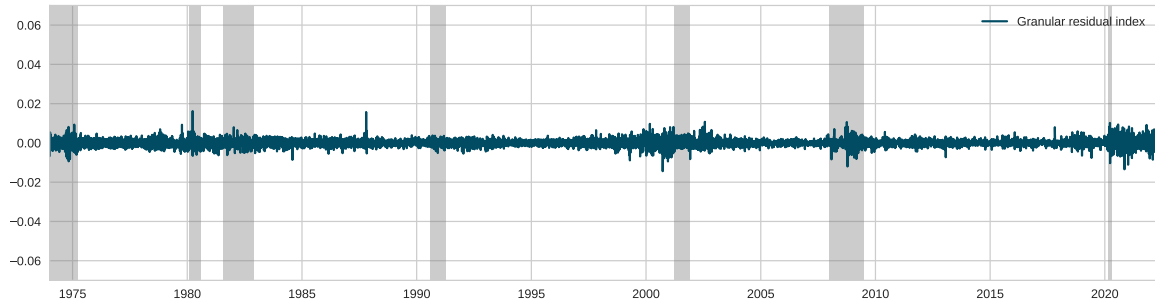
Second, we generate a residual index that distills deviations in the aggregate market stemming from network spillovers. As discussed before, an asset’s aggregate effect is

¹¹Ideally, we would like to construct a weighted index of granular shocks directly. However, we do not have shock identification in our framework. The suggestion of Gabaix (2011) to use factor model residuals is unsuitable for identifying granular shocks in the presence of spillovers. Moreover, factoring out the market aggregate from individual returns risks removing the exact effects that we want to show.

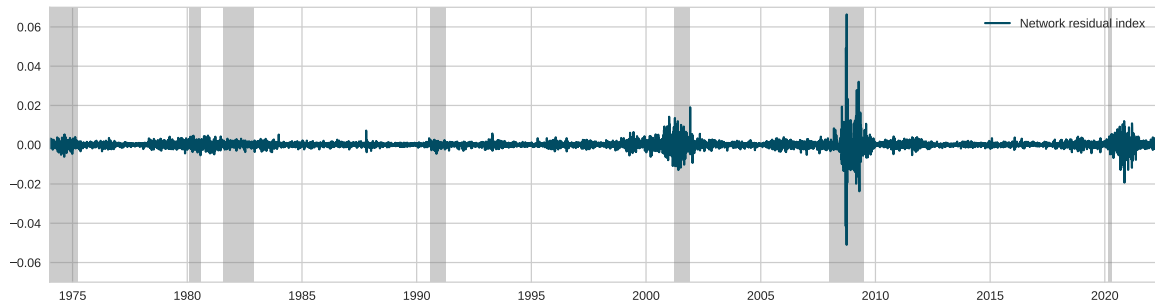
amplified if its *outgoing* spillovers to assets with large weights are sufficiently large. Hence, the additional network-induced wealth impact of each asset can be derived from the weighted network Ω by summing over all outgoing connections. That is, we revisit the concept of *out-degree* as in (2.12) — this time applied to the weighted network Ω — to calculate relative weights in each period. From the resulting influence weights we subtract equal weights as the baseline of a perfectly symmetric network with uniformly distributed contributions. Hence, these weights capture the difference in distribution of network influence from a scenario where all assets are equally influential, in which case all weights will be zero. The resulting *network residual* reads

$$\eta_t^{nw} = \sum_{i=1}^N \left(\frac{\deg_{i,t}^{out,\Omega}}{\sum_{j=1}^N \deg_{j,t}^{out,\Omega}} - \frac{1}{N} \right) r_{it}, \quad (2.18)$$

where $\deg_{i,t}^{out,\Omega}$ is asset i 's *out-connectedness* as in (2.12) on the weighted network Ω . Intuitively, this residual isolates granular effects due to network amplification. That is — consistent Acemoglu, Carvalho, et al. (2012, Theorem 4) who makes out the imbalance of the network structure as a precondition for network effects on the aggregate — it captures unequal exaggerations of granular innovations through the spillover network.



(a) Daily returns on the granular residual η^g .



(b) Daily returns on the network residual η^{nw} .

Figure 2.6: Daily returns on the granular and network residuals. The shaded intervals indicate NBER recessions.

Figure 2.6 displays the daily returns on the two residual indices over the sample period from January 1974 to March 2022. Consistent with Table 2.1, important assets

in the network tend to have larger market capitalization weights, such that some granular events at firms with large weights in (2.17) also appear in (2.18) – hence, the series have a positive correlation of 0.22, while both their means are statistically indistinguishable from zero. Nevertheless, the most distinctive daily returns appear at different times and therefore differ for the two series. In particular, the network residual shows characteristic periods of heightened volatility during the dot-com crash, the Global Financial Crisis, and the COVID-19 pandemic. While the granular residual also captures the heightened volatility level during these episodes, volatility increases are more distinct for the network residual.

2.4.4 Do granular innovations move the market?

Although the return on the market portfolio plays a key role in empirical asset pricing models, there is little evidence regarding the influences causing stock market fluctuations, particularly at higher frequencies. Even though numerous variables have been extensively explored to predict market returns over longer horizons (see Welch and Goyal (2008)), the causes of short-term stock market volatility are often broadly classified as “aggregate”. At the same time, Shiller (1980)’s excess volatility puzzle – the phenomenon of asset prices being significantly more volatile than their fundamental value – remains mostly unexplained. Campbell and Ammer (1993) attempts to explore the roots of those aggregate shocks by attributing a portion of stock market returns to news about future dividends, yet leaving a large proportion of monthly returns unexplained.

Given the efficiency of modern markets, even more frequent price updates likely reflect permanent adjustments to the arrival of new information rather than only short-term trading noise. While it is certainly true that part of the price-relevant information results from innovations on an aggregate level such as changes to macro-economic variables, we believe micro-economic innovations also exert a significant effect on market returns through the concentration of weights and network spillovers. Thus, we aim to test the hypothesis that granular innovations impact market returns through the granular residual η^g and the network residual η^{nw} .

We run time series regressions of the form

$$r_t^{mkt} = b_0 + b_1\eta_t^g + b_2\eta_t^{nw} + u_t, \quad (2.19)$$

where r_t^{mkt} is the return on the value-weighted CRSP index as a proxy for the market portfolio return. First, let us consider ordinary least squares regressions of (2.19) at daily and monthly frequencies over the full sample period. Table 2.2 shows the results of univariate and bivariate regressions of the market excess returns on the residuals.

Together, the granular and network residuals explain close to 10% of daily and 14% of monthly market returns. The network residual has the stronger contribution, which is also evident in the univariate regressions. Further, the network residual is significantly linked to aggregate returns with positive sign, supporting the hypothesis that granular innovations have systemic effects when fueled by the spillover network. At the same time, the granular residual affects aggregate variation negatively, where conditioning on spillover effects in the form of the network residual accentuates the effect. The negative sign is somewhat consistent with the size effect, although the tickers in our samples all qualify as big firms relative to the full cross-section of stocks. Hence overall, the results in Table 2.2 suggest that granular innovations have measurable effects on the aggregate market portfolio when considering network spillovers.

	daily			monthly		
	(1)	(2)	(3)	(4)	(5)	(6)
intercept	0.0007*** (7.4632)	0.0007*** (7.2948)	0.0007*** (7.4536)	0.0133*** (7.5713)	0.0135*** (7.3032)	0.0143*** (7.8791)
granular residual η^g	-0.7369*** (-5.7240)	-0.3023** (-2.2678)		-1.7057*** (-6.4028)	-1.3028*** (-4.2869)	
network residual η^{nw}	1.7002*** (9.4474)		1.5664*** (10.3712)	2.1670*** (7.7928)		1.7275*** (5.6376)
T	12359	12359	12359	588	588	588
R-squared	0.1073	0.0025	0.0933	0.1667	0.0575	0.0729

Table 2.2: Estimates and t-statistics of time series regressions of the form $r_t^{mkt} = b_0 + b_1\eta_t^g + b_2\eta_t^{nw} + u_t$ over the period from 2 January 1974 to 31 March 2022. The dependent variable r_t^{mkt} is the return on the value-weighted CRSP index. t-statistics are calculated from Newey-West standard errors with $T^{1/4}$ lags. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Next, we refine the time period of our analysis by performing the same time series regression as before over a rolling time interval. That is, we want to understand if the relations hold throughout our sample or is driven only by short sub-periods. Additionally, it is helpful to understand at what points in our sample period granular effects on the aggregate market were particularly strong. For that purpose, we report rolling regression statistics in Figure 2.7. First, the negative coefficient on the granular residual η^g seems to be largely driven by periods at the beginning of the sample and after the Global Financial Crisis. On the other hand, the network residual η^{nw} has a positive relation with the overall market throughout, with the exception of the early 1990s. Interestingly, the share of explained variation varies significantly over time, peaking at over 40% shortly after the Financial Crisis likely due to the systemic impact of a small set of firms during that time. Similar but weaker episodes occurred in the early 1980s and early 2000s, possibly related to the influence of energy and new economy stocks, respectively. Given these pockets of heightened explanatory power,

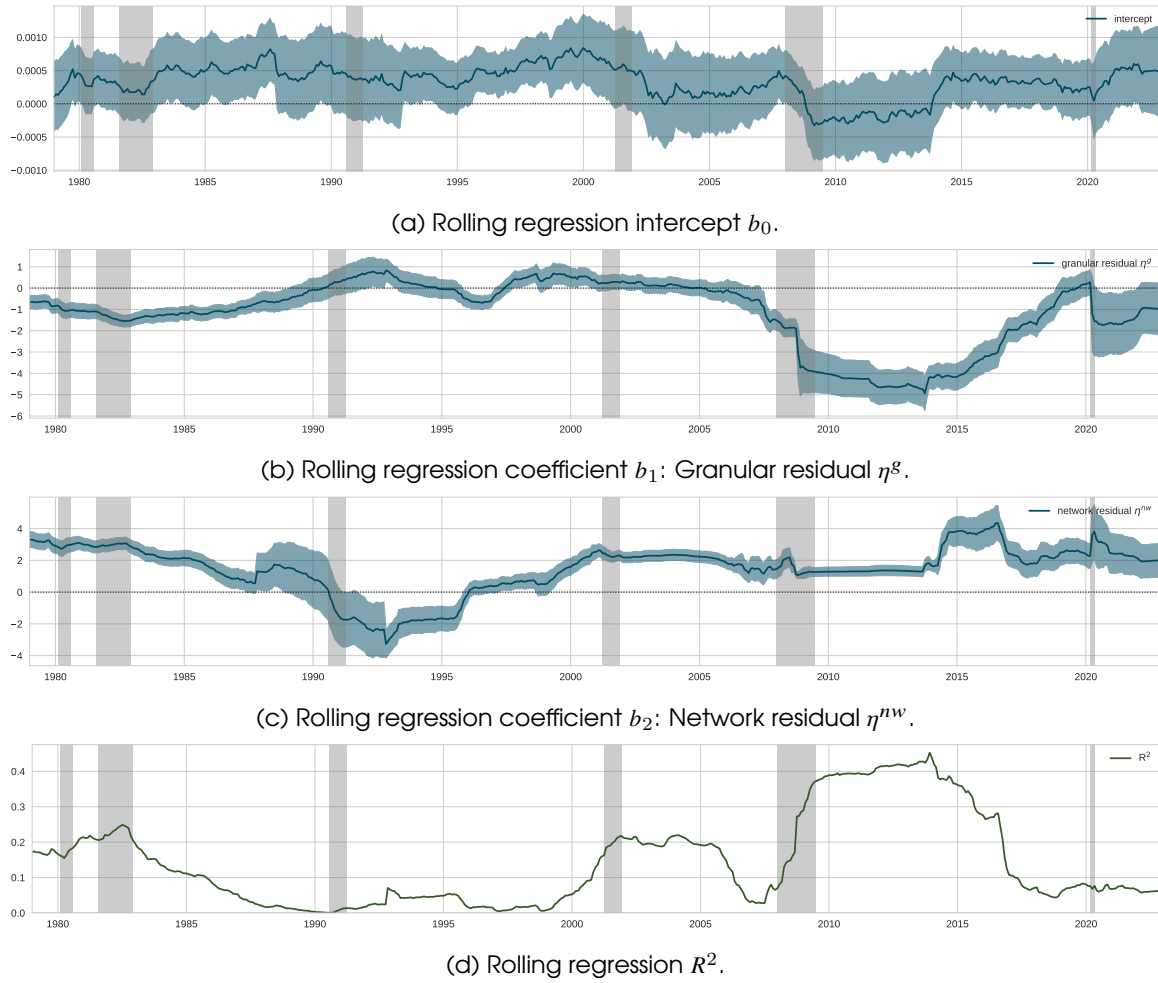


Figure 2.7: Daily OLS regressions in 5-year windows rolling at monthly frequency. Panel A to C display coefficient estimates and 95% confidence bands, Panel D shows the coefficient of determination. t-statistics are calculated from Newey-West standard errors with $T^{1/4}$ lags. The dependent variable r^{mkt} is the return on the value-weighted CRSP index. The shaded intervals indicate NBER recessions.

granular effects on the aggregate market are a particularly important explanation of market movement during times of distress, and promise to significantly contribute to the understanding of such episodes.

2.4.5 Robustness checks

There are some potential problems with the specification in Section 2.4.4 that we would like to address. First, there could be aggregate shocks that simultaneously determine the movement of the aggregate market as well as movement of the granular and network residuals. If omitted from the regressions, this could lead to false conclusions about the effects of the granular innovations on the aggregate. Second, standard asset pricing theory suggests an alternate explanation for the results in Table 2.2: it is possible that the weights in our residual indices are correlated with assets' exposures

to aggregate shocks. In that case, the residual indices would themselves be exposed to the aggregate market return rather than granular innovations affecting the aggregate index. That is, we may face a reverse causality problem with our specification, where high connectedness values of some assets are simply a consequence of a factor structure that transmits aggregate shocks to the the residuals rather than the other way around. To address these concerns, we consider two approaches: First, we include additional conditioning variables in our main specification, and second, we employ instrumental variables to investigate the direction of the effect.

To address the possibility of omitted aggregate shocks, we include various sets of control variables in Table 2.3. More specifically, we include a set of aggregate price variables, uncertainty indices, treasury yields at different maturities, various slopes of the yield curve, and a group common asset pricing factors as conditioning variables in our regressions. Clearly, these control variables are not necessarily exogenous to the performance of the stock market, but they are suitable conditioning variables to assess the stability of the effect of the granular and network residuals nonetheless. Some observations from the table stand out. First, the coefficient on the network residual seems to be robust to all included control variables and stays highly significant with an effect size larger than one in most cases. Second, the negative effect of granular residual by itself gets at least partially crowded out by some macro variables such as yields, and only stays significant when additionally including network spillovers. Third, the regressions in Panel B achieve R^2 values around 50% due to the inclusion of the VIX index, which is likely endogenously linked to movements of the aggregate market. Finally, including common asset pricing factors in Panel E also holds increased explanatory power with regards to the aggregate market, despite these additional factors arguably being independent of market movement. To conclude, our main result on the network residual appears robust to the inclusion of additional macro-economic variables.

Next, given the potential for remaining aggregate influences that we may not have captured with these control variables we ultimately employ an instrumental variables (IV) approach to affirm the causal nature of the regression in (2.19). In other words, with appropriate instruments for the residual indices, we can preclude reverse causality as an alternate explanation for the results in Table 2.2. Interestingly, the classic capital asset pricing model (CAPM) provides a set of instruments with desired properties. That is, we incidentally obtain suitable instruments to estimate the causal effect of the residual variables on the market portfolio by referring to the pricing errors of the market-CAPM. More precisely, the prediction errors of the market-CAPM model are exogenous to the market returns r^{mkt} , and hence meet the exclusion restriction as

instruments for the granular and network residuals. Simultaneously, the pricing error series likely still comprise part of the granular innovations that drive the variation in the residual indices, and therefore have a sufficiently strong first stage. Note that, though related, this IV approach is different from the granular IV approach of Gabaix and Koijen (2020),¹² as we do not use the granular residual to instrument the effect of the aggregate quantity on another variable. Rather, we employ granular innovations that are exogenous to the aggregate quantity to estimate the effect of the granular residual on the aggregate market.

Formally, for each asset included in the respective sample, we run CAPM time series factor regressions in the same period as the spillover networks. We then use the fitted model to calculate prediction errors for the asset returns included in the granular and network residuals. As a result, we obtain a set of 100 instruments per period that contains variation that is exogenous to the market variable. We assemble the pricing errors $\varepsilon_{i,t}$ of all assets i and all periods t in the instrument matrix $Z = [\varepsilon_{i,t}]$. Hence, Z is a matrix that contains a number of columns equal to the number of assets that ever appear in our samples, where each row contains 100 valid nonzero observations. We fill the missing observations by the time series averages to ignore them in the estimation. Given the large number of weak instruments in Z , in addition to the standard two-stage least squares IV estimator, we apply the many-instruments estimator with the standard errors described in Hansen et al. (2008). Note that this approach runs the risk of factoring out too much relevant variation in its instruments, possibly leaving too little of the granular innovations to draw sufficient inference, and thus, resulting in overly conservative estimates. More precisely, the CAPM regressions implicitly assume the variation in the market index to be exogenous to individual assets, which stands in contradiction to our hypothesis. Nevertheless, there may still be sufficient granular-level variation contained in the residuals. At the same time, the asset-level errors are also different from the granular innovations in a network connectedness model since they ignore the effect of spillovers.

Panel A and B in Table 2.4 contain the results of the same regression as in Table 2.2, but residual indices are instrumented by CAPM prediction errors. Panel A shows estimates of two-stage least squares IV regressions, and Panel B estimates the same specifications using the many-IV estimator. While standard IV confirms our previous findings, the many-IV estimates place heavy penalties on the coefficients and standard errors, leaving us with insufficient statistical power to infer significant effects in most specifications. As a result, while preserving the signs of the effects, the estimator shrinks the coefficients heavily towards zero. Hence, despite somewhat

¹²Granular IVs are used for example in Gabaix and Koijen (2021) to instrument for capital flows, or in Galaasen et al. (2020) to instrument for credit risks.

lacking conclusiveness, this IV approach provides minor support for our main findings concerning network amplification of granular effects into the aggregate.

Finally, we consider another instrument that we construct from announcement date returns. Based on the reasoning that some of the information spillovers in our networks originate from announcing firms as in Savor and Wilson (2016), we construct additional indices as in (2.17) and (2.18), but use only announcement day returns instead of r_{it} . The resulting series therefore capture firm-level news which should be exogenous to the aggregate market index. Further, using these instruments mitigates the difficulty of including many instruments for a single variable. Panel C in Table 2.4 contains the results. While announcement day returns are likely unsuitable instruments for the monthly regressions, the positive coefficient on the network residual at the daily frequency confirms our main result: Granular innovations propagate through a network of direct asset-to-asset linkages, such that individual assets temporarily gain systemic importance, and therefore impact the outcome of the aggregate market.

	daily			monthly		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: Conditioning on macro indices (DXY U.S. Dollar index, WTI Crude oil, TNX 10 year yield index)</i>						
intercept	0.0007*** (6.7447)	0.0007*** (6.7424)	0.0007*** (6.7355)	0.0140*** (7.3128)	0.0144*** (7.2849)	0.0146*** (7.5341)
granular residual η^g	-0.5064*** (-3.2652)	-0.0355 (-0.2194)		-1.1990*** (-3.7571)	-0.7599** (-2.1179)	
network residual η^{nw}	1.5224*** (9.1649)		1.4359*** (9.7531)	1.6828*** (6.3072)		1.3409*** (3.9418)
T	9282	9282	9282	444	444	444
R-squared	0.1499	0.0580	0.1443	0.1366	0.0619	0.0951
<i>Panel B: Conditioning on macro uncertainty (TED spread, VIX uncertainty index)</i>						
intercept	0.0010*** (13.1535)	0.0010*** (12.9414)	0.0010*** (13.0686)	0.0180*** (10.8876)	0.0186*** (10.4846)	0.0184*** (11.1804)
granular residual η^g	-0.4761*** (-3.5420)	-0.1532 (-1.1413)		-0.5285** (-2.5480)	-0.1145 (-0.4657)	
network residual η^{nw}	1.0460*** (8.9871)		0.9665*** (9.4519)	1.5743*** (8.3138)		1.4330*** (6.7681)
T	8314	8314	8314	396	396	396
R-squared	0.5433	0.4965	0.5380	0.5322	0.4559	0.5235
<i>Panel C: Conditioning on yields (Fed funds rate, 3-month, 2-year, 10-year, 30-year treasury yields)</i>						
intercept	0.0008*** (7.3095)	0.0008*** (7.3640)	0.0008*** (7.7236)	0.0107*** (5.6651)	0.0107*** (5.5988)	0.0126*** (6.5865)
granular residual η^g	-0.5958*** (-4.0095)	-0.1261 (-0.8489)		-1.0147** (-2.5228)	-0.6463 (-1.3611)	
network residual η^{nw}	1.6160*** (8.1019)		1.5074*** (8.7126)	1.8765*** (5.4658)		1.7152*** (4.6175)
T	8047	8047	8047	496	496	496
R-squared	0.1051	0.0028	0.0964	0.1559	0.0224	0.1243
<i>Panel D: Conditioning on yield curve slope (2year-3month, 10year-2year, 30year-10year yield slopes)</i>						
intercept	0.0008*** (7.3408)	0.0008*** (7.4035)	0.0008*** (7.7314)	0.0133*** (7.6078)	0.0136*** (7.8880)	0.0149*** (8.9839)
granular residual η^g	-0.5895*** (-3.9163)	-0.1248 (-0.8318)		-0.9773** (-2.4333)	-0.6143 (-1.3019)	
network residual η^{nw}	1.6151*** (8.0934)		1.5090*** (8.7389)	1.8725*** (5.2486)		1.7185*** (4.4692)
T	8047	8047	8047	496	496	496
R-squared	0.1046	0.0018	0.0960	0.1534	0.0196	0.1238
<i>Panel E: Conditioning on asset pricing factors (Size, Value, Momentum)</i>						
intercept	0.0008*** (8.7029)	0.0008*** (8.7646)	0.0008*** (8.5887)	0.0154*** (8.7084)	0.0161*** (9.0380)	0.0162*** (8.9879)
granular residual η^g	-1.1955*** (-10.0454)	-0.7429*** (-6.3506)		-1.5302*** (-6.7145)	-1.2039*** (-4.7880)	
network residual η^{nw}	1.6490*** (10.8394)		1.4298*** (11.9143)	1.6030*** (6.0106)		1.1753*** (3.8499)
T	12359	12359	12359	588	588	588
R-squared	0.1763	0.0835	0.1437	0.2646	0.2099	0.1962
<i>Panel F: Conditioning on all of the above simultaneously</i>						
intercept	0.0004*** (4.2959)	0.0004*** (4.3919)	0.0004*** (4.6320)	0.0047*** (2.8386)	0.0038** (2.0246)	0.0049*** (2.9740)
granular residual η^g	-0.4357*** (-4.0462)	-0.0487 (-0.4540)		-0.2214 (-1.0028)	0.1874 (0.8060)	
network residual η^{nw}	0.9492*** (9.3119)		0.8585*** (9.8523)	1.2607*** (8.1254)		1.2106*** (7.7437)
T	6433	6433	6433	396	396	396
R-squared	0.6068	0.5725	0.6031	0.6810	0.6287	0.6799

Table 2.3: Estimates and t-statistics of OLS time series regressions of the form $r_t^{mkt} = b_0 + b_1\eta_t^g + b_2\eta_t^{nw} + u_t$ over all available observations between 2 January 1974 and 31 March 2022. t-statistics are calculated from Newey-West standard errors with $T^{1/4}$ lags. The dependent variable r_t^{mkt} is the return on the value-weighted CRSP index. Data sources for the control variables are [Yahoo! Finance](#) for Panel A, [FRED](#) for Panels B to D, and [French's data library](#) for Panel E. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

	daily			monthly		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: 2-stage least squares estimator</i>						
intercept	0.0007*** (7.0072)	0.0007*** (6.8415)	0.0007*** (7.1281)	0.0135*** (7.6980)	0.0137*** (7.3585)	0.0143*** (7.7892)
granular residual η^g	-0.5805*** (-7.5919)	-0.4049*** (-5.2967)		-1.3819*** (-5.8709)	-0.9450*** (-3.8767)	
network residual η^{nw}	0.6004*** (9.2836)		0.4789*** (7.5845)	2.1519*** (8.2639)		1.8086*** (6.7767)
T	12359	12359	12359	588	588	588
R-squared	0.0630	0.0022	0.0483	0.1632	0.0531	0.0727
<i>Panel B: Many instruments estimator</i>						
intercept	0.0007*** (6.9011)	0.0007*** (6.8302)	0.0007*** (6.9974)	0.1285 (1.0051)	0.1169 (0.9231)	0.1190 (0.7719)
granular residual η^g	-0.5084*** (-5.5405)	-0.4388*** (-4.8925)		1.6922 (0.9982)	1.0820 (0.8350)	
network residual η^{nw}	0.2294*** (2.9805)		0.1087 (1.4484)	-1.9277 (-0.9953)		-1.2600 (-0.7635)
T	12359	12359	12359	588	588	588
R-squared	0.0283	0.0020	0.0125	-6.4527	-4.9906	-5.2696
<i>Panel C: Announcement date return index</i>						
intercept	0.0007*** (7.1540)	0.0007*** (6.9073)	0.0007*** (7.2890)	0.0128*** (6.8257)	0.0127*** (6.4127)	0.0143*** (7.4887)
granular residual η^g	-0.5032 (-1.5467)	-0.0635 (-0.1950)		-2.5454*** (-2.9229)	-2.5754*** (-2.7810)	
network residual η^{nw}	1.4072*** (5.5593)		1.2635*** (5.2473)	1.2735 (1.1043)		0.0314 (0.0248)
T	12359	12359	12359	588	588	588
R-squared	0.1035	0.0009	0.0898	0.1138	0.0026	0.0026

Table 2.4: Estimates and t-statistics of instrumental variables time series regressions of the form $r_t^{mkt} = b_0 + b_1\eta_t^g + b_2\eta_t^{nw} + u_t$ over the period from 2 January 1974 to 31 March 2022. Residual index series η^g and η^{nw} are instrumented by CAPM residuals or announcement day returns. The dependent variable r^{mkt} is the return on the value-weighted CRSP index. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

2.5 Conclusions

In this paper, we explore the implications of direct information spillovers on the asset-level for the overall market aggregate. First, we find evidence of a strong and time-varying spillover network, and describe the impact of individual assets within the network. Second, we link the structure of the network to the aggregate market return and argue that innovations on the granular level propagate to the extent that they affect the overall market. We find such effects to significantly account for 5-10% of the market variation over our full sample from 1974 to 2022, and temporarily for more than 40% of the market variation in the aftermaths of the Great Financial Crisis. We conclude that the granular hypothesis has explanatory power in financial markets.

These insights highlight the need to better connect macro-finance to events on the corporate level, and thereby building unifying bridges between micro and macro economics. Our findings are relevant to modern day financial markets. For example, to prevent contagion of granular events to systemically relevant portions of the overall markets, financial supervision bodies need to be wary of forming network concentration. Similarly, a directed spillover network affects diversification and optimal portfolio choice for investors.

Our results have broader implications for the asset pricing literature. More precisely, the existence of granular effects in equity markets implies one of two things for the compensation for risks: Either, the market risk premium contains components that constitute compensation for granular risks, or, only exposure to aggregate shocks is compensated by a risk premium and beta estimates are biased by granular effects. While we find this question highly relevant, we leave the analysis to future research. Additionally, the existence of directed spillovers puts a new spin on portfolio optimization, where directed measures could replace symmetric correlations as a means to approximate out-of-sample mean-variance efficiency.

2.A Appendices

2.A.1 Data specifics

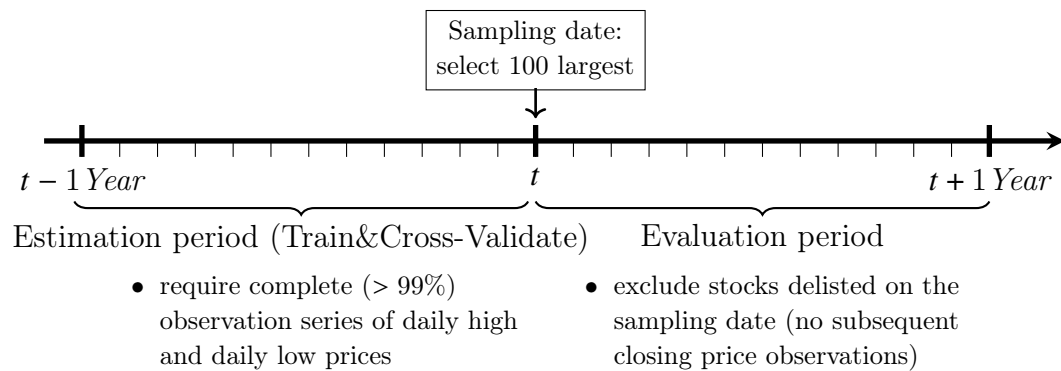


Figure 2.8: Illustration of the sampling logic. At the end of each month we select the 100 largest stocks by market capitalization. We require the selected sample to have complete observation series of daily high and low price data. Additionally, we exclude stocks that are delisted on the sampling date.

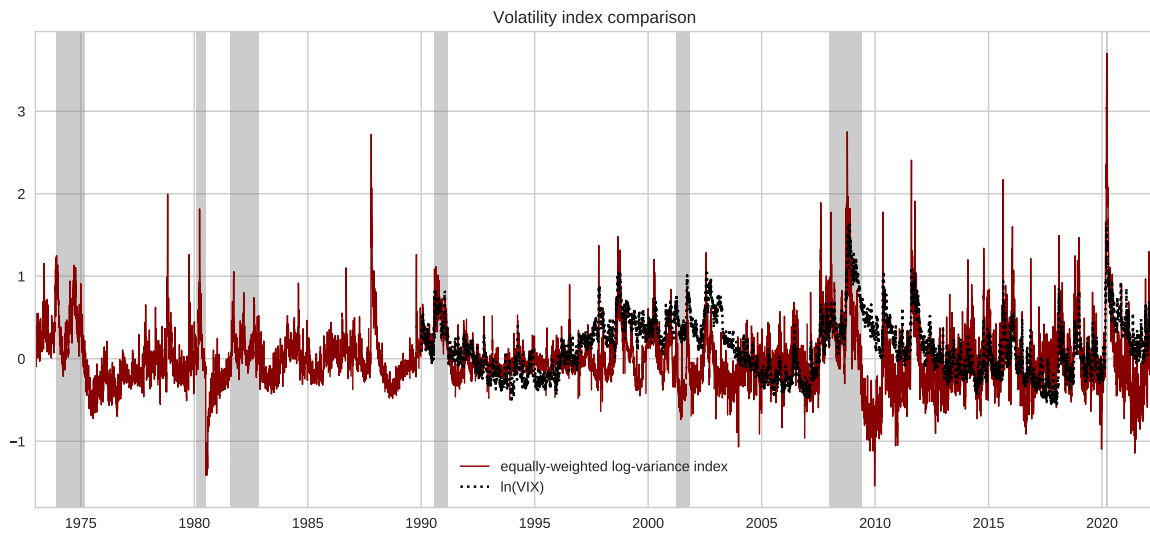


Figure 2.9: Evolution of the daily volatility over time and the logarithm of the VIX for comparison. The shaded intervals indicate NBER recessions.

2.A.2 Standardized VAR estimation

Consider the following VAR(1) model:

$$y_t = \mu + \Gamma y_{t-1} + \varepsilon_t \quad (2.20)$$

Row i in (2.20) has the following regression form:

$$y_{i,t} = \mu_i + \sum_{j=1}^N \gamma_{i,j} y_{j,t-1} + \varepsilon_{i,t} \quad (2.21)$$

After standardizing $y_{i,t}$ by mean $\bar{y}_i$ and standard deviation σ_{y_i} as $\tilde{y}_{i,t} = \frac{y_{i,t} - \bar{y}_i}{\sigma_{y_i}}$, we estimate a transformed model without intercepts:

$$\tilde{y}_{i,t} = \sum_{j=1}^N \tilde{\gamma}_{i,j} \tilde{y}_{j,t-1} + \tilde{\varepsilon}_{i,t} \quad (2.22)$$

We can then recover the VAR(1) coefficients on the original data in the following manner:

$$\frac{y_{i,t} - \bar{y}_i}{\sigma_{y_i}} = \sum_{j=1}^N \tilde{\gamma}_{i,j} \frac{y_{j,t-1} - \bar{y}_j}{\sigma_{y_j}} + \tilde{\varepsilon}_{i,t} \quad (2.23)$$

$$y_{i,t} = \bar{y}_i + \sum_{j=1}^N \tilde{\gamma}_{i,j} \frac{\sigma_{y_i}}{\sigma_{y_j}} (y_{j,t-1} - \bar{y}_j) + \sigma_{y_i} \tilde{\varepsilon}_{i,t} \quad (2.24)$$

$$y_{i,t} = \underbrace{\bar{y}_i - \sum_{j=1}^N \tilde{\gamma}_{i,j} \frac{\sigma_{y_i}}{\sigma_{y_j}} \bar{y}_j}_{\mu_i} + \underbrace{\sum_{j=1}^N \tilde{\gamma}_{i,j} \frac{\sigma_{y_i}}{\sigma_{y_j}} y_{j,t-1}}_{\gamma_{i,j}} + \underbrace{\sigma_{y_i} \tilde{\varepsilon}_{i,t}}_{\varepsilon_{i,t}} \quad (2.25)$$

Now let the VAR(1) equation additionally contain common exogenous factors as regressors:

$$y_{i,t} = \mu_i + \sum_{k=1}^K \beta_{i,k} f_{k,t} + \sum_{j=1}^N \gamma_{i,j} y_{j,t-1} + \varepsilon_{i,t} \quad (2.26)$$

Then we obtain from the standardized specification (now additionally with $\tilde{f}_{k,t} = \frac{f_{k,t} - \bar{f}_k}{\sigma_{f_k}}$):

$$y_{i,t} = \underbrace{\bar{y}_i - \sum_{k=1}^K \tilde{\beta}_{i,k} \frac{\sigma_{y_i}}{\sigma_{f_k}} \bar{f}_k}_{\mu_i} - \sum_{j=1}^N \tilde{\gamma}_{i,j} \frac{\sigma_{y_i}}{\sigma_{y_j}} \bar{y}_j + \underbrace{\sum_{k=1}^K \tilde{\beta}_{i,k} \frac{\sigma_{y_i}}{\sigma_{f_k}} f_{k,t}}_{\beta_{i,k}} + \underbrace{\sum_{j=1}^N \tilde{\gamma}_{i,j} \frac{\sigma_{y_i}}{\sigma_{y_j}} y_{j,t-1}}_{\gamma_{i,j}} + \underbrace{\sigma_{y_i} \tilde{\varepsilon}_{i,t}}_{\varepsilon_{i,t}} \quad (2.27)$$

2.A.3 Exemplary Network Estimation: 31 March 2022

In this appendix, we illustrate the estimation procedure as detailed in Section 2.2 by showcasing auxiliary plots and tables based on the latest available sample period from 1 April 2021 to 31 March 2022. We consider one calendar year worth of data preceding the sampling date, resulting in a window of 253 daily observations.

Sampling & Data

First, we calculate intraday realized volatilities $\tilde{\sigma}_{i,t}$ as in Section 2.2.1 for all assets in the CRSP universe and select 100 assets based on the criteria described in Section 2.2.1. We provide a list of the tickers of the selected assets in Table 2.5. Additionally, Table 2.6 contains an overview of the used sector abbreviations.

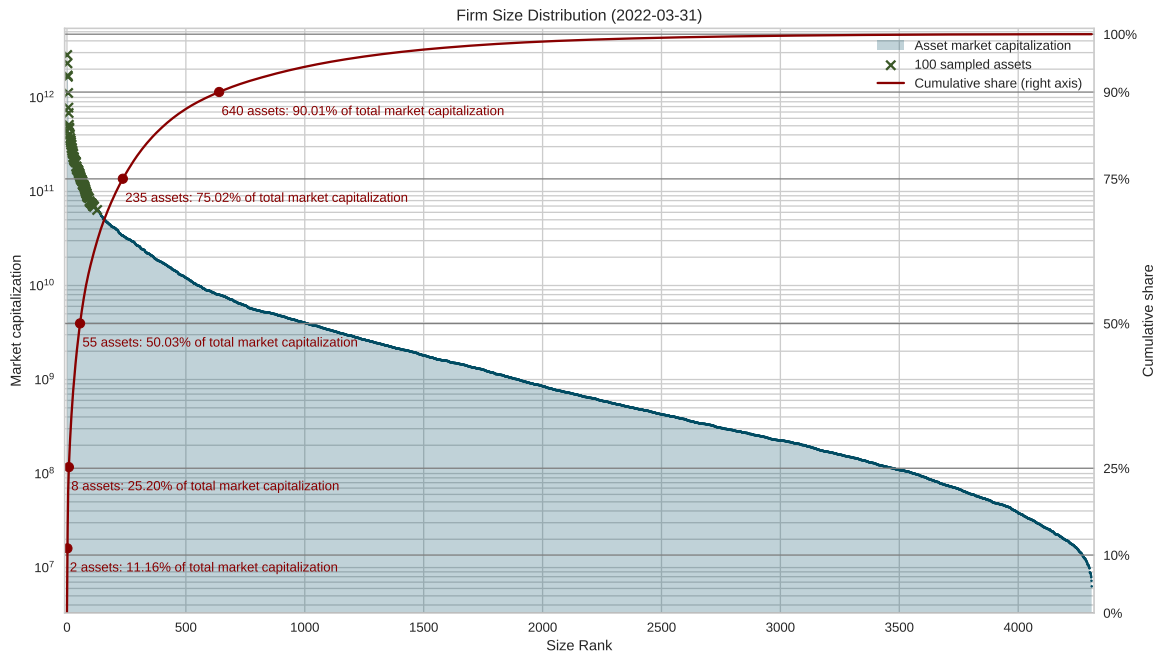


Figure 2.10: Distribution of firm size (market capitalization) in the CRSP universe at the end of March 2022.

Figure 2.10 displays the distribution of firm sizes in the CRSP universe and the position of the 100 assets selected into our sample. Notably, there is a strong concentration in market capitalization into few assets such that already 54 assets make up more than 50% percent of the total market value, despite there being over 4000 assets in the cross-section. The assets ordered by size almost form a straight line on the logarithmic scale, indicating a power-law distribution. Additionally, there seems to be an extra curve up for the largest assets and a curve down for the smallest, which leads to additional concentration. Given our criteria to select 100 assets into the sample, we

Ticker	Company Name	Ticker	Company Name
AAPL	Apple Inc	LRCX	Lam Resh Corp
ABBV	Abbvie Inc	MA	Mastercard Inc
ABT	Abbott Laboratories	MCD	Mcdonalds Corp
ADBE	Adobe Inc	MDLZ	Mondelez International Inc
ADP	Automatic Data Processing Inc	MMM	3M Co
AMAT	Applied Materials Inc	MO	Altria Group Inc
AMD	Advanced Micro Devices Inc	MRK	Merck & Co Inc New
AMGN	Amgen Inc	MRNA	Moderna Inc
AMZN	Amazon Com Inc	MS	Morgan Stanley Dean Witter & Co
ANTM	Anthem Inc	MSFT	Microsoft Corp
AVGO	Broadcom Inc	MU	Micron Technology Inc
AXP	American Express Co	NEE	Nextera Energy Inc
BA	Boeing Co	NFLX	Netflix Inc
BAC	Bank Of America Corp	NKE	Nike Inc
BKNG	Booking Holdings Inc	NOW	Servicenow Inc
BLK	Blackrock Inc	NVDA	Nvidia Corp
BMJ	Bristol Myers Squibb Co	ORCL	Oracle Corp
BRK	Berkshire Hathaway Inc Del	PEP	Pepsico Inc
BX	Blackstone Inc	PFE	Pfizer Inc
C	Citigroup Inc	PG	Procter & Gamble Co
CAT	Caterpillar Inc	PM	Philip Morris International Inc
CHTR	Charter Communications Inc New	PNC	P N C Financial Services Grp Inc
CMCSA	Comcast Corp New	PYPL	Paypal Holdings Inc
COP	Conocophillips	QCOM	Qualcomm Inc
COST	Costco Wholesale Corp New	RTX	Raytheon Technologies Corp
CRM	Salesforce Com Inc	SBUX	Starbucks Corp
CSCO	Cisco Systems Inc	SCHW	Schwab Charles Corp New
CVS	C V S Health Corp	SNOW	Snowflake Inc
CVX	Chevron Corp New	SPGI	S & P Global Inc
DE	Deere & Co	SQ	Block Inc
DHR	Danaher Corp	SYK	Stryker Corp
DIS	Disney Walt Co	T	A T & T Inc
FB	Meta Platforms Inc	TFC	Truist Financial Corp
GE	General Electric Co	TGT	Target Corp
GILD	Gilead Sciences Inc	TJX	T J X Companies Inc New
GM	General Motors Co	TMO	Thermo Fisher Scientific Inc
GOOGL	Alphabet Inc	TMUS	T Mobile U S Inc
GS	Goldman Sachs Group Inc	TSLA	Tesla Inc
HD	Home Depot Inc	TXN	Texas Instruments Inc
HON	Honeywell International Inc	UBER	Uber Technologies Inc
IBM	International Business Machs Cor	UNH	Unitedhealth Group Inc
INTC	Intel Corp	UNP	Union Pacific Corp
INTU	Intuit Inc	UPS	United Parcel Service Inc
ISRG	Intuitive Surgical Inc	USB	U S Bancorp Del
JNJ	Johnson & Johnson	V	Visa Inc
JPM	Jpmorgan Chase & Co	VZ	Verizon Communications Inc
KO	Coca Cola Co	WFC	Wells Fargo & Co New
LLY	Lilly Eli & Co	WMT	Walmart Inc
LMT	Lockheed Martin Corp	XOM	Exxon Mobil Corp
LOW	Lowe's Companies Inc	ZTS	Zoetis Inc

Table 2.5: Tickers and corresponding company names of the included stocks in the 31 March 2022 estimation.

end up with the assets marked by a green cross, including mostly the largest assets in the cross-section by market capitalization.

Figure 2.11 displays the distribution of the pooled daily intraday volatilities $\tilde{\sigma}_{i,t}$ of

Sector ticker	Sector
NoDur	Consumer Nondurables – Food, Tobacco, Textiles, Apparel, Leather, Toys
Durbl	Consumer Durables – Cars, TVs, Furniture, Household Appliances
Manuf	Manufacturing – Machinery, Trucks, Planes, Off Furn, Paper, Com Printing
Enrgy	Oil, Gas, and Coal Extraction and Products
Chems	Chemicals and Allied Products
BusEq	Business Equipment – Computers, Software, and Electronic Equipment
Telcm	Telephone and Television Transmission
Utils	Utilities
Shops	Wholesale, Retail, and Some Services (Laundries, Repair Shops)
Hlth	Healthcare, Medical Equipment, and Drugs
Money	Finance
Other	Other – Mines, Constr, BldMt, Trans, Hotels, Bus Serv, Entertainment

Table 2.6: Overview of the sector abbreviations used in this paper. The classification corresponds to the 12-sector industry mapping of Fama and French (1997).

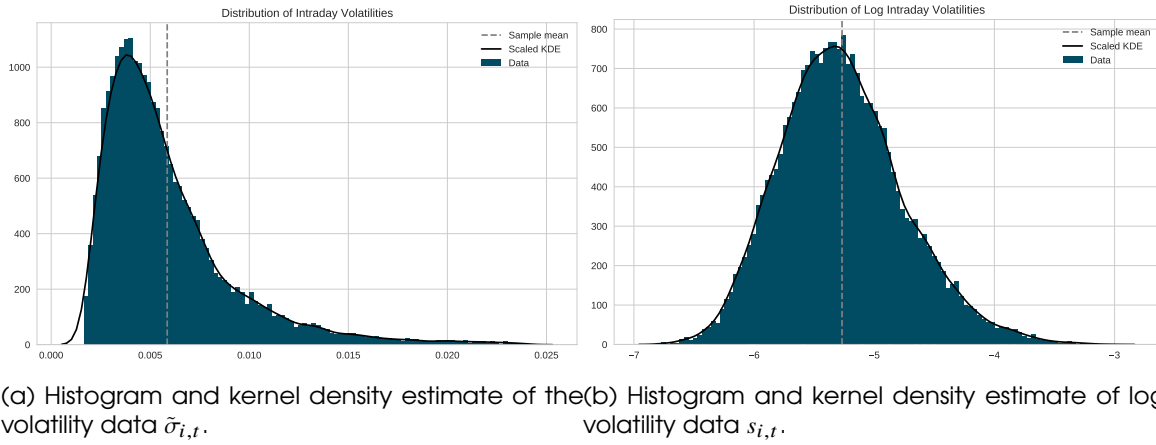


Figure 2.11: Histograms and Gaussian kernel density estimates of the distribution of the intraday volatility data before and after the log transformation. Note that the distributions are mixtures of the distributions of the individual series.

the 100 assets before and after the log-transformation, resulting in an approximately normal distribution of the variable $s_{i,t} = \ln(\tilde{\sigma}_{i,t})$.

Checking the correlations and first-order auto-correlations of this panel of intraday volatility data, we observe a strong comovement structure in the heatmaps in Figure 2.12. Elements in the correlation matrix are standard Pearson correlation coefficients. Analogously, we define elements in the auto-correlation matrix as

$$\phi_{i,j}^{AR1} = \frac{\sum_{t=2}^T (s_{i,t} - \bar{s}_i)(s_{j,t-1} - \bar{s}_j)}{\sum_{t=2}^T |s_{i,t} - \bar{s}_i| \cdot \sum_{t=2}^T |s_{j,t-1} - \bar{s}_j|}.$$

VAR estimation

We model valuation volatility $s_{i,t}$ as a VARX(1) according to $s_{i,t} = \mu_i + \beta_i f_t + \sum_{j=1}^N \gamma_{i,j} s_{j,t-1} + u_{i,t}$ and apply Adaptive Elastic Net estimation. Figure 2.13 illustrates the search for the hyperparameters λ and κ . We define a logarithmic grid over suitable values and select the best value pair through cross-valuation. Panel 2.13a

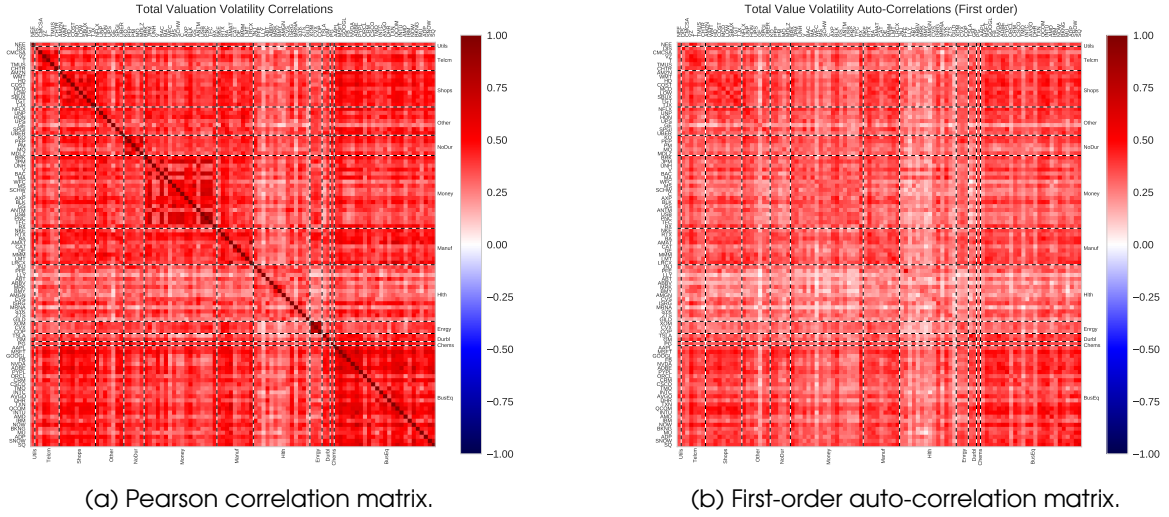
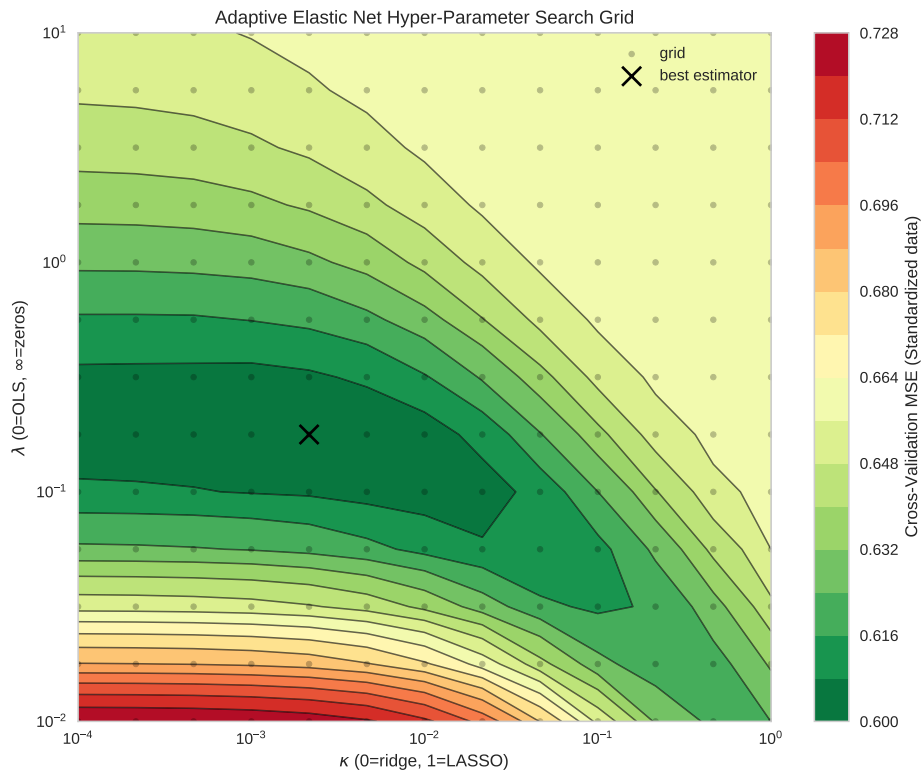


Figure 2.12: Contemporaneous correlation and auto-correlation heatmaps of intraday volatility data $s_{i,t}$. Series are grouped by sector and ordered by average market capitalization over the sample period within sectors.

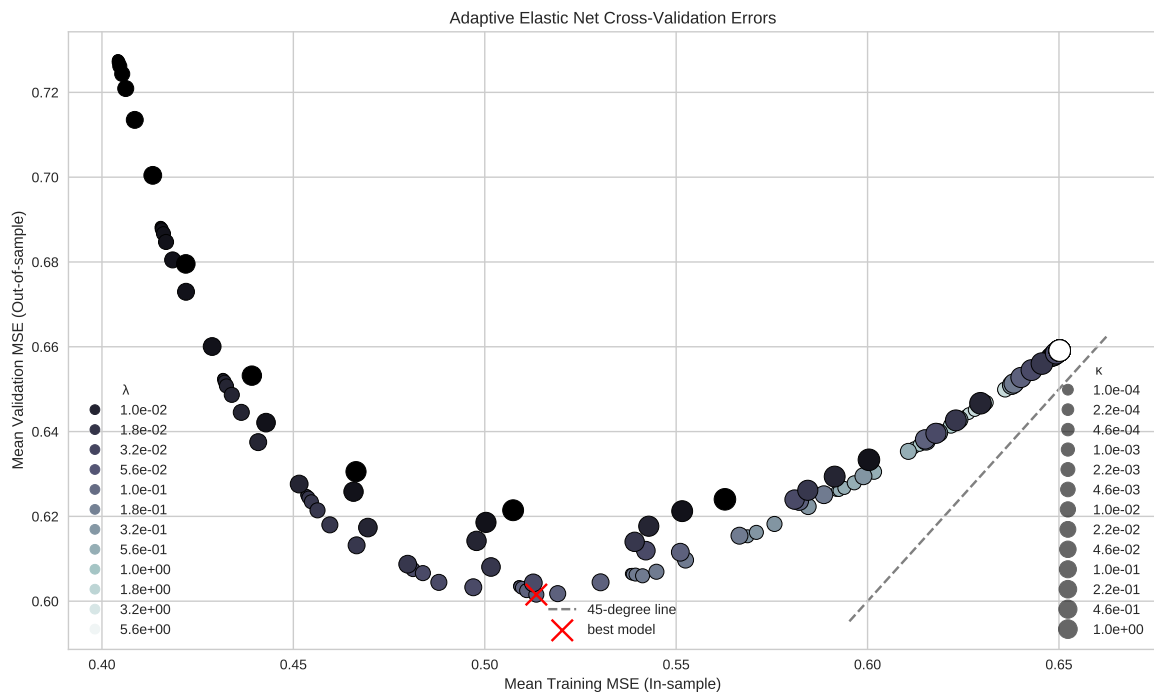
displays the hyperparameter grid and cross-validation MSE loss used in the Adaptive Elastic Net estimation of the VAR(1) regression. The best estimator is the one that minimizes the cross-validation MSE. Note that several special cases are nested for specific hyperparameter values - λ equal to zero results in the OLS estimate of the model, a κ equal to zero corresponds to ridge regression, and a κ equal to one gives the LASSO estimator. Panel 2.13b then plots the cross-validation mean training MSE against the mean validation MSE. Note how the validation set performance deteriorates as the level of regularization approaches the unregularized OLS estimates.

Figure 2.14 contrasts heatmaps of the VARX(1) coefficients to illustrate how regularized estimation changes the estimates. On the left, Panel 2.14a displays the OLS estimate of the coefficient matrix. Panel 2.14b then shows the same figure for the optimized Adaptive Elastic Net estimator. While conventional estimation results in what appears mostly noise, regularized estimation recovers a pattern with a pronounced diagonal and a sparse representation of inter-asset links.

Next, we visualize how lagged spillovers in the regularized VAR(1) coefficient matrix capture auto-cross-correlations after accounting for common factors and spillovers in Figures 2.15 and 2.16. Panel 2.15a displays the distribution of the data after removing common factors, and Panels 2.16a and 2.16b show the corresponding correlation and auto-correlation matrices. Comparing with the plots in Figures 2.11 and 2.12 we see that, despite having factored out common variation as systematic components f_t , there is still a substantial part of the correlation structure left to be captured. However, after additionally accounting for spillovers, there is only a meaningful portion of correlation left in the contemporaneous matrix in Panel 2.16c, whereas the remaining



(a) VARX(1) regression hyperparameter grid and cross-validation MSE loss of the Adaptive Elastic Net estimation.



(b) Mean training MSE against the mean cross-validation MSE. Dots correspond to the grid in Panel 2.13a with the size and colour indicating the respective combination of the κ and λ hyperparameters. The dashed 45-degree line indicates models that do equally well in-sample and out-of-sample.

Figure 2.13: Grid search for the Adaptive Elastic Net hyperparameters in the VARX(1) estimation.

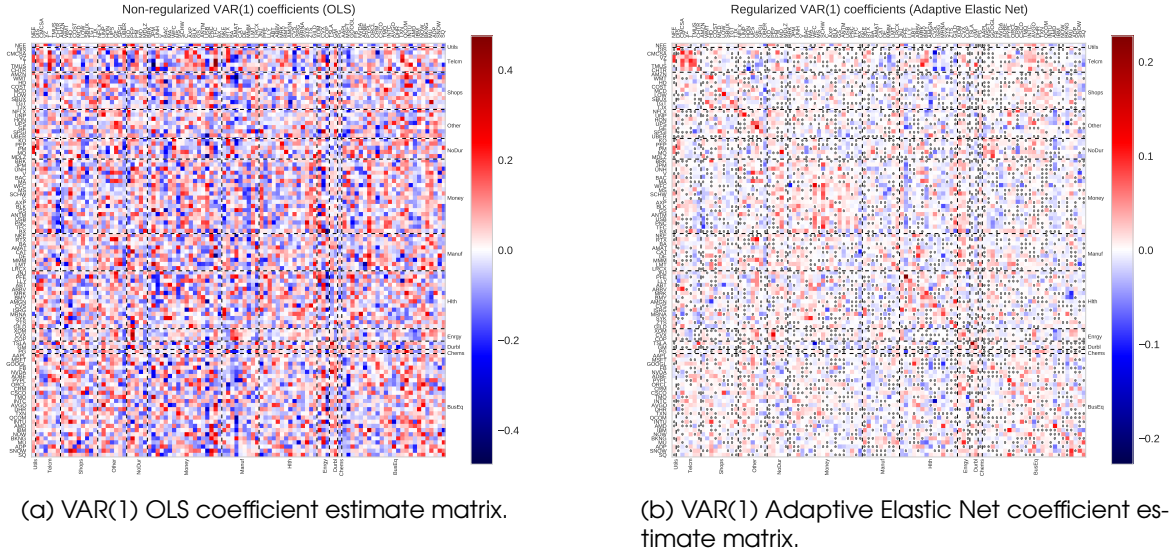


Figure 2.14: Comparison of heatmap plots of the VAR coefficient estimates via OLS and Adaptive Elastic Net. Exact zero values are marked by a gray zero.

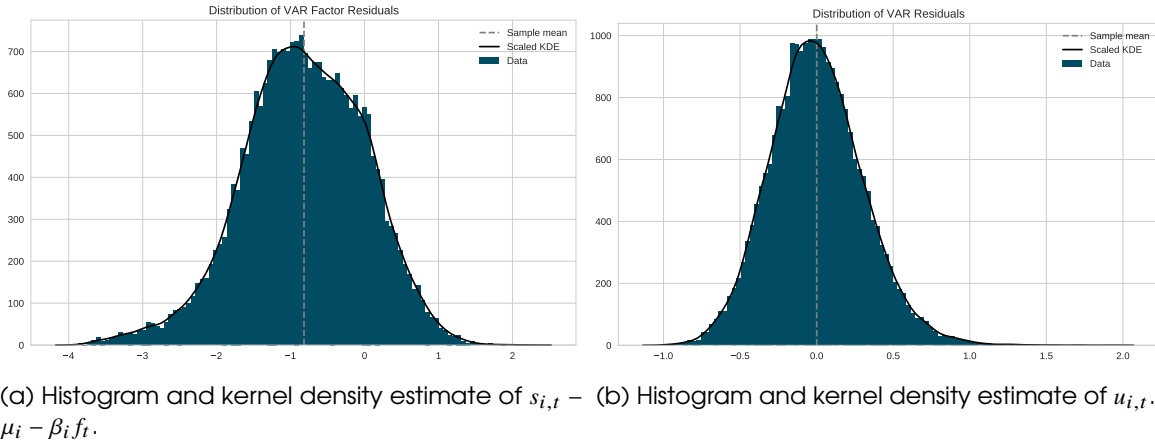


Figure 2.15: Histograms and Gaussian kernel density estimates of VAR residuals of the sample drawn on 31-03-2022. We model valuation volatility $s_{i,t}$ according to (2.4): $s_{i,t} = \mu_i + \beta_i f_t + \sum_{j=1}^N \gamma_{i,j} s_{j,t-1} + u_{i,t}$.

auto-correlations have mostly faded in Panel 2.16d and the error distribution in Panel 2.15 closely approximates a Gaussian distribution.

Finally, Figure 2.17 displays heatmaps of the VMA matrices at increasing lags. Expectedly, the magnitude of the MA coefficients decreases at increasing horizon. Also visible is the magnifying effect of larger horizons, resulting in the column-wise concentration of coefficients over time, indicating the importance of some series in the panel.

Covariance matrix estimation

As a second input to the network construction, we estimate the high-dimensional residual covariance matrix through regularized GLASSO estimation. Analogous to the

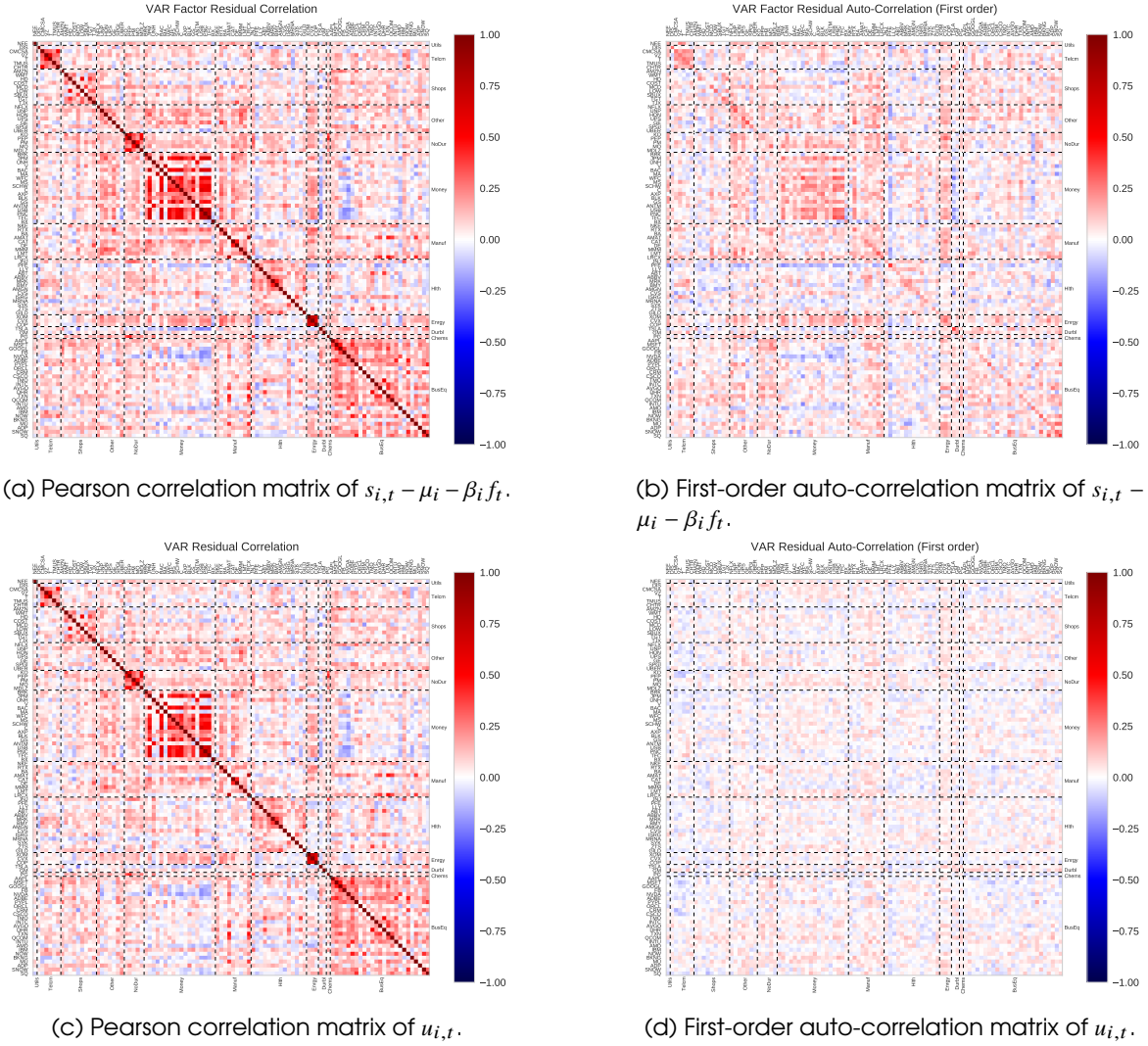


Figure 2.16: Contemporaneous correlation and auto-correlation heatmaps of VAR residuals before and after subtraction of spillovers.

regularized estimation of the regression coefficients, we apply a grid search approach to cross-validate the ρ hyperparameter of the GLASSO estimator. Figure 2.18 illustrates the hyperparameter search. We choose the estimator such that it minimizes the cross-validation loss. Note that as ρ goes to zero, the estimate approaches the sample covariance estimate. Likewise, if ρ goes to infinity we obtain a diagonal covariance matrix estimate with sample variances on the diagonal and zeros at the covariance terms.

Examining the resulting covariance estimates in Figure 2.19, a large share of the values has been visibly shrunk. That is, the classical sample estimate of the covariance matrix overfits outlier observations that do not generalize well within the observation window. Although the Graphical LASSO does not force entries in the covariance matrix to zero, the resulting covariance matrix in Panel 2.19b appears more focused when

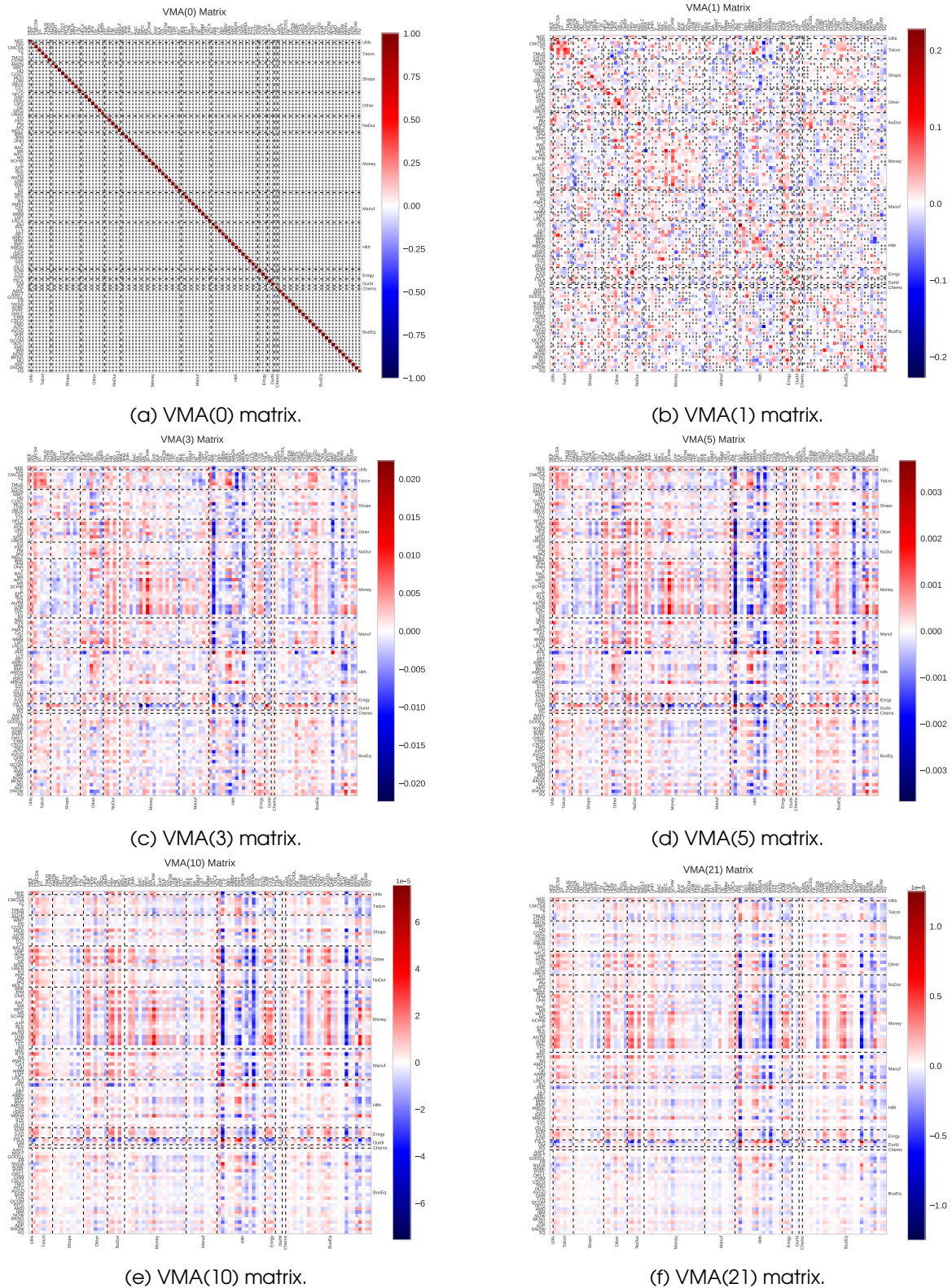


Figure 2.17: Heatmaps of VMA matrices. Values are obtained from inversion of the estimated VAR(1) process.

compared to the sample covariance matrix in Panel 2.19a. More clearly, a comparison

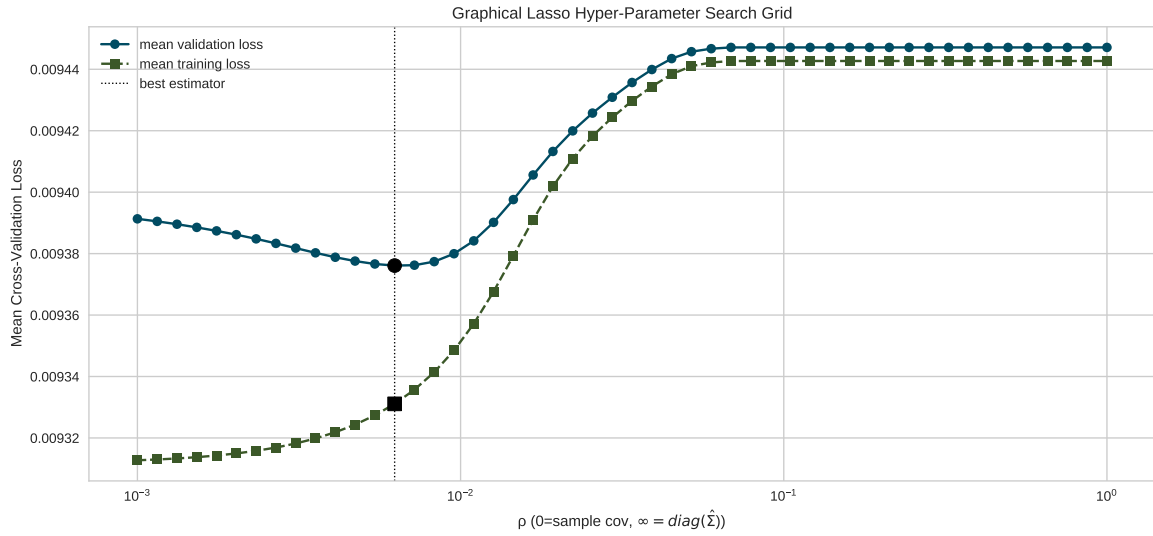


Figure 2.18: Hyperparameter search grid and cross-validation loss used in the Graphical LASSO estimation of the VAR(1) residual covariance matrix.

of the corresponding partial correlation matrices in Panels 2.19c and 2.19d underlines how the applied regularization induces conditional independence, and hence sparsity on many edges in the contemporary network graph. We define off-diagonal elements in this matrix by $\hat{\varphi}_{i,j} = -\frac{\hat{\theta}_{i,j}}{\sqrt{\hat{\theta}_{i,i}\hat{\theta}_{j,j}}}$, where $\hat{\theta}_{i,j}$ are elements in $\hat{\Theta}$. The illustration suggests that ordinary estimation fails to distinguish persistent relations between variables from purely statistical artifacts, and is therefore unable to uncover the underlying structure.

FEVD Network Construction

We then use the VMA matrices as in Figure 2.17 and the residual covariance matrix from Figure 2.19b to construct multivariate generalized impulse response matrices at increasing horizons. We display heatmaps of such matrices in Figure 2.20.

Finally, aggregating the impulse response matrices up to horizon $H = 21$ days we obtain a matrix of forecast error variances (FEV). With the help of the force-directed algorithm, we obtain a first network representation of this matrix in Figure 2.21. That is, links in this network represent the flow of forecasting uncertainty with regards to return volatility of individual tickers. Lastly, applying a row-wise decomposition, we obtain the FEVD network as displayed in Section 2.3.1.

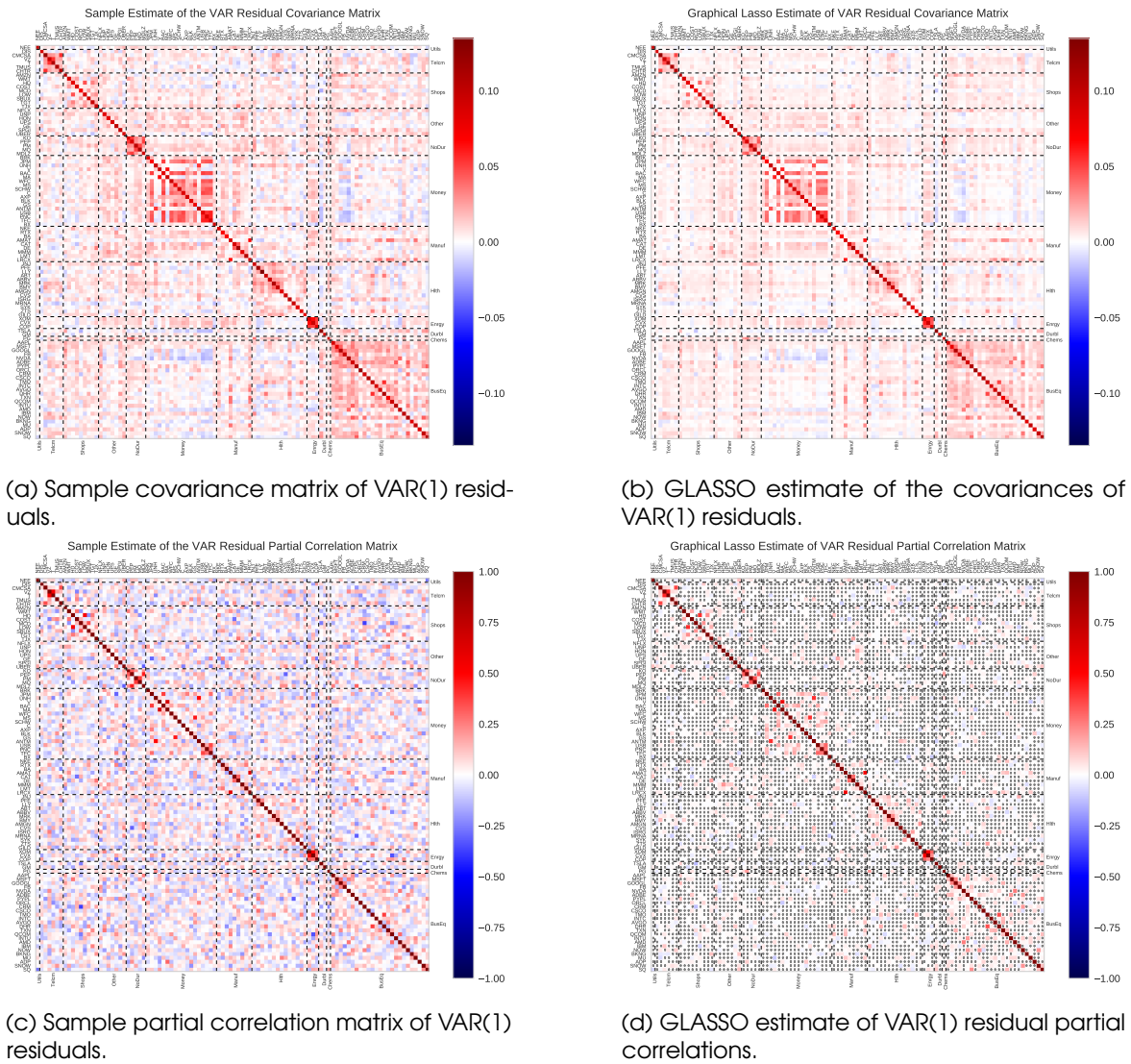
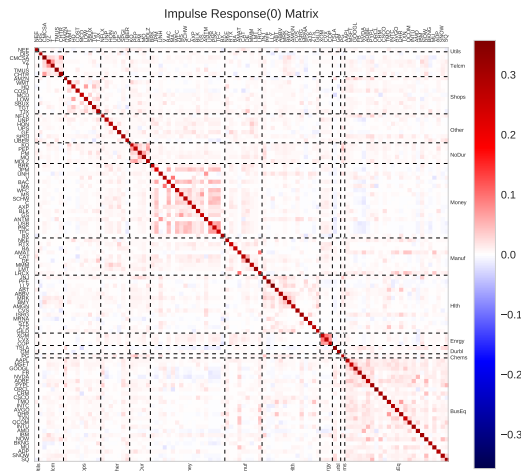
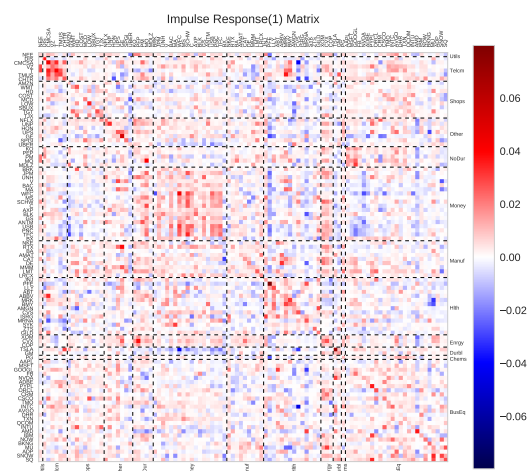


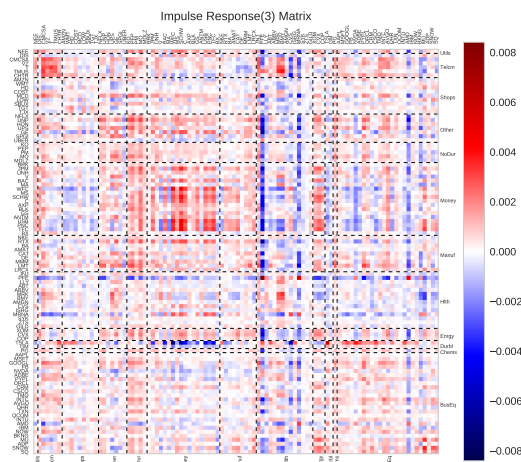
Figure 2.19: Heatmap plots of the sample covariance matrix of the VAR(1) residuals and its cross-validated Graphical LASSO estimate. Exact zero values are marked by a gray zero. Additionally, we show heatmap plots of the corresponding partial correlation matrices.



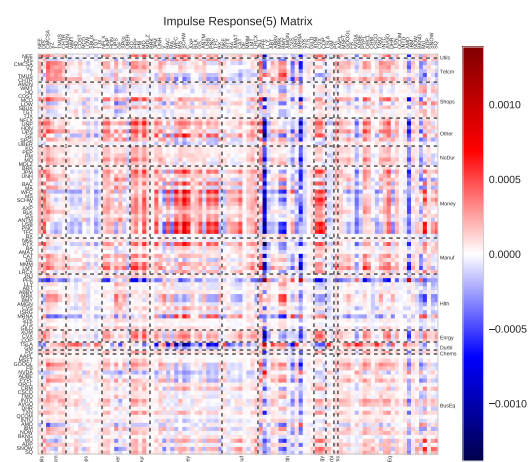
(a) Impulse response(0) matrix.



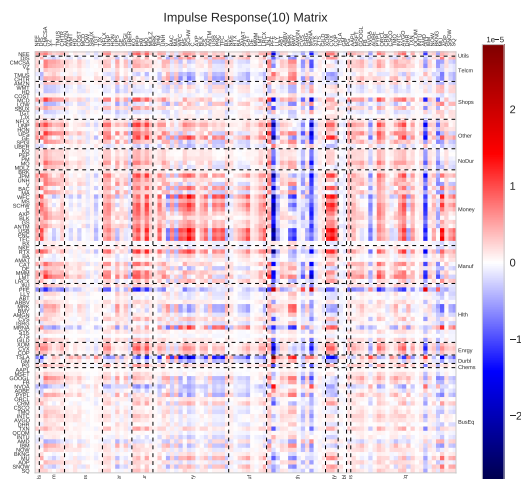
(b) Impulse response(1) matrix.



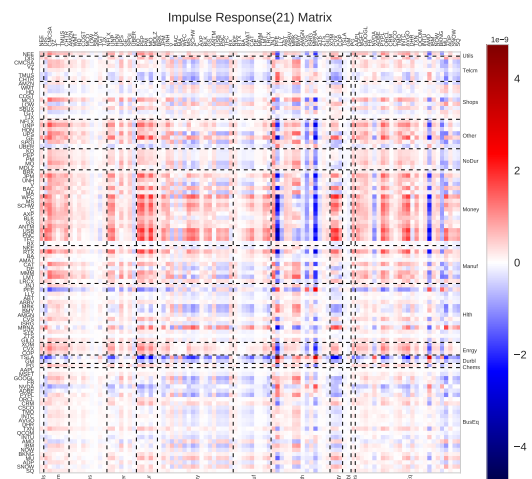
(c) Impulse response(3) matrix.



(d) Impulse response(5) matrix.



(e) Impulse response(10) matrix.



(f) Impulse response(21) matrix.

Figure 2.20: Heatmaps of the multivariate generalized impulse response matrices.

FEV Network (2022-03-31)

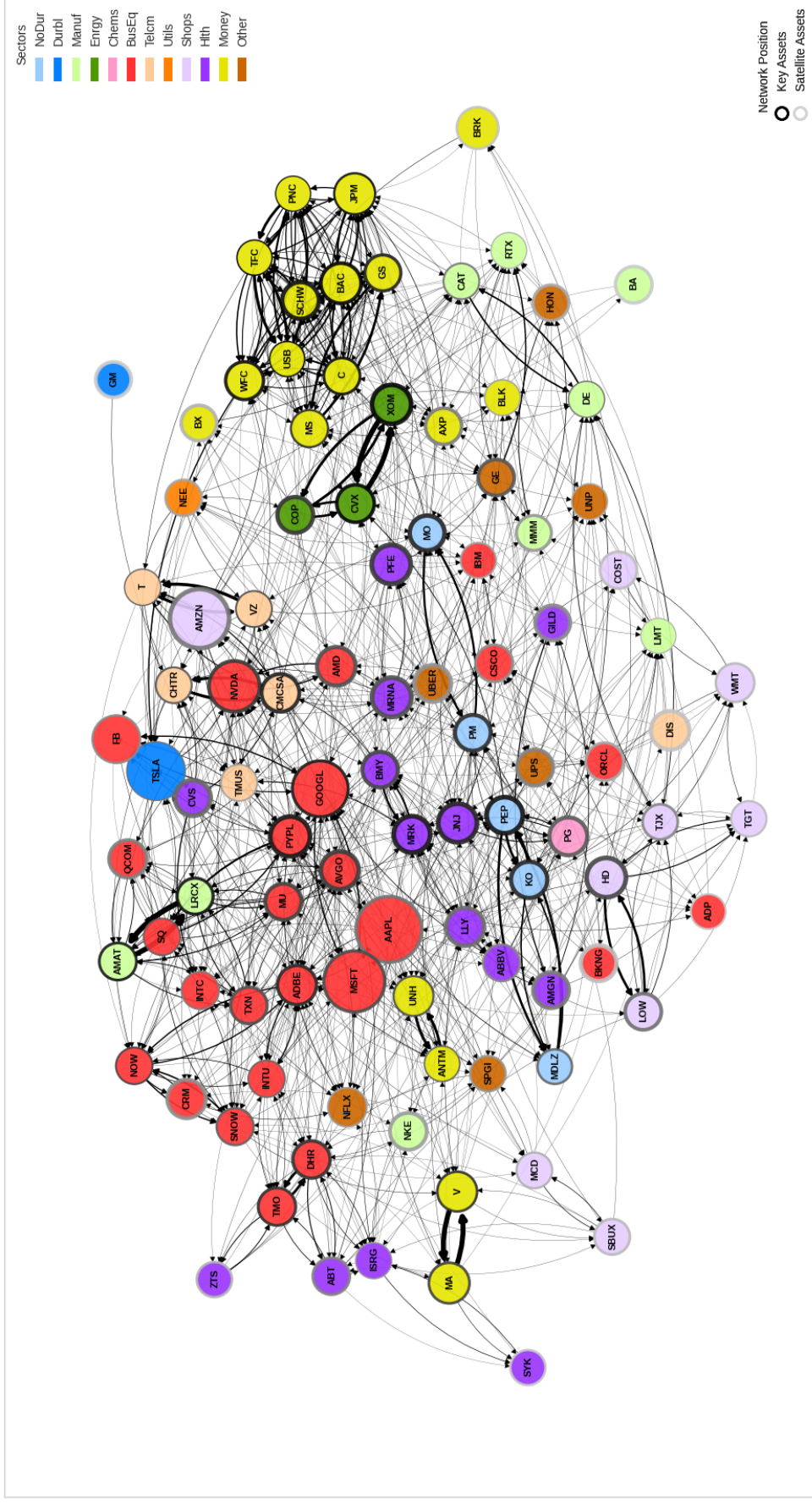


Figure 2.21: Network graph of the forecast error variance adjacency matrix. The visualization is created with the help of the Fruchterman-Reingold force-directed algorithm (see Fruchterman and Reingold (1991)). Tickers are colored according to their SIC codes in the Fama/French 12-sector industry classification. Edge thickness is determined by the strength of the respective directed link. The nodes with the largest net connectedness (key assets) are circled in black, and the nodes with the smallest net connectedness (satellites) are circled in light grey. The thickness of the circle boundaries is determined by the net value of *out*-connectedness minus *in*-connectedness. A list of ticker-company name pairs is provided in Appendix 2.A.3

2.A.4 Rolling window statistics

Estimation statistics

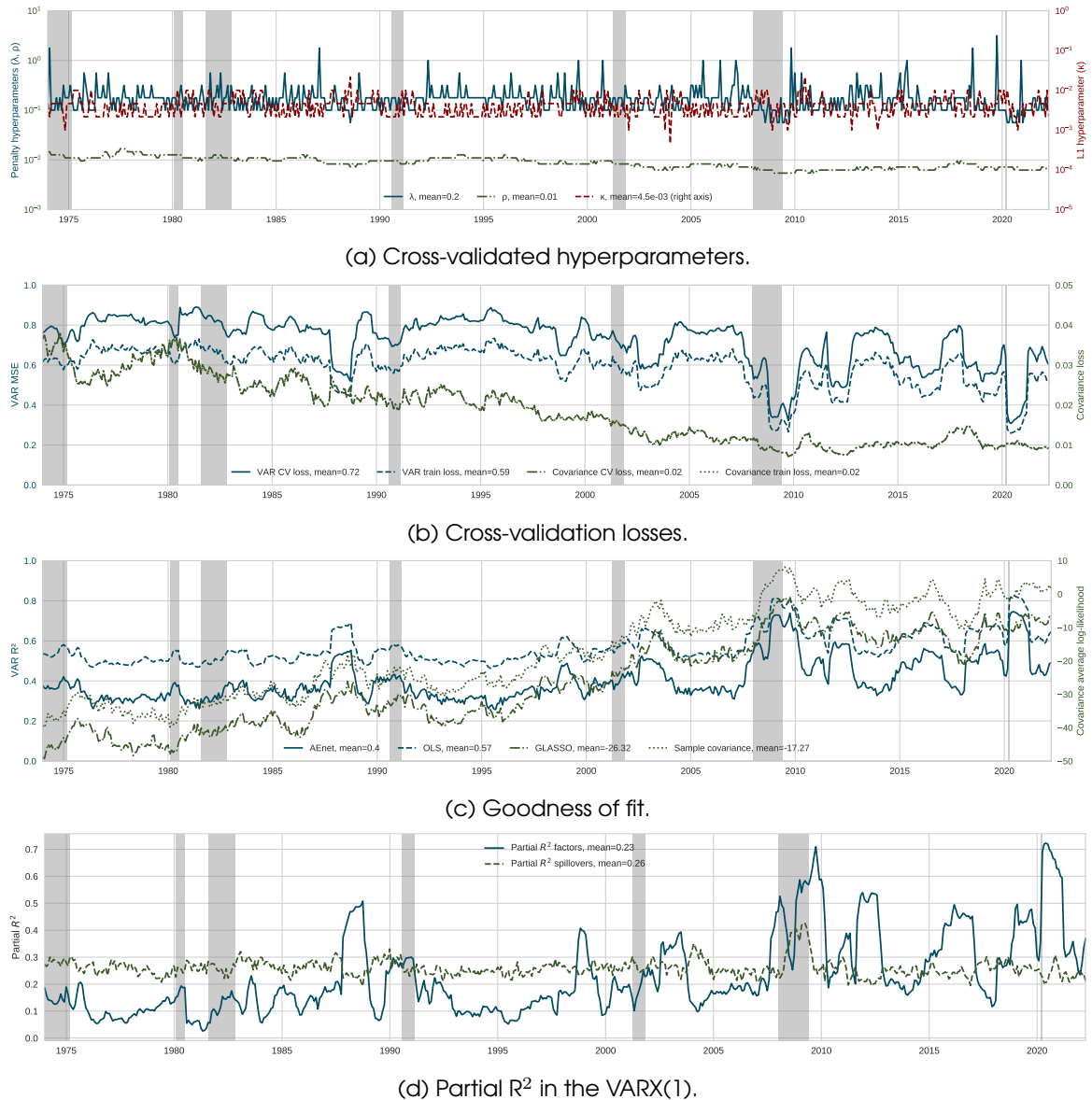


Figure 2.22: Summary statistics of the rolling estimation. Estimates are obtained from regularized VARX(1) and covariance matrix estimation over rolling periods of one calendar year. The shaded intervals indicate NBER recessions.

Figure 2.22 monitors the evolution of several statistics related to the rolling estimation of FEV networks.

First, we observe the evolution of the chosen hyperparameters over the estimation windows in Panel 2.22a. Recall that these hyperparameters are chosen to optimize the out-of-sample performance in a CV setting. Both the AENET hyperparameters λ and κ , and the GLASSO hyperparameter ρ are fairly stable throughout the sample period.

Second, we look at the optimized losses for the estimators in each cross-validation in Panel 2.22b. For the standardized data in the VAR estimation, the losses have a direct interpretation as the share of unexplained variation. Unsurprisingly, the model performs considerably better in-sample than out-of-sample. Nevertheless, the stability of the explained variation gives us confidence in the procedure. Contrarily for the covariance estimation, due to their scaling, the losses in the training samples versus the validation samples are almost indistinguishable.

Third, the goodness of fit for the first-stage VAR(1) estimation reaches a reasonable level of approximately 30% to 70% in Panel 2.22c. Clearly, the goodness-of-fit statistics for both estimation stages trail their unregularized equivalents. Recall that the goal is not the best in-sample fit, but rather we are trading off a lower in-sample variation (bias) in favour of a better out-of-sample prediction (variance). Hence, these levels are speaking to the ability of the regularized estimators to counteract the overfitting of conventional methods.

And fourth, Panel 2.22d compares the partial R^2 values for the factor and the spillover coefficients of the VARX(1). While spillovers have an almost constant level of explanatory power beyond the common factor, there is considerable variation in the reverse case. Here the common volatility factor appears to particularly add explanatory value towards the end of the considered period, with some pronounced periods for example post-Great Recession and post-COVID.

Regularization statistics

Next, we look at the outcomes of regularization in the rolling estimation. Figure 2.23 contains three panels with rolling statistics that describe the impact of regularization on the estimation and the estimates.

First, we break down the degrees of freedom used during the estimation steps in Panel 2.23a. The total area (red) shows the number of data points available for estimation. Nonregularized estimation techniques bind more than half of the degrees of freedom (illustrated by the dashed lines). The variable selection techniques reduce this number distinctively (illustrated by the filled areas in blue and green). In all periods, the applied techniques fall back on the usage of fewer degrees of freedom than nonregularised estimators. Therefore, the confidence in the estimates improves considerably for our empirical analysis.

Second, we examine the sparsity level as an outcome of variable selection in Panel 2.23b, that is, the share of zeros in the estimated matrices. Here, we define $sparsity = \frac{|\{\theta_i: \theta_i=0\}_i^K|}{|\{\theta_i\}_i^K|}$, where $\{\theta_i\}_i^K$ is the set of estimated parameters in the respective estimation step, and $|\cdot|$ is a set's cardinality. On average, the estimator sets about a

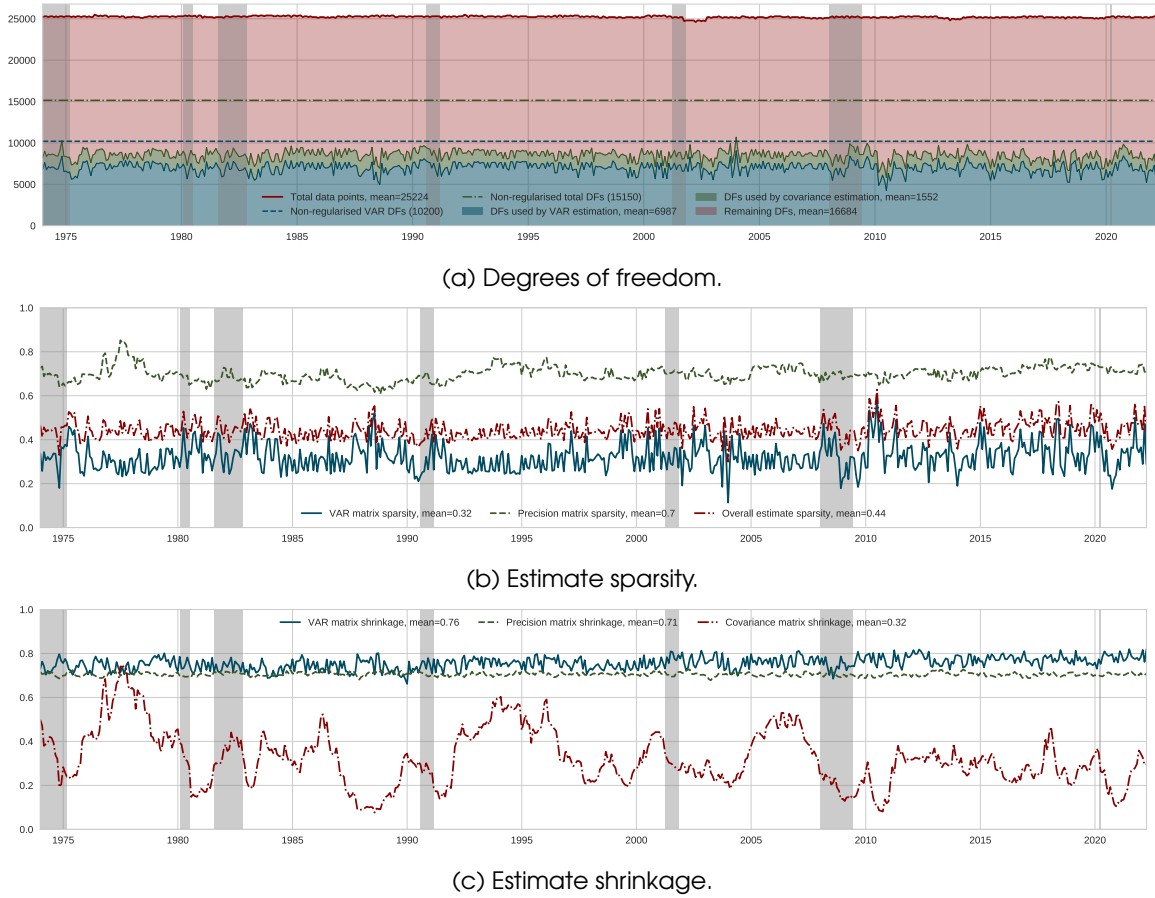


Figure 2.23: Summary statistics of the rolling regularization. Estimates are obtained from regularised VARX(1) and covariance matrix estimation over rolling periods of one calendar year. The shaded intervals indicate NBER recessions.

third of the VAR matrix entries to zero, indicating the lack of lagged spillovers between the respective assets. Similarly, the precision matrix estimated by the GLASSO selects only about a third of the connections to have weights greater than zero. Note however, that this does not imply any sparsity on its inverse, the covariance matrix. We also report the mean value of the two densities weighted by their nonregularised number of parameters.

And third, in Panel 2.23c we analyze the shrinkage factors applied to the nonzero estimates of the regularized estimators in a similar way. For that purpose, let $shrinkage = 1 - \frac{\sum_{i \in \mathcal{N}} |\hat{\theta}_i^{shrinkage}|}{\sum_{i \in \mathcal{N}} |\hat{\theta}_i^{standard}|}$, where $\mathcal{N} = \{i; \hat{\theta}_i^{shrinkage} \neq 0\}$ is the set of all indices i corresponding to nonzero shrinkage parameter estimates. We report the shrinkage factors for the VAR(1) matrix obtained via AENET estimation relative to the conventional OLS estimates. Then, we apply the same metric to the GLASSO estimate of the precision matrix relative to its sample equivalent. The shrinkage of the regression coefficients and the precision matrix entries are fairly stable, while the resulting shrinkage factor of the covariance matrix exhibits a more volatile dynamic behavior.

2.A.5 Auxiliary analysis

Individual asset weights

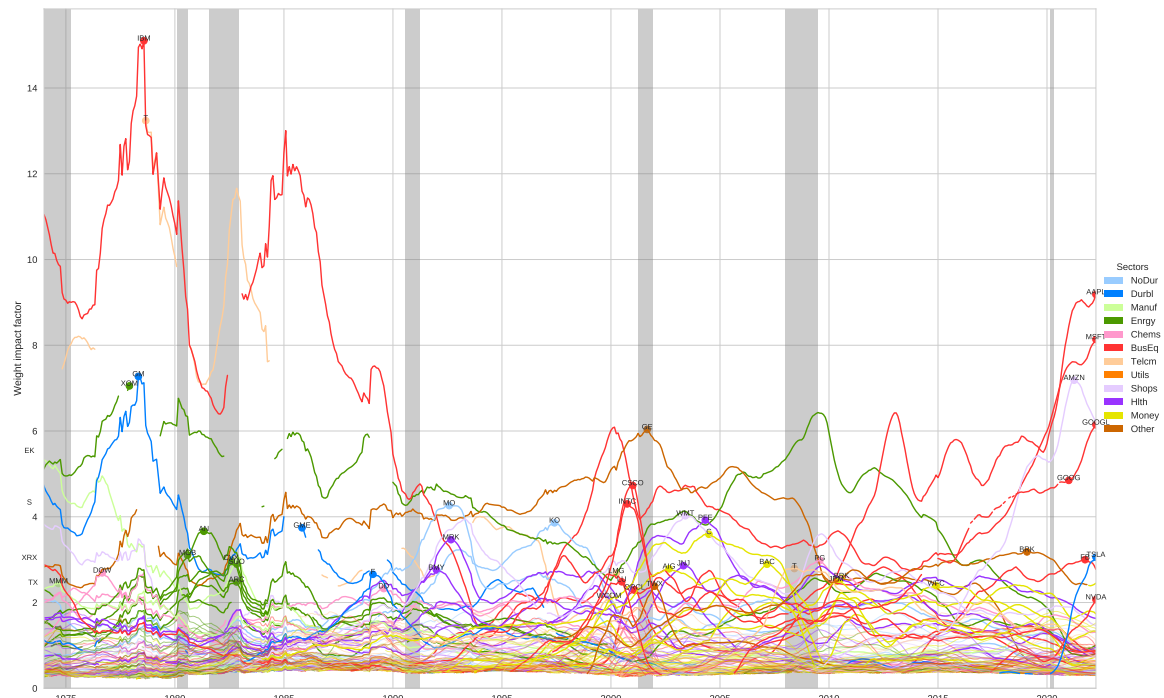


Figure 2.24: Rolling size weight impact of individual assets. Weight impact factor measures the actual market weight in terms of market capitalization relative to a uniform weight distribution, that is, where all assets receive an weight impact factor equal to one. Hence, a score of two implies an impulse to the respective asset has twice the impact compared to the perfectly diversified case. The shaded intervals indicate NBER recessions.

Granular network events

Date	Ticker	Full company name	Sector	Peak	Date	Ticker	Full company name	Sector	Peak
1974-03	FG	International Flavors & Frag Inc	Money	2.92	1996-05	FNM	Federal National Mortgage Assn	Money	2.94
1974-04	AVP	Avon Products Inc	Chems	4.22	1997-06	KMB	Kimberly Clark Corp	Manuf	3.91
1974-07	BDK	Black & Decker Mfg Co	Manuf	4.60	1997-07	WMX	Waste Management Inc New	Other	3.12
1974-11	KO	Coca Cola Co	NoDur	2.75	1997-08	BA	Boeing Co	Manuf	3.03
1975-03	GLW	Schlitz Jos Brewing Co	BusEq	3.29	1997-10	CPB	Campbell Soup Co	NoDur	3.18
	SQB	Squibb Corp	Hlth	4.27	1999-10	MOB	Mobil Corp	Enrgy	3.15
1975-04	STY	Sterling Drug Inc	Hlth	2.77	2000-08	AMAT	Applied Materials Inc	Manuf	2.71
1975-06	ROH	Rohn & Haas Co	Chems	2.74	2001-02	YHOO	Yahoo Inc	BusEq	2.81
1976-01	IP	International Paper Co	Manuf	2.69	2001-03	PCS	Sprint Corp	Telecm	4.69
	MCD	McDonalds Corp	Shops	3.07	2001-04	GLW	Corning Inc	BusEq	4.03
1976-03	CBM	Chesebrough Ponds Inc	Chems	2.51		JDSU	J D S Uniphase Corp	BusEq	3.09
1976-05	DEC	Digital Equipment Corp	BusEq	3.46	2001-05	SUNW	Sun Microsystems Inc	BusEq	2.94
1976-06	DIS	Disney Walt Co	Telecm	5.24	2001-06	EMC	E M C Corp Ma	BusEq	3.79
1976-09	ENE	Enron Corp	Enrgy	5.83	2001-09	HON	Honeywell International Inc	Other	3.76
1976-12	APD	Air Products & Chemicals Inc	Chems	2.60	2001-12	QCOM	Qualcomm Inc	BusEq	2.69
1977-01	MER	Merrill Lynch & Co Inc	Money	2.55	2002-01	DELL	Dell Inc	BusEq	3.61
1977-04	RTX	Raytheon Technologies Corp	Manuf	3.98	2002-02	WFC	Wells Fargo & Co New	Money	4.80
1977-06	RTN	Raytheon Co	BusEq	2.79	2003-11	BLS	Bellsouth Corp	Telecm	3.26
1977-09	CIL	Continental Ill Corp	Money	2.67		EBAY	Ebay Inc	BusEq	5.11
1977-12	HAL	Halliburton Co	Enrgy	3.61	2003-12	CMCSA	Comcast Corp New	Telecm	2.85
1978-01	HPQ	H P Inc	BusEq	2.81		T	A T & T Inc	Telecm	3.72
1978-03	PHA	Pharmacia Corp	Hlth	3.14		TWX	Time Warner Inc New	Other	4.38
1978-04	SY	Sperry Corp	BusEq	3.29	2004-02	FBF	Fleetboston Financial Corp	Money	2.85
1978-05	REV	Revlon Inc	Chems	3.51	2004-06	AWE	A T & T Wireless Svcs Inc	Telecm	8.09
1978-08	ABT	Abbott Laboratories	Hlth	3.83	2004-10	MMC	Marsh & McLennan Cos Inc	Money	3.07
1978-09	APD	Southern Railway Co	Chems	3.07	2005-08	COP	Conocophillips	Enrgy	4.01
	BKO	Baker International Corp	Manuf	3.12	2005-10	KRB	M B N A Corp	Money	2.81
	MDE	Mcdermott J Ray & Co Inc	Manuf	4.31		OXY	Occidental Petroleum Corp	Enrgy	4.56
1978-11	PGL	American Broadcasting Cos Inc	Utils	3.33	2006-09	DVN	Devon Energy Corp New	Enrgy	2.63
1979-02	DD	Du Pont E I De Nemours & Co	Chems	3.66	2006-12	VLO	Valero Energy Corp New	Enrgy	2.93
	EK	Eastman Kodak Co	Manuf	4.38	2007-11	CVX	Chevron Corp New	Enrgy	3.76
1979-03	CS	Cities Service Co	Enrgy	2.65	2008-07	BAC	Bank Of America Corp	Money	9.97
1979-05	TDY	Teledyne Inc	Manuf	4.52	2008-08	WB	Wachovia Corp 2Nd New	Money	6.46
1979-07	NCR	N C R Corp	BusEq	3.52	2008-09	AIG	American International Group Inc	Money	4.76
1979-08	TWX	Time Warner Inc	Other	2.58		BUD	Anheuser Busch Cos Inc	NoDur	7.55
1980-01	HON	Honeywell Inc	BusEq	3.88		ELV	Elevance Health Inc	Money	3.12
1980-05	GO	Gulf Corp	Enrgy	3.34		JPM	Jpmorgan Chase & Co	Money	4.51
1980-11	HBO	Hudsons Bay Oil & Gas Ltd	Enrgy	3.38	2008-10	APA	Apache Corp	Enrgy	2.51
1981-04	SUO	Shell Oil Co	Enrgy	3.02	2009-01	BRK	Berkshire Hathaway Inc Del	Other	4.66
1981-08	TXN	Texas Instruments Inc	BusEq	2.66	2009-02	AFL	Aflac Inc	Money	3.11
1981-10	AMX	Amx Inc	Manuf	2.86		DNA	Genentech Inc	Hlth	3.35
1982-04	AN	Amoco Corp	Enrgy	3.29		USB	U S Bancorp Del	Money	5.39
	UCL	Unocal Corp	Enrgy	3.31	2009-03	GE	General Electric Co	Other	8.93
1982-09	JPM	Morgan J P & Co Inc	Money	2.56	2009-04	DOW	Dow Chemical Co	Chems	2.59
1982-10	CMB	Chase Manhattan Corp	Money	3.82		WYE	Wyeth	Hlth	6.16
1982-11	PZL	Pennzoil Co	Enrgy	3.12	2009-12	BNI	Burlington Northern Santa Fe Cp	Other	4.95
	TIC	Travelers Corp	Money	2.91	2010-03	CSCO	Cisco Systems Inc	BusEq	2.73
1983-03	JCP	Penney J C Co Inc	Shops	4.14	2010-05	XTO	X T O Energy Inc	Enrgy	5.35
1983-04	GET	Getty Oil Co	Enrgy	4.77	2010-06	BK	Bank Of New York Mellon Corp	Money	3.28
	PEP	Pepsico Inc	NoDur	3.66	2010-07	TRV	Travelers Companies Inc	Money	2.69
	PG	Procter & Gamble Co	Chems	3.51	2010-08	CAT	Caterpillar Inc	Manuf	3.09
1983-05	MCIC	M C I Communications Corp	Telecm	6.87	2010-12	UNH	Unitedhealth Group Inc	Money	3.67
1983-11	GTE	G T E Corp	Telecm	2.98	2011-02	C	Citigroup Inc	Money	5.49
1984-03	MO	Altria Group Inc	NoDur	3.72	2011-05	AXP	American Express Co	Money	2.58
1984-10	MOT	Motorola Inc	BusEq	3.06	2011-10	FCX	Freeport Mcmoran Copper & Gold	Other	2.51
1984-11	AHS	American Hospital Supply Corp	Hlth	3.20	2012-10	YUM	Yum Brands Inc	Shops	2.57
1984-12	F	Ford Motor Co Del	Durbl	3.48	2013-12	MS	Morgan Stanley Dean Witter & Co	Money	3.63
	GM	General Motors Corp	Durbl	3.94	2014-02	GS	Goldman Sachs Group Inc	Money	2.55
1985-08	UIS	Unisys Corp	BusEq	3.51	2014-06	PNC	P N C Financial Services Grp Inc	Money	2.69
1986-04	HCA	Hospital Corp America	Hlth	2.94	2015-02	DTV	Directv	Telecm	3.68
	RCA	R C A Corp	BusEq	3.54	2015-09	PM	Philip Morris International Inc	NoDur	2.70
1987-01	WLA	Warner Lambert Co	Hlth	2.79	2015-10	KMI	Kinder Morgan Inc	Utils	2.55
1987-05	SRD	Standard Oil Co Oh	Enrgy	6.38	2017-05	RAI	Reynolds American Inc	NoDur	7.80
1987-07	TX	Texaco Inc	Enrgy	3.39	2017-07	MON	Monsanto Co New	Chems	3.15
1987-08	PPG	P P G Industries Inc	Chems	3.25	2018-01	META	Meta Platforms Inc	BusEq	3.93
1987-09	NSC	Norfolk Southern Corp	Other	2.78	2018-10	FOXA	Twenty First Century Fox Inc	NoDur	5.57
1988-04	INTC	Intel Corp	BusEq	3.16		XOM	Exxon Mobil Corp	Enrgy	2.71
1988-07	IBM	International Business Machs Cor	BusEq	3.10	2019-08	V	Visa Inc	Money	2.62
1989-03	RJR	R J R Nabisco Inc	NoDur	3.73	2020-03	MA	Mastercard Inc	Money	2.97
1989-04	MMM	3M Co	Manuf	2.77	2020-04	CRM	Salesforce Inc	BusEq	2.62
1989-08	CNA	C N A Financial Corp	Money	2.67	2020-09	MSFT	Microsoft Corp	BusEq	5.63
1990-06	XRX	Xerox Corp	BusEq	3.65		TSLA	Tesla Inc	Durbl	3.31
1990-11	ARC	Atlantic Richfield Co	Enrgy	2.92	2020-10	AAPL	Apple Inc	BusEq	12.99
1991-12	AIT	Ameritech Corp	Telecm	3.00		AMZN	Amazon Com Inc	Shops	8.90
1992-01	MRO	Marathon Oil Corp	Enrgy	4.12	2020-11	NVDA	Nvidia Corp	BusEq	3.36
1992-04	GCI	Food Lion Inc	Telecm	2.52	2020-12	AMD	Advanced Micro Devices Inc	BusEq	3.01
1993-08	SLE	Sara Lee Corp	NoDur	3.78		GOOG	Alphabet Inc	BusEq	2.70
1993-09	PFE	Pfizer Inc	Hlth	3.41	2021-01	NFLX	Netflix Inc	Other	4.85
1993-12	SGP	Schering Plough Corp	Hlth	2.59	2021-04	ADBE	Adobe Inc	BusEq	4.21
1994-02	BMJ	Bristol Myers Squibb Co	Hlth	2.88	2022-05	GOOGL	Alphabet Inc	BusEq	3.71
1994-10	CCB	Capital Cities Abc Inc	Telecm	2.80					

Table 2.7: List of peaks that surpass an effect size of 2.5 in Figure 2.5.

2.B Bibliography

- Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). “The network origins of aggregate fluctuations.” In: *Econometrica* 80.5, pp. 1977–2016.
- Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015a). *Networks, shocks, and systemic risk*. Tech. rep. National Bureau of Economic Research.
- Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015b). “Systemic risk and stability in financial networks.” In: *American Economic Review* 105.2, pp. 564–608.
- Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2017). “Microeconomic origins of macroeconomic tail risks.” In: *American Economic Review* 107.1, pp. 54–108.
- Acharya, V. V., L. H. Pedersen, T. Philippon, and M. Richardson (2017). “Measuring systemic risk.” In: *The review of financial studies* 30.1, pp. 2–47.
- Adrian, T. and M. K. Brunnermeier (2016). “CoVaR.” In: *The American Economic Review* 106.7, p. 1705.
- Ahern, K. R. (2013). “Network centrality and the cross section of stock returns.” In: *Available at SSRN 2197370*.
- Ahern, K. R. (2017). “Information networks: Evidence from illegal insider trading tips.” In: *Journal of Financial Economics* 125.1, pp. 26–47.
- Ahern, K. R. and J. Harford (2014). “The importance of industry links in merger waves.” In: *The Journal of Finance* 69.2, pp. 527–576.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). “Modeling and forecasting realized volatility.” In: *Econometrica* 71.2, pp. 579–625.
- Andrei, D., H. Friedman, and N. B. Ozel (2023). “Economic uncertainty and investor attention.” In: *Journal of Financial Economics* 149.2, pp. 179–217.
- Balakumar, B., T. Hastie, J. Friedman, and N. Tibshirani Simon (2016). *Glmnet for Python*. http://www.stanford.edu/~hastie/glmnet_python/.

- Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006). “Who’s who in networks. Wanted: The key player.” In: *Econometrica* 74.5, pp. 1403–1417.
- Banerjee, O., L. E. Ghaoui, and A. d’Aspremont (2008). “Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data.” In: *Journal of Machine learning research* 9.Mar, pp. 485–516.
- Barigozzi, M. and C. Brownlees (2019). “Nets: Network estimation for time series.” In: *Journal of Applied Econometrics* 34.3, pp. 347–364.
- Baruník, J. and T. Křehlík (2018). “Measuring the frequency dynamics of financial connectedness and systemic risk.” In: *Journal of Financial Econometrics* 16.2, pp. 271–296.
- Battiston, S. et al. (2012). “Debtrank: Too central to fail? financial networks, the fed and systemic risk.” In: *Scientific reports* 2.1, pp. 1–6.
- Bianchi, D., M. Billio, R. Casarin, and M. Guidolin (2015). “Modeling contagion and systemic risk.” In: *Available at SSRN*.
- Billio, M., M. Getmansky, A. W. Lo, and L. Pelizzon (2012). “Econometric measures of connectedness and systemic risk in the finance and insurance sectors.” In: *Journal of financial economics* 104.3, pp. 535–559.
- Blank, S., C. M. Buch, and K. Neugebauer (2009). “Shocks at large banks and banking sector distress: The Banking Granular Residual.” In: *Journal of Financial Stability* 5.4, pp. 353–373.
- Bonacich, P. (1987). “Power and centrality: A family of measures.” In: *American journal of sociology* 92.5, pp. 1170–1182.
- Brownlees, C. and R. F. Engle (2017). “SRISK: A conditional capital shortfall measure of systemic risk.” In: *The Review of Financial Studies* 30.1, pp. 48–79.
- Brunner, F. and R. Hipp (2023). “Estimating large-dimensional connectedness tables: The Great Moderation through the lens of sectoral spillovers.” In: *Quantitative Economics* 14.3, pp. 1021–1058.
- Campbell, J. Y. and J. Ammer (1993). “What moves the stock and bond markets? A variance decomposition for long-term asset returns.” In: *The journal of finance* 48.1, pp. 3–37.

- Carvalho, V. M. (2014). "From micro to macro via production networks." In: *Journal of Economic Perspectives* 28.4, pp. 23–48.
- Chuluun, T., A. Prevost, and A. Upadhyay (2017). "Firm network structure and innovation." In: *Journal of Corporate Finance* 44, pp. 193–214.
- Demirer, M., F. X. Diebold, L. Liu, and K. Yilmaz (2018). "Estimating global bank network connectedness." In: *Journal of Applied Econometrics* 33.1, pp. 1–15.
- Diebold, F. X. and K. Yilmaz (2009). "Measuring financial asset return and volatility spillovers, with application to global equity markets." In: *The Economic Journal* 119.534, pp. 158–171.
- Diebold, F. X. and K. Yilmaz (2012). "Better to give than to receive: Predictive directional measurement of volatility spillovers." In: *International Journal of forecasting* 28.1, pp. 57–66.
- Diebold, F. X. and K. Yilmaz (2014). "On the network topology of variance decompositions: Measuring the connectedness of financial firms." In: *Journal of Econometrics* 182.1, pp. 119–134.
- Fama, E. F. and K. R. French (1997). "Industry costs of equity." In: *Journal of financial economics* 43.2, pp. 153–193.
- Foerster, A. T., P.-D. G. Sarte, and M. W. Watson (2011). "Sectoral versus aggregate shocks: A structural factor analysis of industrial production." In: *Journal of Political Economy* 119.1, pp. 1–38.
- Friedman, J., T. Hastie, and R. Tibshirani (2010). "Regularization paths for generalized linear models via coordinate descent." In: *Journal of statistical software* 33.1, p. 1.
- Friedman, J., T. Hastie, and R. Tibshirani (2008). "Sparse inverse covariance estimation with the graphical lasso." In: *Biostatistics* 9.3, pp. 432–441.
- Fruchterman, T. M. and E. M. Reingold (1991). "Graph drawing by force-directed placement." In: *Software: Practice and experience* 21.11, pp. 1129–1164.
- Gabaix, X. (2011). "The granular origins of aggregate fluctuations." In: *Econometrica* 79.3, pp. 733–772.
- Gabaix, X. (2016). "Power laws in economics: An introduction." In: *Journal of Economic Perspectives* 30.1, pp. 185–206.

- Gabaix, X. and R. S. Koijen (2020). *Granular instrumental variables*. Tech. rep. National Bureau of Economic Research.
- Gabaix, X. and R. S. Koijen (2021). *In search of the origins of financial fluctuations: The inelastic markets hypothesis*. Tech. rep. National Bureau of Economic Research.
- Galaasen, S., R. Jamilov, R. Juelsrud, and H. Rey (2020). *Granular credit risk*. Tech. rep. National Bureau of Economic Research.
- Hansen, C., J. Hausman, and W. Newey (2008). “Estimation with many instrumental variables.” In: *Journal of Business & Economic Statistics* 26.4, pp. 398–422.
- Hasler, M. and C. Ornathanalai (2018). “Fluctuating attention and financial contagion.” In: *Journal of Monetary Economics* 99, pp. 106–123.
- Herskovic, B. (2018). “Networks in production: Asset pricing implications.” In: *The Journal of Finance* 73.4, pp. 1785–1818.
- Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh (2016). “The common factor in idiosyncratic volatility: Quantitative asset pricing implications.” In: *Journal of Financial Economics* 119.2, pp. 249–283.
- Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh (2020). “Firm volatility in granular networks.” In: *Journal of Political Economy* 128.11, pp. 4097–4162.
- Herskovic, B., T. Kind, and H. Kung (2023). “Micro uncertainty and asset prices.” In: *Journal of Financial Economics* 149.1, pp. 27–51.
- Malevergne, Y., P. Santa-Clara, and D. Sornette (2009). *Professor zipf goes to wall street*. Tech. rep. National Bureau of Economic Research.
- Manski, C. F. (1993). “Identification of endogenous social effects: The reflection problem.” In: *The Review of Economic Studies* 60.3, pp. 531–542.
- Nelson, D. B. (1991). “Conditional heteroskedasticity in asset returns: A new approach.” In: *Econometrica: Journal of the Econometric Society*, pp. 347–370.
- Newman, M. (2018). *Networks*. Second. Oxford university press.
- Ozsoylev, H. N. and J. Walden (2011). “Asset pricing in large information networks.” In: *Journal of Economic Theory* 146.6, pp. 2252–2280.
- Page, L., S. Brin, R. Motwani, and T. Winograd (1999). *The PageRank citation ranking: Bringing order to the web*. Tech. rep. Stanford InfoLab.

- Parkinson, M. (1980). "The extreme value method for estimating the variance of the rate of return." In: *Journal of business*, pp. 61–65.
- Pesaran, H. H. and Y. Shin (1998). "Generalized impulse response analysis in linear multivariate models." In: *Economics letters* 58.1, pp. 17–29.
- Petajisto, A. (2009). "Why do demand curves for stocks slope down?" In: *Journal of Financial and Quantitative Analysis* 44.5, pp. 1013–1044.
- Rapach, D. E., J. K. Strauss, and G. Zhou (2013). "International stock return predictability: what is the role of the United States?" In: *The Journal of Finance* 68.4, pp. 1633–1662.
- Savor, P. and M. Wilson (2016). "Earnings announcements and systematic risk." In: *The Journal of Finance* 71.1, pp. 83–138.
- Shiller, R. J. (1980). *Do stock prices move too much to be justified by subsequent changes in dividends?*
- Tambe, P. and L. M. Hitt (2014). "Job hopping, information technology spillovers, and productivity growth." In: *Management science* 60.2, pp. 338–355.
- Welch, I. and A. Goyal (2008). "A comprehensive look at the empirical performance of equity premium prediction." In: *The Review of Financial Studies* 21.4, pp. 1455–1508.
- Zenou, Y. (2016). "Key players." In: *Oxford Handbook on the Economics of Networks*, pp. 244–274.
- Zou, H. and H. H. Zhang (2009). "On the adaptive elastic-net with a diverging number of parameters." In: *The Annals of Statistics* 37.4, p. 1733.

Chapter 3

Granular risk premia

Abstract

In a network connectedness model of US equity prices, granular innovations at the firm level can have systemic effects for two reasons: Index weights are too concentrated for diversification arguments to apply, and spillovers propagate non-systematic innovations to a large portion of investor wealth. In this study, I introduce a pricing kernel that includes structural connections and show theoretically that inter-asset connectedness has implications for expected returns. I empirically investigate various conjectures to validate this framework and its quantitative relevance by using estimated spillover networks of 100 US equities. My results imply that risk premia for firm-level innovations exist temporarily and relate to mean returns in the cross-section of stocks. I collect evidence that such granular risk premia could be an overlooked determinant of asset prices, consistent with the continuous emergence of anomalies relative to reduced-form factor pricing models.

3.1 Introduction

In 2022, just 10 stocks made up more than 25% of the value of the entire US equity market. Given this level of concentration, investors face considerable exposure to the nonsystematic risk of companies with large weights in their portfolios. Yet, limits to diversification stand in contrast to the paradigm of the empirical asset pricing literature to approximate the complexity of asset markets through linear factor models which neglect the importance of firm-specific events. In addition, residuals of empirical factor models often exhibit correlations, contradicting the theoretical assumption of independent idiosyncratic innovations. Therefore, the assumptions underlying factor models disallow systemically important risks to spread from granular sources. Instead, I relax these assumptions and let compensated idiosyncratic risk affect equilibrium asset prices through concentration and inter-asset linkages.

The goal of this paper is to understand if cross-asset dependencies promote the formation of return premia for nonsystematic risk, which I term ‘granular risk premia’. In my framework, granular innovations can have systemic effects due to a combination of two reasons: First, index weights are too concentrated for diversification arguments to apply, and second, structural spillovers propagate nonsystematic innovations to a large portion of investor wealth. For example, a firm-level shock that hits the asset with the largest index weight may show equal systemic impact as a shock to a small firm that spreads to the valuations of all firms in its sector. As a result, granular innovations influence equilibrium asset prices in the Capital Asset Pricing Model (CAPM) and command their own risk premia, which I attempt to seek out empirically. Hence, my analysis revisits asset pricing in light of the ‘granular hypothesis’ of Gabaix (2011), where the compositions of equity markets may be such that micro-economic fluctuations can lead to macro-economic swings, in particular when amplified in structural spillover networks in line with Acemoglu, Carvalho, et al. (2012). I investigate the hypothesis that return premia for granular risks arise due to structural cross-dependencies between assets that reduced-form factor models are not able to capture. That is, I aim to provide insights into asset prices from the perspective of spillovers. Ultimately, I investigate if nonsystematic variations are as immaterial to asset prices as standard factor models suggest, or rather, if such models are unable to capture important sources of compensation for risk.

By neglecting the existence of direct inter-asset links, factor pricing models implicitly assume independence of nonsystematic innovations. However, this assumption is hard to justify empirically as nonsystematic innovations are not orthogonal (for example, see Herskovic, Kelly, et al. (2016)), affect aggregates (for example, see Gabaix (2011)), and predict asset returns (for example, see Goyal and Santa-Clara (2003) and

Ang et al. (2006)). My analysis offers an alternative conceptual framework to the dominion of linear factor models by reconciling empirical challenges with risk-based asset pricing rationales. That is, I provide a framework that incorporates micro to macro effects through concentration and spillovers, and thus embeds systemic effects into equilibrium asset pricing theory. The framework predicts that some granular innovations carry risk premia and offers a new perspective on the ever-evolving ‘zoo’ of newfound factors in the empirical finance literature (compare Feng et al. (2020)). In contrast to factors proxying for exogenous aggregate shocks, the spillover network framework attributes asset comovement to the spread of granular innovations. Therefore, the frequent appearance and disappearance of asset pricing factors could be in line with factors summarizing granular news, where the importance of specific sources of risk changes dynamically.

To that effect, a reoccurring insight of recent contributions to the finance literature is that the workhorse empirical factor models come short of capturing the complexity of asset markets. For example, multiple papers claim to have made out yet another source of systematic variation in the cross-section of returns ¹. Counteractingly, an expanding body of work reconciles this ‘factor zoo’ into small but comprehensive sets of latent factors with statistical methods. For example, Freyberger et al. (2020) selects independent predictors via a group LASSO, Kelly, Pruitt, et al. (2019) proposes a variation of principal component analysis, Kozak et al. (2020) combines principal components with Bayesian shrinkage, Gu et al. (2021) employs a deep auto-encoder network to model latent factors, Bryzgalova et al. (2022) develops a Bayesian approach for factor selection, and Giglio and Xiu (2021) proposes a method to deal with omitted pricing factors when estimating risk premia. Despite all efforts, the linear factor pricing literature has yet to agree on a narrative to explain the incidence of transient pricing factors, and accordingly, no factor model has been universally accepted to describe the dynamics of asset markets.² By incorporating nonsystematic innovations and temporary systemic effects into a new formulation of the stochastic discount factor (SDF), I attempt to better understand granular effects on asset prices.

This paper augments the asset pricing framework with decentralized links between individual assets in place of a fully centralized factor structure. Introducing such links between returns has implications for the SDF, and consequently for equilibrium asset prices. For example, the framework predicts that risk premia for firm-level innovations can evolve through systemic effects, while exposures to granular risks are determined by a directed network of inter-asset linkages. Hence, the granular spillover structure leads

¹Among others, Harvey, Liu, and Zhu (2016) and Feng et al. (2020) survey this literature and discuss the pitfalls of accepting such anomalies as asset pricing factors.

²Compare, for example, Barillas and Shanken (2018) and Baba Yara et al. (2021)

to differences in expected returns across assets and time. My goal is to validate such predictions with an empirical approximation. For that purpose, I obtain connectedness tables from intraday log-volatility data to proxy for the most important inter-asset dependencies. Then, I transfer the estimated volatility spillover networks to return data and use relative weights in aggregate wealth to connect granular variation with the wealth portfolio. On the basis of these approximations, I investigate various conjectures that result from the framework, and collect evidence for the empirical relevance of the proposed economic mechanism.

Tracing back asset price variation to micro-economic innovations empirically is challenging as the structure of the spillover network is unobserved. In light of the large-dimensional setting of the cross-section of asset returns, approaches to structural estimation fail. However, advances in the statistical learning literature have led to the adoption of forecast error variance decompositions (FEVDs, see Pesaran and Shin (1998) and Diebold and Yilmaz (2014)) to large dimensions. Here, regularized estimation opens up the possibility to approximate network structures with many nodes for example in Demirer et al. (2018), Barigozzi and Brownlees (2019), and in Chapter 1. An additional challenge arises from widely efficient financial markets, where lagged cross-dependencies barely exist, even at high frequencies. Hence, estimations frequently resort to second moments to capture the flow of information³. In this paper, I therefore carry out an additional estimation step that proxies the spillover structure in first moments by network estimates obtained from second moments.

On the theoretical level, my framework adds to the literature that seeks to incorporate micro-level shocks into macro-economic frameworks. In that line of work, Gabaix (2011) and Acemoglu, Carvalho, et al. (2012) set precedents that pave the way for idiosyncratic shocks to be considered as drivers of aggregate movement, in the latter case also in connection with networks. Versions of this mechanism have been since picked up mainly in connection with sectoral output, for example in Carvalho (2014) and Acemoglu, Ozdaglar, et al. (2017). In financial markets, the spread of granular events to aggregate outcomes via networks is closely related to the idea of contagion among financial institutions. For example, Elliott et al. (2014) develop a model of cascading failures of financial organization transmitted through direct inter-dependencies amongst organizations. Similarly, Acemoglu, Ozdaglar, et al. (2015) develop a framework to study the fragility of the financial system where an inter-bank network transmits counter-party risk to the extent of systemic failures. I extend this theoretical literature by incorporating the “granular mechanism” into the asset pricing framework. As a novel result, I obtain an expression that links systemically relevant

³compare Demirer et al. (2018), Barigozzi and Brownlees (2019), Chapter 2.

micro-economic innovations to equilibrium expected returns via the SDF. That way, my work offers a unifying view of macro-finance and the micro-economic narratives of corporate finance.

Empirical approximations of the spillover structure of individual asset returns in this paper imply granular risk premia sometimes twice as large as risk premia on common shocks. Historically, the most notable granular risk premia occurred in the energy sector in the 1980s, in the telecommunication and technology sectors around the turn of the millennium, and in the financial sector around the Global Financial Crisis. Lately, as of 2022, material granular risk premia have again arisen in various sectors, most prominently in the technology and healthcare sectors. These empirical results square well with documents of rising firm-level concentration in Autor et al. (2020), and with the idea of information spillovers in first and second moments driven by investors' attention to news as in Hasler and Ornathanalai (2018). While spillovers in my estimations are not limited to announcement days, supporting evidence for price responses to earnings announcements across assets has been presented in Savor and Wilson (2016) and Andrei, Friedman, et al. (2023). Consistent with the periodic occurrence of granular risk premia in this paper, Andrei and Hasler (2015) provide a theory around investors' attention that predicts rising variances and risk premia in environments of increased uncertainty and attention to news. Further, my results are connected to works that measure systemic risk in financial networks. For example, Bianchi et al. (2015) propose a method to measure systemic risk in a model that infers the strength of cross-firm network linkages from linear factor model residuals.

Finally, I find cross-sectional differences in in-sample returns in line with the predictions of the spillover network framework. At the same time, aggregate market returns, and therefore the equity risk premium, have an insignificant relation to granular risk premia. The latter finding supports the idea that granular risks evoke asset comovement independent of the aggregate market, which factor pricing models then register as additional priced factors. By linking asset prices to firm-level spillovers, my work also builds on a diverse literature on concentration and networks in financial markets. For instance, Hou and Robinson (2006) documents that firms in more concentrated industries earn lower returns, and Herskovic, Kelly, et al. (2020) finds that network effects explain the evolution of firm size concentration and the co-movement of firm volatilities. Employing input-output tables as networks, Ahern (2013) shows that stocks belonging to more central industries in the network of intersectoral trade earn higher returns. Similarly, Herskovic (2018) study the asset pricing implications of changes in the sectoral input-output network, where network concentration and sparsity relate to priced systematic risk.

The rest of this research paper is structured as follows: Section 3.2 explains the idea of inter-asset linkages in asset pricing frameworks. Section 3.3 presents the methodology used to approximate the spillover network. Section 3.4 exhibits the estimated spillover models and presents descriptive evidence. Finally, I empirically investigate the predictive power of the spillover network model and discuss some findings in light of classical asset pricing models in Section 3.5, before Section 3.6 concludes.

3.2 Framework

The Capital Asset Pricing Model (CAPM), proposed by Sharpe (1964) and Lintner (1965), ties the stochastic discount factor (SDF) m_t to the return on aggregate wealth R_t^W :

$$m_t = a + bR_t^W. \quad (3.1)$$

It follows from the fundamental asset pricing equation, $1 = \mathbb{E}[m_t R_t]$, that only comovement with the return on wealth will be compensated with an excess return, where $R_t = 1 + r_t$ is the gross return on the asset.

To make the return on wealth empirically traceable, the empirical finance literature has adopted linear factor pricing models as the standard workhorse. Such models approximate the SDF by one or more observed common factors F that summarize asset returns and are commonly thought to indicate times of distress to aggregate wealth. Most prominently, the market CAPM uses the return on the stock market portfolio as a single proxy variable $R_t^W \approx R_t^M = \sum_{i=1}^N w_i R_{i,t}$. Then, the factor loadings β_i , that is, the covariances with the pricing factors, determine individual asset returns r_i together with the nonsystematic disturbances ε_i as $r_{i,t} = \beta_i F_t + \varepsilon_{i,t}$. Hence, only assets that correlate with the common factors through β_i will receive a risk premium, while the independent mean-zero innovations ε_i do not affect expected returns. Panel 3.1a of Figure 3.1 illustrates this concept in simplified form.

However, the assumptions underlying factor pricing models are too restrictive to adequately represent the complex interconnectedness of financial markets. Particularly, it is implausible that innovations at the asset level ε_i are always confined to the individual asset. For example, this assumption prohibits contagion of news such that they impact assets of a certain industry and not of another, unless the model contains a factor that captures that exact sectoral shock. Therefore, a major shortcoming in the centralized structure of factor pricing models is the fact that they are silent about inter-asset relations that go beyond systematic comovement. This also forbids the possibility of asset-level news turning systemic as widely presumed during the

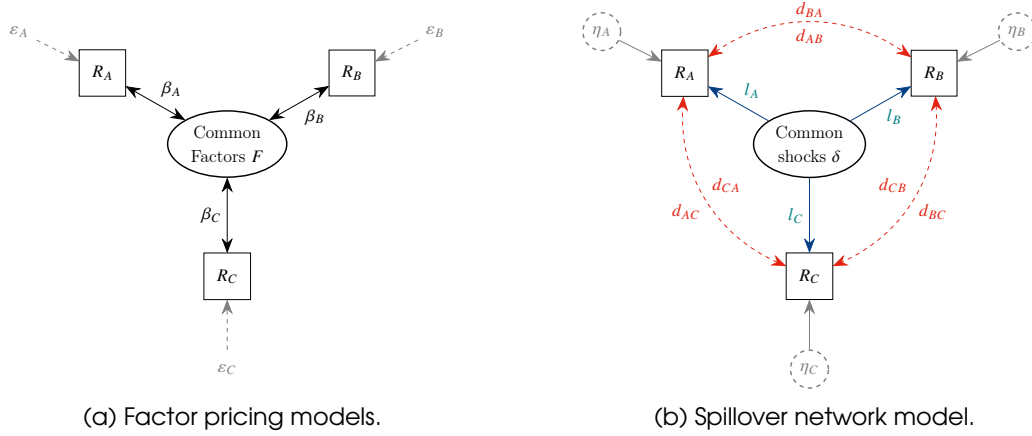


Figure 3.1: Graph representations of a conventional linear factor model and a network model.

Global Financial Crisis. Hence, I propose a return model that includes additional direct linkages between assets.

More specifically, factor pricing models assume that assets are only related through their exposure to common factors and that there are no direct connections or effects between individual assets. However, the assumption of fully independent shocks is questionable in an increasingly interconnected world. Collusion or competitive relations, overlapping suppliers or customers, cross-corporate holdings, exposure to financial intermediaries, or sentiment spillovers all make presumably idiosyncratic asset-level innovations interconnected. In support of this notion, Herskovic, Kelly, et al. (2016) find that there is a strong factor structure in the nonsystematic volatilities of individual assets. Entertaining the idea of such direct linkages between assets opens up the possibility for firm-level events to turn systemic if their influence on total wealth through the cross-section of assets is sufficiently large. That is, following arguments of Acemoglu, Carvalho, et al. (2012), not only innovations to constituents with large index weights may affect the aggregate, but the structure of a network of direct spillovers between assets can lead to an outsized effect of any asset-level innovation on aggregate wealth. An implication of this is that there are nonsystematic innovations that can only be partially diversified away due to their impact on the wealth portfolio, and thus, directed linkages between assets play a crucial role in explaining asset prices.

To understand how spillovers affect asset returns, consider a model that adds direct linkages between returns. That is, let asset i 's return be a function of all other asset's returns in addition to common shocks δ_t and a granular disturbance $\eta_{i,t}$. Formally,

$$r_{i,t} = l_i \delta_t + \sum_{j \neq i} d_{ij} r_{j,t} + \eta_{i,t}, \quad (3.2)$$

where l_i is a loading coefficient on the common shock δ_t , and d_{ij} is a coefficient measuring the direct spillover strength from asset j 's return to asset i . Panel 3.1b of Figure 3.1 illustrates this network connectedness model graphically, where the additional transmission channels d_{ij} between assets are depicted as dashed arrows. As a result, adding spillovers between asset returns makes them subject to granular innovations of other assets $\eta_{j,t}$ next to the common innovations δ_t . Rewrite (3.2) in vector form to obtain

$$r_t = L\delta_t + D^- r_t + \eta_t \quad (3.3)$$

$$= \Gamma L\delta_t + \Gamma\eta_t, \quad (3.4)$$

where D^- is a spillover matrix with zeros on the diagonal and entries d_{ij} otherwise, and $\Gamma = (I - D^-)^{-1}$. Now it is easy to see that in the absence of spillovers; that is, if $D = I \Rightarrow D^- = 0$, (3.3) reduces to a factor model, with the common innovations δ_t being the factors.

Next, rewrite the market return as the return on the value-weighted portfolio of all available assets, that is, $r_t^M = \sum_{i=1}^N w_i r_{i,t}$:

$$r_t^M = w' \Gamma L \delta_t + w' \Gamma \eta_t. \quad (3.5)$$

As a consequence, the effect of the granular innovation vector η_t on the market return is negligible only if all of the spillover-adjusted weights $w' \Gamma$ are sufficiently small. Given the large number of assets available to investors, this “diversification argument” is reasonable as long as the concentration of asset weights in the cross-section is low (see Gabaix, 2011). In today’s concentrated markets, however, a small set of stocks make up a significant portion of the aggregate investment portfolio.⁴ Hence, when approximating the SDF through factors tied to aggregate wealth, it is unlikely that nonsystematic innovations to assets with large spillover-adjusted weights never impact the market return. That is, granular innovations are not always orthogonal to the SDF due to nonnegligible reverse transmission channels from the assets to the aggregate.⁵ Consistent with this idea are the findings of Goyal and Santa-Clara (2003) and Ang et al. (2006), which suggest that nonsystematic innovations carry information about expected asset returns.

Given the extended formulation of the return equation in (3.3), the market return in (3.5) consists of two exogenous components. First, it is subject to common innovations δ_t , where the size of the effect of δ_t on the market return not only depends on the

⁴Figure 3.6 in Appendix 3.A.5 illustrates equity market concentration in the CRSP universe over time. Similarly, Bae et al. (2021) document rising levels of concentration in international equity markets.

⁵This is also directly supported by results in Chapter 2.

weighted average of individual exposures L , but also on the structure of the spillovers in Γ . Second, the N -dimensional granular innovation vector η_t can also affect the market return. More precisely, if Γ is such that a granular innovation transmits to assets with sufficient combined market weights, some elements in η_t will gain systemic impact on the aggregate market return. Note that the interplay of both the structure of the spillover network and the concentration of market capitalization weights plays an important role in evaluating the systemic relevance of individual assets. For example, if a granular innovation originates from an asset with a small market weight, it can still have a large aggregate effect if it causes variation in assets with large weights. Returning to the SDF in (3.1), I obtain a refined CAPM pricing kernel with directed spillovers between assets:

$$m_t = a + bw'\Gamma L\delta_t + bw'\Gamma\eta_t. \quad (3.6)$$

Note that this kernel still represents the market return as a single proxy for the return on wealth, but allows a more detailed look at the sources of variation through a decomposition into common and granular components. Given this formulation, both common and granular innovations can influence asset prices in principle. More precisely, the fundamental asset pricing equation implies that only assets that co-move with the driving forces of the SDF are compensated by an excess return. Hence, it immediately follows from (3.6) that potential risk premia evolve from the co-movement with common innovations as well as with the most impactful granular innovations. The former is in line with the classical notion of linear factor models, while the latter is new in this specification. In other words, if the spillover network is such that an asset affects a large share of aggregate wealth as measured by the market weights, any asset that co-moves with the granular innovations to this systemically relevant asset earns an additional expected excess return.

To better understand how the decomposed CAPM kernel offers a finer interpretation of asset prices, consider the following beta representation of the expected excess returns of any asset in the economy:⁶

$$\begin{aligned} \mathbb{E}[R_{i,t}] - R^f &= \underbrace{-bR^f w'\Gamma L\Sigma_\delta}_{\text{common risk premium}} \underbrace{\Sigma_\delta^{-1} \text{Cov}(r_{i,t}, \delta_t)}_{\text{exposure to common risks}} + \underbrace{-bR^f w'\Gamma\Sigma_\eta}_{\text{granular risk premium}} \underbrace{\Sigma_\eta^{-1} \text{Cov}(r_{i,t}, \eta_t)}_{\text{exposure to granular risks}} \quad (3.7) \\ &= \lambda_t \beta_i + \xi_t \gamma_i. \end{aligned}$$

Here, assets earn compensation for exposure to common shocks in the form of the common risk premia λ_t similar to factor risk premia in factor pricing models, where

⁶See Appendix 3.A.1 for a more detailed derivation of this equation.

the magnitude of the risk premia is proportional to the sensitivity of the market portfolio to the respective shocks. Any asset's common shock exposure is captured through the transformed loadings $\beta_i = [\Gamma L]_i$, that is, factoring in indirect exposure and feedback effects through the spillover network as in (3.4). Novelly, assets can earn additional compensation for exposure to granular shocks, given that the particular shock is systemically relevant for the market portfolio, and hence, carries a granular risk premium. Given the large number of assets in the cross-section, the granular risk premia vector ξ_t is approximately sparse with few elements being nonzero. Asset i 's exposure to the individual granular risks is then given by the i -th row in Γ , γ_i .

The hypothesis of direct return spillovers has several implications for the validity of linear factor pricing models. When approximating the return on wealth by a small set of common factors, one simplifies the complex interactions underlying its movement and cannot capture the more nuanced nature of a multitude of potentially systemically relevant granular innovations. Note that the negligence of granular effects is common to all pricing models that approximate the return on wealth through a set of common factors, including empirical applications of the CAPM that use a market proxy as the single factor. Furthermore, while multi-factor models such as those motivated by arbitrage pricing theory or statistical latent factor models offer closer approximations to the return on wealth, this may be due to the additional factors capturing some systemically relevant granular risks instead of common innovations to all assets.

Contrary to the bulk body of the literature that assumes correlation is sufficient to understand pricing behavior, I investigate the role of causality in asset pricing. To understand why that matters, consider two assets, A and B, where nonsystematic innovations in η_A have a causal effect on the return of asset B, r_B , but not vice versa. Then, the correlation of the returns r_A and r_B is the outcome of nonsystematic innovations on the asset level at A, but not at B. The standard correlation argument implies that there is limited diversification potential for both assets. However, the spillover framework predicts that nonsystematic innovations in B can be diversified away while the innovations to A are systemic and cannot be diversified. Thus, assets with greater causal effects on others are more likely to be systemically relevant and to induce risks to total wealth. As a consequence, adding the notion of causality to asset pricing frameworks enhances our understanding of the sources of priced risks, and therefore, the sources of risk premia and asset returns.

In the following, I attempt to collect evidence of granular effects being at play in equity markets. A major difficulty here is the approximation of the structural parameters and shocks, namely the spillover network D^- and the innovation vectors δ_t and η_t .

3.3 Empirical strategy

Probing hypotheses connected to the above framework is challenging due to two reasons. First, the presence of the structural parameters D^- only allows identification of (3.3) under identifying restrictions. Second, the number of parameters in the spillover network is $N^2 - N$, that is, quadratic in the width of the cross-section. Therefore, the large-dimensional nature of the problem impedes the econometrician's ability to come up with trustworthy estimates for D^- while further aggravating the search for suitable identifying restrictions. Therefore, in this paper, I use two assumptions that make the problem empirically traceable. First, I restrict my analysis to a selected subset of the cross-section of equities to reduce the number of nodes N in the structural network. This assumes that the sampling procedure is representative of the wider cross-section in that the most impactful assets are likely to be included. Second, I use a proxy for the spillover matrix D^- that I obtain from a separate estimation on second moments. More precisely, I assume that the spillover matrix is proportional to a proxy table $\tilde{D}^-$ as $D^- \approx c\tilde{D}^-$. Finally, I estimate the model parameters through maximum likelihood estimation.

3.3.1 Data & rolling samples

In light of the limitations associated with estimations in high-dimensional parameter spaces, it is not feasible to estimate the spillover network model for the full cross-section of stock returns. This is aggravated by the dynamic nature of stock markets where tickers constantly start and cease to be listed. Therefore, I restrict the cross-section to be considered in the empirical analysis to a subset of 100 stocks derived from the full CRSP universe. However, to be able to draw broader conclusions about aggregate effects, the goal is to select the assets that account for the largest swings in investor wealth and are therefore most representative of the overall market. More specifically, since I am ultimately interested in the effects on investor wealth, I want to maximize the chances to include those assets that induce the most variation in the aggregate market portfolio. Therefore, I calculate the average market capitalization of each stock available in the cross-section of CRSP,⁷ and select the 100 assets with the largest weights by that measure.

To account for dynamic additions and subtractions to the most impactful stocks, I conduct the sampling procedure at the end of each calendar month from December 1973

⁷I consider securities with exchange code $\in \{1, 2, 3\}$ and share code $\in \{11, 12\}$. Additionally, I require assets to have an intraday high price and low price available for at least 99% of the trading days in the last 12 months. Finally, I exclude assets that are delisted on the sampling day and therefore do not have subsequent price information.

to December 2022. More specifically, I employ rolling windows with a length of one calendar year to perform the sample selection such that I obtain overlapping samples of 100 stocks, which I will perform my analysis on. That is, samples are overlapping sets of dates and assets, where each sample contains a cross-section of the 100 most important stocks by means of average market capitalization in one year's worth of daily observations. Despite comprising only around 2% of the number of stocks in the CRSP universe, the samples make up half of the total market capitalization on average, while this share never drops below 41% (see Figure 3.2). Consequently, I consider the samples representative of the entire equity market. Finally, in addition to daily pricing data, I also match each ticker to one of 12 sectors following the industry definitions of Fama and French (1997).

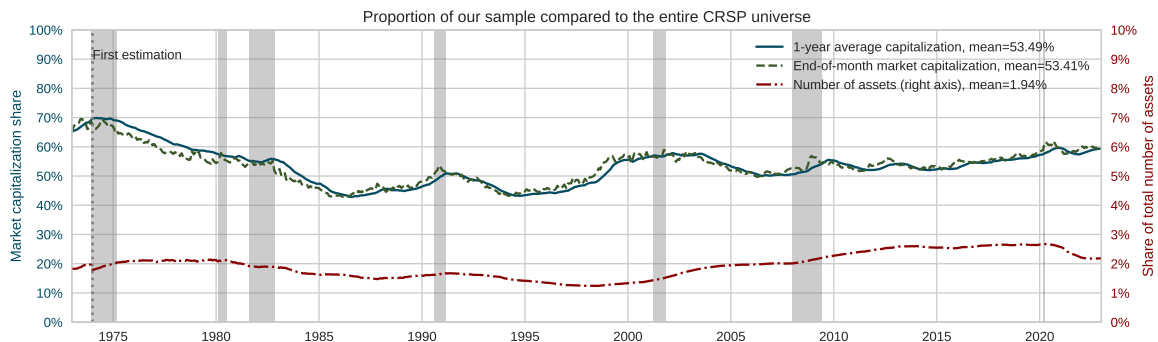


Figure 3.2: Fraction of the CRSP dataset contained in the rolling samples. The total number of assets in CRSP ranges from 3795 in April 2020 to 8174 in June 1998. At the end of each month, I sample the 100 assets with the largest market capitalization in the previous 12 months. The shaded intervals indicate NBER recessions.

3.3.2 Approximating the spillover table

A central part of the spillover network model are the structural linkages in the spillover matrix D^- . However, even after having limited the size of the network to 100 in the sampling step, structural estimation is still infeasible. Therefore, it is crucial for any empirical analysis of the model to find an adequate spillover network that approximates the flow of information between assets. That is, the proxy network should contain links between all assets contained in our samples, ideally in the form of a directed graph. While the absolute magnitude of the edges is of secondary importance with the global scaling parameter c , a suitable proxy graph should contain edges that describe well the relative strength of the direct spillovers. On the asset level, the literature has proposed to construct networks from observable variables, for example, from board interlocks in Chuluun et al. (2017), from common ownership in Azar et al. (2018), from multinational production data in Boehm et al. (2019), or from customer-supplier linkages in Herskovic, Kelly, et al. (2020). However, since I am primarily interested in

the flow of price-relevant information and not so much in a specific micro-economic mechanism that leads to the connections, statistical procedures that infer linkages directly from prices offer more appropriate approximations.

To that end, I resort to the methodology of FEVDs, which has successfully been employed in other studies of spillover networks before. That is, I estimate generalized FEVDs as in Pesaran and Shin (1998) for the 100 included assets, and follow the arguments of Diebold and Yilmaz (2014) to establish the FEVD estimates as suitable connectedness tables that represent structural links. Given the fact that there is little forecastability in return observations but substantial evidence in favor of the predictability of second moments, return volatilities make up a more informative basis for the estimation of directed connections. Hence, as do Barigozzi and Brownlees (2019) and Chapter 2, I employ volatility data at the daily frequency to infer asset-level connectedness tables. Appendix 3.A.2 presents detailed descriptions of how to apply the methodology to construct the tables. Henceforth, let the fitted FEVD tables be denoted as $\hat{D}$.

With the number of nodes being fixed at 100, the estimation of the FEVD networks poses a dimensionality problem in both estimation stages. More precisely, fitting a VAR(1) model with 100 assets requires the estimation of 10,000 parameters, and the residual covariance matrix contains an additional 5,050 free parameters. To mitigate the inaccuracies that result from the large-dimensional estimation, the statistical learning literature has proposed several techniques. In this application, I use the Adaptive Elastic Net of Zou and Zhang (2009) to estimate the regression coefficients and the Graphical LASSO of Banerjee et al. (2008) and Friedman et al. (2008) for the residual covariance matrix with cross-validated regularization parameters. The use of large-dimensional techniques in similar contexts has previously been demonstrated in Demirer et al. (2018) and in Chapter 1, where the latter presents instructions for large-dimensional estimation. Appendix 3.A.3 contains detailed accounts of the employed estimators. Note that even though the estimation steps assume sparsity in the coefficient matrix and the inverse error covariance matrix, the resulting FEVD tables are generally not sparse. This is consistent with an unobstructed flow of information between assets, countering the criticism of sparsity in Giannone et al. (2021).

The FEVD estimate $\hat{D}$ is a network matrix whose entries $\hat{d}_{ij}$ quantify the percentage of variation of i that can be explained by variation in j . This fits well with the intuition of information spillovers, where the release of new information about one asset affects the valuation of other assets. That is, I understand connectedness in the sense of one asset's investors reacting to price changes of other assets, be it for fundamental or behavioral reasons. While this leads to instantaneous return correlations, dependencies

also become visible in return volatilities. I leverage this fact for inference on the connectedness structure from second moments. Hence, embedded in the use of FEVD spillover tables to approximate the structural coefficients on the return level is the assumption that inter-asset relations are similar in the first and second moments such that stronger links in $\hat{D}$ correspond to stronger links in D^- . While Hasler and Ornathanalai (2018) provide a theoretical mechanism and empirical evidence for volatility spillovers and return spillovers to occur simultaneously, a shortcoming exists in the inability to detect negative linkages on the return level. However, based on the overall level of correlation in the stock market, I suspect that such relationships are seldom and the approximation error due to missed negative signs is small.

3.3.3 Maximum likelihood estimation

To obtain estimates for the model parameters, I estimate an empirical counterpart of (3.3) via a maximum likelihood specification. That is, given a proxy for the spillover table $\tilde{D}^-$ and for the common shock observations $\tilde{\delta}_t$, I maximize the likelihood of observing the sampled asset returns under the assumption of independent Gaussian errors.⁸ More precisely, I require estimates of the parameters $\theta = \{\mu, L, c, \Sigma_\eta\}$ in:

$$r_t = \mu + L\tilde{\delta}_t + c\tilde{D}^-r_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \Sigma_\eta), \quad \forall t = 1, \dots, T. \quad (3.8)$$

Here, μ is a $N \times 1$ vector of intercepts, L is a $N \times K$ matrix of common shock loadings, c is a scalar scaling parameter representing the spillover strength, and Σ_η is a diagonal matrix with the variances of the granular shocks on its diagonal. Note that despite this equation being structural, the spillover structure has a fixed network architecture and is only scaled via c . To be precise, there are $N(2 + K) + 1$ free parameters in μ , L , c and Σ_η , compared to $N(N + 1)/2$ degrees of freedom. Hence, as long as the number of factors K is small relative to the number of assets N , using the proxy matrix $\tilde{D}^-$ imposes a sufficient number of identifying restrictions for the identification of (3.8) following the necessary condition for identification in Rothenberg (1971). I derive the maximum likelihood estimator $\hat{\theta}_{MLE}$ in more detail in Appendix 3.A.4.

It is worth looking into the employed proxies $\tilde{\delta}_t$ and $\tilde{D}^-$. In the former case, any proxy for the common shocks δ_t is admissible in the regression in principle. For instance, standard asset pricing factors that arguably proxy common movements among assets are suitable candidates. However, such factors are constructed under the notion of no network effects. For example, any factor that applies a value-weighting

⁸Note that the estimator can be interpreted as the quasi-maximum likelihood estimator under more general distributional assumptions (compare White (1982)). I account for potential misspecification of the errors through the use of robust standard errors in subsequent analyses.

risks capturing aggregate movements that stem from granular sources with systemic impact. Therefore, including such factors impairs the ability to cleanly separate the effects, and thus, I suggest to instead use equally weighted asset returns as a single gauge for common shocks. Second, the FEVD tables $\tilde{D}^-$ as discussed in Section 3.3.2 are suitable proxy matrices for the spillover structure in D^- as they establish the relative size of the contributions to variations in the respective asset. Precedents of employing connectedness tables that are not directly inferred from the modeled data to parameterize structural models can be found across the literature that makes use of input–output tables in production networks.⁹ By scaling $\tilde{D}$ to the data, my approach goes one step further to bring the data and the external table together.

Note that there are two shortcomings of estimating the structural model in (3.8) with the use of FEVD tables as a proxies. First, the fact that the FEVD tables are estimated from second moments falls short of the goal of inferring the spillover network of returns, and thus, first moments. However, there is likely a close relationship of returns and return volatilities which allows to adequately estimate relative size effects of return spillovers from volatility spillovers. Second, FEVD tables report variance decompositions, that is, they only have positive entries. Consequently, the FEVD proxy matrices limit the structural approximation to positive comovement. The usual procedure to perform a likelihood-ratio test for over-identifying restrictions as in Lütkepohl (2005, Chapter 9) to find out whether any of those assumptions are too strict is impaired by the large-dimensionality. I report the results of such tests nonetheless.

3.4 Tracing granular risk premia

In this and the following section, I investigate several conjectures based on the framework in Section 3.2 with the help of the rolling empirical approach from Section 3.3. First, I attempt to establish the suitability of the empirical spillover tables to describe the return data. Specifically, I look into the explanatory power of the tables with regards to the samples of individual asset returns but also examine how the tables relate to the aggregate equity market. I also present descriptive results about the origins of the risk premia implied by the estimates.

Conjecture 1 *Variance spillover tables have explanatory power with regards to return spillovers.*

The first conjecture probes the suitability of the method to gain insight into asset returns rather than just second moments. A first approach tests if the respective

⁹For a review see Carvalho and Tahbaz-Salehi (2019).

spillover table should be included in the return formulation of (3.8) via the restriction $H_0 : c = 0$. However, in line with the vast literature of factor asset pricing models, the residuals after accounting for common shocks are still correlated and a corresponding Wald test statistic rejects the null hypothesis with a probability indistinguishable from 1 in all periods. While this is first evidence that it is sensible to add spillovers to the common shock structure in our samples, it does not immediately follow that the tables $\tilde{D}$ tables are necessarily good proxies for the spillover structure. Therefore, it is helpful to consider some alternative tables that could serve as proxies for the return spillovers. As the analysis is performed over rolling samples, I compare the explanatory power of various candidate tables in each sample period separately and then aggregate the results of all periods.

	free para- meters	un- stable model	failed esti- mation	over- identi- fied	log-likelihood				AIC			
					mean	rank	average rank	beats FEVD	mean	rank	average rank	beats FEVD
Panel A: Comparison of candidate proxy tables												
FEVD $\tilde{D}$	301	0	0	589	73,103.92	1	1.69	-	-145,605.85			-
equal links $[1/N]_{N \times N}$	301	0	0	589	72,173.26	5	4.80	21	-143,744.53			21
random links $\mathcal{U}(0, 1)$	301	0	2	589	72,142.32	6	5.78	17	-143,682.79			17
symmetric $(\tilde{D} + \tilde{D}')/2$	301	0	0	589	73,076.19	2	2.04	209	-145,550.38			209
square root $\tilde{D}^{-1/2}$	301	0	0	589	72,752.38	4	3.17	126	-144,902.77			126
FEV $FEV(21)$	301	0	0	589	72,925.80	3	3.52	32	-145,249.60			32
Panel B: Model restrictions using FEVD proxy table $\tilde{D}$												
full model	301	0	0	589	73,103.92			-	-145,605.85	2	2.26	-
unconstrained c	301	283	0	589	72,746.94			69	-144,911.16	3	2.85	69
$\mu = 0$	201	0	0	589	73,061.60			-	-145,721.20	1	1.03	581
$L = 0$	201	0	5	589	72,204.01			-	-144,004.51	4	3.87	-
$c = 0$	300	0	0	589	70,537.31			-	-140,474.61	5	4.99	-

Table 3.1: Model comparison with various candidate proxy tables and restrictions. In Panel A, each alternative table is used as a proxy in (3.8). Panel B estimates restricted models with the FEVD table $\tilde{D}$ as proxy in all cases. Models are fitted at 589 end-of-month sampling dates from Dec-1973 to Dec-2022. The columns ‘unstable model’, ‘failed estimation’, ‘overidentified’, and ‘beats FEVD’ count the number of times where the respective model converged to an unstable SVAR process, did not converge, fails a likelihood-ratio test for overidentifying restrictions with 99% confidence, and outperformed the baseline specification. ‘Average rank’ is calculated from ranking specifications in each period and averaging across samples.

Table 3.1 contains the results over the 589 sample periods of the analysis. It compares the model likelihoods for a variety of candidate tables to explain the samples. First, looking at log-likelihoods suggests that the chosen proxy $\tilde{D}$ has the best fit to the data overall, followed by a symmetric version of $\tilde{D}$. Even though there are periods in which symmetric tables excel, this comparison also testifies to the importance of directed spillovers rather than just correlations in most periods. Ranking the candidate models in each period yields additional support for the proxies of choice, with FEVD tables scoring the best average rank (1.69) across periods. Next, Panel B compares various model restrictions.¹⁰ Restricting the model to not include a common component

¹⁰For illustration, not imposing a stability condition on the structural process (see Appendix 3.A.4) leads to unstable estimates in 283 periods, sometimes with higher in-sample likelihoods.

($L = 0$) or spillovers ($c = 0$) confirms that both add relevant information to the model. The latter case is interesting as the restricted model is the same as a one-factor model with the common shock as the factor. In a direct model comparison, excluding spillovers from the model noticeably decreases the model's informativeness with a large increase in AIC. Hence, despite rejecting the validity of its restrictions in a likelihood-ratio test in all periods, the spillover model outperforms a one-factor model by a notable margin. Note that the even stricter independence assumption underlying factor-pricing models is also always rejected by a test for overidentifying restrictions for the restricted model with $c = 0$. Finally, it is interesting to note that a model that excludes asset intercepts is preferred in terms of AIC in most periods. However, to stay consistent with the standard practice of factor models to include intercepts, I will conduct subsequent analyses based on the full model instead of the restricted one.

The explanatory power of the proxy tables for returns is remarkable as it supports the use of spillover tables recorded from second moments in applications to model first moments. This result is novel as most modeling of asset-level relationships has been confined to correlations. Further, it opens up the possibility for causal analyses of asset prices without relying on exogenous factors. Given the information content of the volatility spillover tables, they may serve to uncover new insight into the mechanics of asset markets.

Conjecture 2 *The network architecture of return spillovers relates to aggregate return and volatility.*

Next, (3.5) links the structural tables Γ to the return on the aggregate market r^M . Hence, if the estimated tables were to appear in that equation rather than just the identity matrix, it should be possible to relate properties of the estimated tables $\hat{\Gamma} = (I - \hat{c}\tilde{D})^{-1}$ to moments of the aggregate return in the respective samples. I consider three such properties in my analysis: connectedness, which is the average absolute value of off-diagonal elements $\frac{1}{N(N-1)} \sum_{i \neq j} |\hat{\gamma}_{ij}|$, asymmetry, which is the fraction of the off-diagonal entries in $\hat{\Gamma}^-$ that is asymmetrical $\sum_{i \neq j} |(\hat{\gamma}_{ij} - \hat{\gamma}_{ji})/2| / \sum_{i \neq j} |\hat{\gamma}_{ij}|$, and concentration, which is the Herfindahl index over the column sums of $\hat{\Gamma}$.¹¹ I consider the sample-period volatility and return of the CRSP value-weighted index as response variables. In other words, I want to understand whether return spillover tables help to predict aggregate volatilities and returns.

Table 3.2 presents the results of contemporaneous regressions with volatilities and returns of aggregate indices over the 1-year samples as the dependent variables and

¹¹See Figures 3.9, 3.10, and 3.11 in Appendix 3.A.5 for plots of these variables over time.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Panel A: Volatility (annualized)</i>							
constant	0.0446*** (3.3898)	0.1091** (2.1036)	-0.1104*** (-3.1299)	-0.1000* (-1.8194)	-0.1646*** (-4.2457)	-0.0828** (-2.1069)	-0.1605*** (-3.8489)
connectedness	3.6356*** (7.8424)			4.5601*** (7.3096)	2.8443*** (6.4108)		2.7615*** (4.8247)
asymmetry		0.1998 (0.8387)		0.5135*** (2.7472)		-0.3696*** (-3.2765)	-0.0344 (-0.2807)
concentration			22.4417*** (7.3067)		19.7513*** (5.5368)	27.2213*** (7.8159)	20.2739*** (4.9054)
T	589	589	589	589	589	589	589
R-squared	0.2317	0.0183	0.3739	0.3377	0.5104	0.4197	0.5107
<i>Panel B: Return (annualized)</i>							
constant	0.1060** (2.3934)	0.2390*** (2.6618)	0.3660*** (3.3972)	0.3303** (2.5552)	0.3584*** (3.1577)	0.3832*** (3.3834)	0.3896*** (2.9269)
connectedness	-0.5580 (-0.3675)			-1.9923 (-1.1001)	0.3967 (0.2423)		-0.2282 (-0.1298)
asymmetry		-0.6596* (-1.6558)		-0.7966* (-1.8760)		-0.2316 (-0.7073)	-0.2593 (-0.7550)
concentration			-23.4550** (-2.5282)		-23.8302*** (-2.6075)	-20.4605** (-2.2888)	-19.8864** (-2.4692)
T	589	589	589	589	589	589	589
R-squared	0.0009	0.0328	0.0672	0.0429	0.0676	0.0701	0.0702

Table 3.2: Regressions of aggregate index moments on characteristics of the estimated tables $\hat{\Gamma}$ for the samples spanning from Dec-1973 to Dec-2022. The dependent variables are the annual volatility and annual return of the CRSP value-weighted index over the respective sample period. t-statistics are calculated from Newey-West standard errors with $T^{1/4}$ lags.

traits of $\hat{\Gamma}$ as the predictors. The number of samples is equal to the number of end-of-month sampling dates spanning from Dec-1973 to Dec-2022. Panel A testifies that periods with larger connectedness and concentration exhibit larger aggregate volatility, where asymmetry only has a significant relation with volatility when conditioned on connectedness. Given the link of the market return with the SDF, this could mean that a denser spillover network implies a more volatile SDF. Turning to returns in Panel B, larger spillover asymmetry and concentration is associated with lower aggregate returns. An interpretation of this finding could be that return transmissions mainly occurs when there are large adverse shocks on the granular level, that consequently drag down the market overall. Contrarily, the evidence on the relation of connectedness with aggregate returns is inconclusive.

Overall, there is sufficient evidence in Table 3.2 to conclude that the return spillover table estimates $\hat{\Gamma}$ link to moments of the market aggregate, which is a necessary precondition for the subsequent analysis of asset prices in the spillover framework. Put differently, if return spillovers are a relevant predictor of aggregate wealth, then they may also be a meaningful ingredient in asset pricing frameworks. In heuristic terms, the results in Panel A suggest that the risk of tail-events is higher if the network is

dense and concentrated. In line with Kelly and Jiang (2014), big aggregate movements are therefore more likely if individual firms are linked. Further, Panel B indicates that negative aggregate returns often appear in conjunction with elevated levels of network concentration. Thus, return spillovers open up an explicit channel for granular shocks to drive aggregate movement, similar to the findings of Gabaix (2011) for GDP growth, of Blank et al. (2009) for banking distress, and of Chapter 2 for asset markets.

Conjecture 3 *Granular risk premia exist and have have economically significant magnitude, at least temporarily.*

After the evidence on Conjectures 1 and 2 showed that the spillover tables are useful in their application to returns, I want to trace out the granular risk premia as implied by (3.7). That is, given the fitted model parameters, I calculate empirical estimates for the common risk premia $\hat{\lambda}_t$ and granular risk premia $\hat{\xi}_t$:

$$\begin{aligned}\hat{\lambda}_t &= -bR^f w' \hat{\Gamma} \hat{L} \hat{\Sigma}_\delta \\ \hat{\xi}_t &= -bR^f w' \hat{\Gamma} \hat{\Sigma}_\eta\end{aligned}\tag{3.9}$$

Here, I employ the estimates $\hat{\Gamma}$, $\hat{L}$, $\hat{\Sigma}_\delta$, and $\hat{\Sigma}_\eta$, which I obtain from the estimations of (3.8) in each sample period. Additionally, I let b and R^f be constant throughout the samples for demonstrative purposes.¹²

Note that cross-sectional regressions such as the Fama-MacBeth procedure are unsuitable methods in this context to determine the magnitude of the risk premia λ_t and ξ_t for statistical reasons. First, the number of potential risk premia exceeds the degrees of freedom in the respective cross-section. That is, applying cross-sectional procedures would require estimating $K + N$ risk premia from just N cross-sectional observations, where K is the number of common shocks in the model and N is the number of included assets. Second, granular risk premia are likely short-lived, such that it is unlikely to achieve convincing statistical power when aggregating cross-sectional results over several periods. The time-varying nature of the granular risk exposures γ further aggravates this problem, as cross-sectional procedures rely on the consistency of the estimated exposures. Finally, the dynamic composition of asset markets generally, and the cross-sectional samples in particular, makes it impossible to construct a fixed panel to perform standard procedures.

Figure 3.3 displays the calculated risk premia for all common and granular shocks in the rolling samples. Using the equally weighted return index as a single proxy for

¹²Numeric values for a and b can be obtained by solving the system $\mathbb{E}[(a + bR^M)R^M] = 1$, $\mathbb{E}[(a + bR^M)R^f] = 1$, which results from the fundamental asset pricing equation. I assume values of $R^f = 1.02$, $\mathbb{E}[R^M] = 1.07$, and $\text{Var}(R^M) = 0.2$ for the approximation, resulting in the values $a \approx 1.2427$ and $b \approx -0.2451$. The specific assumption about R^f , R^M , and $\text{Var}(R^M)$ only has a minor impact on the magnitude of the calculated premia.

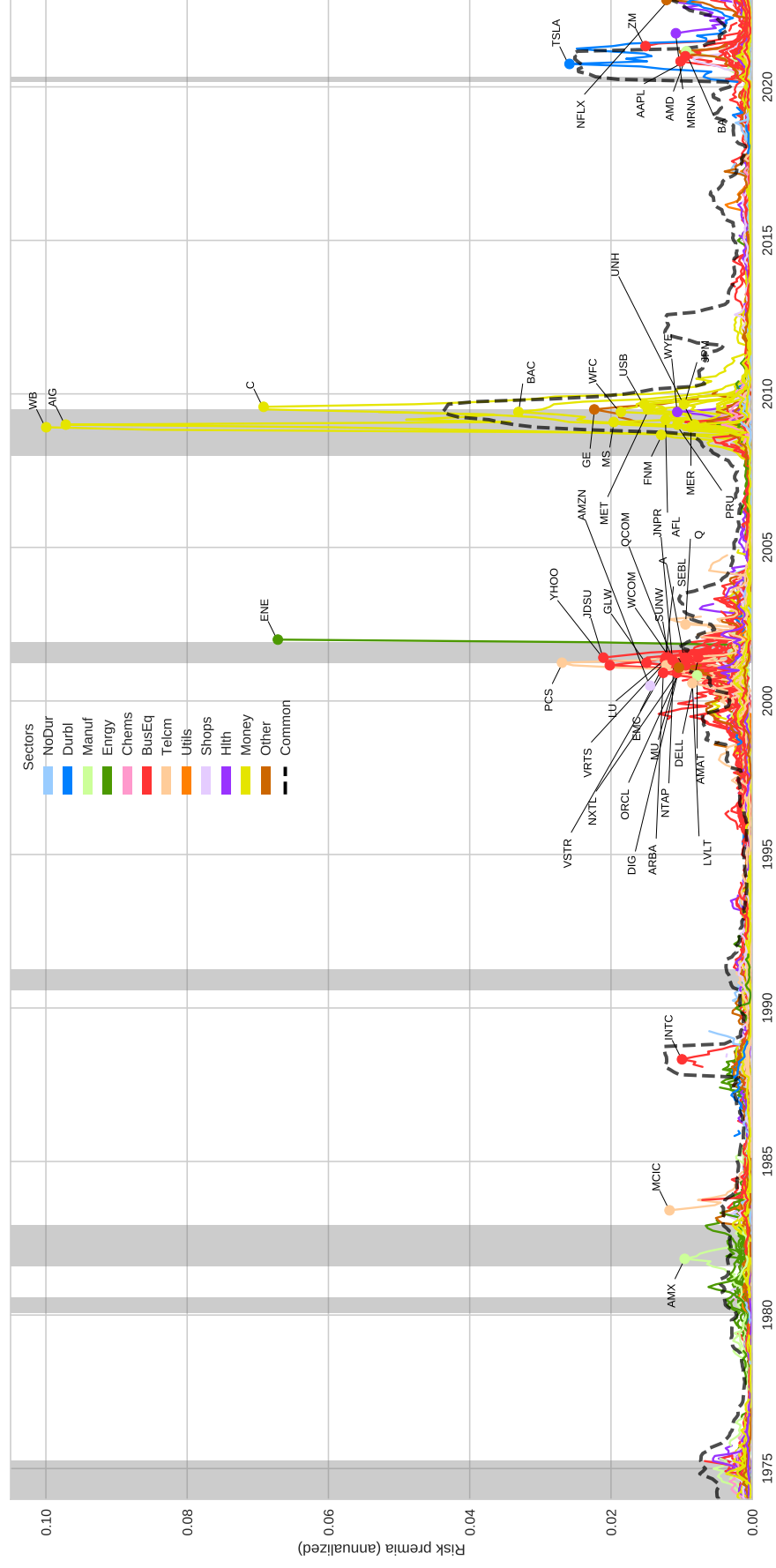


Figure 3.3: Annualized common and granular risk premia estimates for samples spanning from Dec-1973 to Dec-2022. Granular risk premia are matched to one of 12 sectors following the Fama and French (1997) industry definitions. Peaks above 0.075 are labeled with the respective ticker. The shaded intervals indicate NBER recessions.

common shocks, the corresponding common risk premium is the most dominant force in most periods. However, individual granular risk premia often spike up for short time intervals and temporarily exceed the common risk premium. In line with systemic risks accumulating in the financial sector as in Billio et al. (2012), the largest granular risk premia appear during the financial crisis of 2008, where risks associated with Wachovia, AIG, and Citi command large compensations. Similarly, the premia for risks of new economy firms such as Amazon, Yahoo, and Sprint Corporation peak at the height of the Dot-com boom in the early 2000s, only surpassed by the collapse of Enron at the end of 2001. More recently, the early 2020s saw granular premia for exposure to technological innovators such as Tesla, AMD or Apple. Additionally, companies with business models particularly affected by the Covid-19 pandemic carried significant premia lately, examples being Netflix, Zoom, or Moderna. Therefore, investigating individual granular risk premia offers insight into additional sources of systemic risk and their compensation.

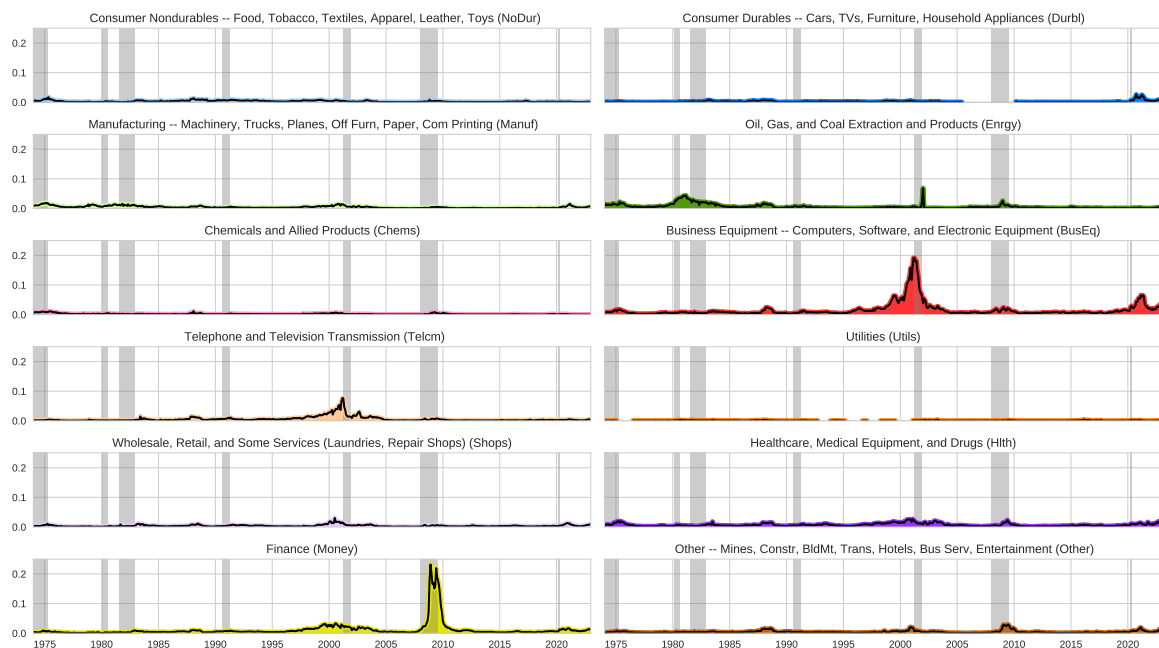


Figure 3.4: Sectoral sums of granular (asset-level) risk premia. Lines are interrupted if no asset from the given sector is sampled at the respective date. The shaded intervals indicate NBER recessions.

While the spillover network framework allows to examine premia at the most granular stock level, additional insight can be gained from aggregating premia into sectoral groups. I use the industry definitions of Fama and French (1997) to attribute each premium to one of 12 sectors and sum their values. Note that the resulting sectoral numbers should not be misunderstood as sectoral premia as earning them would imply unity exposures to each one of the sectoral constituents. Nevertheless, the aggregation provides a view into the most important drivers of risk over time. Figure 3.4 shows

the results for each sector. Consistent with widely held beliefs about the economic history of the last 50 years, the 1980s brought premia in the manufacturing and energy sectors, the 1990s and early 2000s had telecommunication and computer firms, and the late 2000s markets were dominated by financial firms. There are additional premia in other sectors that evolve simultaneously fueled by the network linkages at the time, and occasional bumps throughout the timeline that correspond to isolated events.

The idea of risk premia in the spillover network framework is straightforward: Shocks — common and granular — are redistributed in the cross-section via spillover linkages such that the structure of the spillover network determines which ones have a system-wide impact, and therefore command a risk premium. While the order of magnitude of the theoretically calculated granular premia is insufficient to explain historically observed return levels, this is also true for the common risk premium and in line with the well-known equity premium puzzle.¹³ Nevertheless, the relative size of the granular premia, when compared to the common premium, could indicate that granular risks may be an important mechanism to explain asset prices. This new element is particularly promising for the cross-section, where there is significant variation in the granular exposure tables Γ .

3.5 Granular risks and asset prices

Having collected supporting evidence for the suitability of the empirical approach to describe the return data in Section 3.4, in this section, I scrutinize the asset pricing implications of the granular framework. That is, if the spillover network framework lives up to its promise, it should be possible to validate its predictions on observed asset prices, both in-sample and out-of-sample. To that end, I first demonstrate the explanatory power of the spillover network model with regards to unexplained variation in linear factor models. Then, I test the predictions the model makes about expected returns both on the individual asset level cross-sectionally and on the aggregate market level in the time series.

Conjecture 4 *A substantial fraction of nonsystematic variance in reduced-form factor models stems from spillovers.*

Factor pricing models are reduced-form models of structural spillover networks. To see that this is true, consider the following transformation of (3.3) with additional

¹³Compare Mehra and Prescott (1985).

intercepts μ :

$$\begin{aligned} r_t &= \mu + L\tilde{\delta}_t + D^-r_t + \eta_t \\ &= \underbrace{\Gamma\mu}_{\alpha} + \underbrace{\Gamma L}_{\beta} \tilde{\delta}_t + \underbrace{\Gamma\eta_t}_{\varepsilon_t} \\ &= \alpha + \beta\tilde{\delta}_t + \varepsilon_t. \end{aligned}$$

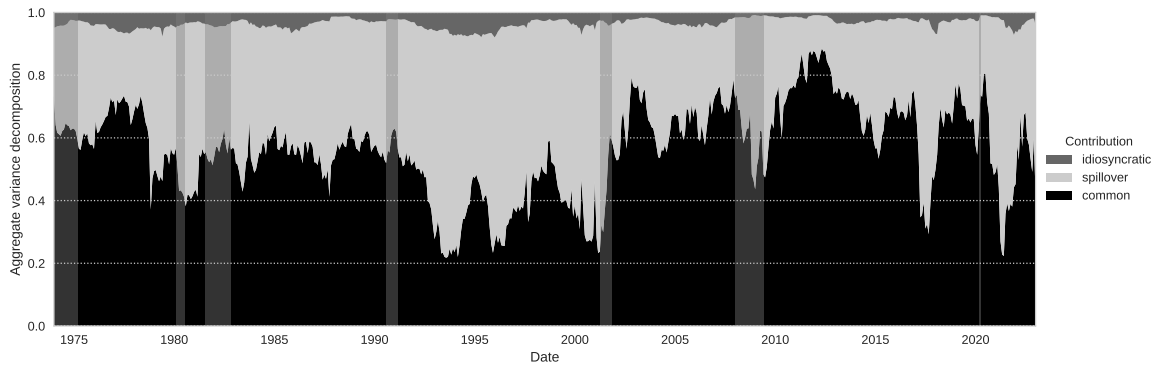
This identity holds for any choice of proxies for the common shocks δ_t , where the proxies act as common factors in the reduced-form formulation. Note also that reduced-form estimation yields unbiased factor model parameters α and β even in the presence of spillovers, which is easily confirmed by the fact that the transformed spillover network model estimates are identical to the reduced-form OLS factor model estimates. While factor models typically attribute a fraction of time-series variation in asset returns to “systematic” factor risk,¹⁴ there is often a larger part of unexplained variation termed as ‘idiosyncratic’ or ‘nonsystematic’ risk. However, given extensive evidence contradicting the independence of the reduced-form errors ε_t (compare for example Herskovic, Kelly, et al. (2016)), it is an open question how much variation is truly idiosyncratic and how much is transmitted from sources other than the common factors and innovations to the asset itself. A related comparison of reduced-form factor approaches with structural models with sectoral linkages can be found in Foerster et al. (2011) in the context of industrial production. In light of the structural formulation in the spillover network model, new insight can be derived from a decomposition of return variances. Let $\Gamma^+ = \text{diag}(\text{diag}(\Gamma))$ be the diagonal part of the matrix Γ . Then, the covariance matrix of the return vector Σ_r has the following decomposition:

$$\Sigma_r = \underbrace{\Gamma L \Sigma_\delta L' \Gamma'}_{\text{common}} + \underbrace{\Gamma \Sigma_\eta \Gamma' - \Gamma^+ \Sigma_\eta \Gamma'^+}_{\text{granular spillovers}} + \underbrace{\Gamma^+ \Sigma_\eta \Gamma'^+}_{\text{idiosyncratic}}. \quad (3.10)$$

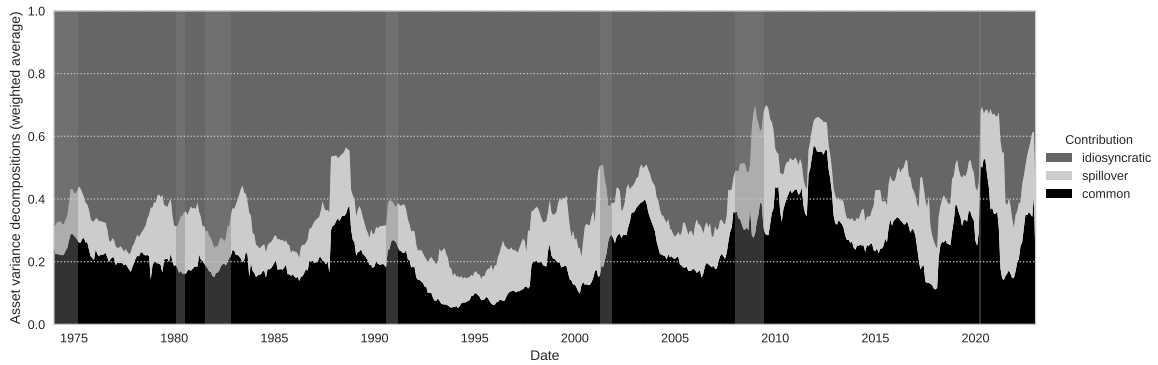
Here, the diagonality of Σ_η and Γ^+ ensures that the idiosyncratic components are now truly idiosyncratic, while all return correlation stems from common shocks or granular spillovers. A variance decomposition for the value-weighted portfolio of the included assets r_M follows from multiplying weights as $\sigma(r_M) = w' \Sigma_r w$.

Figure 3.5a shows decompositions of the value-weighted portfolio of the included assets over time. Unsurprisingly, common variation makes up the largest part of the variation. However, contrary to common belief, nonsystematic shocks also contribute about a third of aggregate movement on average, sometimes substantially more. Most of these contributions are a consequence of spillovers that induce correlated reduced-form residuals, while direct contributions of idiosyncratic components through the

¹⁴Estimates are typically in the range of 20% to 50%.



(a) Decomposition of aggregate index variance $\sigma(r_{\dot{M}}) = w' \Sigma_r w$.



(b) Weighted average of decompositions of individual asset return variances $w' \text{diag}(\Sigma_r)$.

Figure 3.5: Asset return variance decomposition over time. Variances are decomposed as in (3.10), where the aggregate return is weighted proportionally to the assets' weights in the market portfolio. The shaded intervals indicate NBER recessions.

concentration of weights are small. Turning to the asset level, Figure 3.5b shows that most return variances of individual assets are still largely dominated by independent idiosyncratic components, followed by the common component. Spillovers only account for about 15% of asset-level variation, on average, peaking at 41% at the heights of the Global Financial Crisis, with the spillover share of single assets sometimes being as high as 80%.

Observing these levels of spillover contributions to the variances of a value-weighted portfolio and individual assets suggests that spillovers could be a relevant mechanism in explanations of asset prices. If granular risks drive aggregate returns and the returns of a relevant fraction of individual assets simultaneously, such assets should earn additional compensation. Further, if the assets that originate spillovers exhibit distinct characteristics, expected returns should correlate with the characteristics. Then, the asset characteristics develop predictive power with regards to asset returns, showing up as anomalies with regards to reduced-form factor pricing models. Hence, granular risk premia could be a potential reason for ever new factors to appear in the literature (compare, for example, Feng et al. (2020) and Harvey and Liu (2021)), where the

returns on the newfound factors merely capture a changing set of granular risk premia instead of common risk premia.

Conjecture 5 *Assets with direct exposure to priced granular risks earn higher returns, on average.*

The next question I ask is: Does the spillover network framework as a whole, and granular risk in particular, help to price individual assets? (3.7) implies that the expected return on any asset is a linear function of its exposures to several premia for common and granular risks. More specifically, total expected returns in the model consist of a common component from multiplying common risk premia λ_t with the asset's common risk exposures β_i , and a granular component from multiplying granular risk premia ξ_t with the asset's granular risk exposures γ_i :

$$\underbrace{\mathbb{E}[R_{i,t}] - R^f}_{\text{total}} = \underbrace{\lambda_t \beta_i}_{\text{common}} + \underbrace{\xi_t \gamma_i}_{\text{granular}}. \quad (3.11)$$

Here, λ_t and ξ_t are prices of risks, and β_i and γ_i denote the quantities of the respective risks pertinent to asset i .

To test this relationship empirically, I first check its predictive power in the cross-section. More specifically, do assets with exposure to relevant sources of risk earn additional compensation? Given the calculated risk premia from Section 3.4, I use the estimated exposures $\hat{\beta} = \hat{\Gamma} \hat{L}$ and $\hat{\Gamma}$ to calculate the components of the model-implied expected returns as in (3.11). I relate the calculated expected returns to observed returns in the respective one-year sample, and out-of-sample in the year following the estimation date. To that end, I conduct one-way and two-way portfolio sorts of the 100 assets in each sample.

	in-sample			out-of-sample		
	total	common	granular	total	common	granular
Low	0.0825	0.0842	0.0979	0.0864	0.0854	0.0846
2	0.0980	0.0911	0.1020	0.0825	0.0827	0.0825
3	0.1057	0.1107	0.1087	0.0793	0.0805	0.0769
4	0.1243	0.1203	0.1276	0.0703	0.0762	0.0798
High	0.1717	0.1758	0.1460	0.0897	0.0834	0.0844
H-L	0.0891*** (3.9406)	0.0916*** (3.9530)	0.0481*** (2.9049)	0.0032 (0.2119)	-0.0020 (-0.1328)	-0.0002 (-0.0195)

Table 3.3: One-way cross-sectional portfolio sorts. The 100 assets contained in each sample are sorted into five equally weighted portfolios according to the model's expected return predictions as in (3.11). Observations stem from 1-year end-of-month samples spanning from Dec-1973 to Dec-2022. Measured outcomes of the portfolios are one-year in-sample and out-of-sample returns ending and starting at the respective sampling date. t-statistics are calculated from Newey-West standard errors with $T^{1/4}$ lags.

Table 3.3 presents the results of one-way portfolio sorts on total, common, and granular expected returns. The in-sample analysis shows that all three are significantly related to relative outperformance. That is, for all three sorting variables, assets with higher expected returns under the model indeed exhibit returns in excess of those with lower expected returns. As a novel result, this differential also extends to the granular component of the expected return, while the return premium on common risk exposure is in line with standard factor model frameworks. However, none of these differences is confirmed in the subsequent year, which hints at the elusive nature of short-lived return premia. Note that having restricted the samples to only 100 assets may impair the power of the out-of-sample results in addition to the instability of loadings and spillover linkages.

		in-sample				out-of-sample			
		granular				granular			
		1	2	3	3-1	1	2	3	3-1
<i>Panel A: Simultaneous sort of common & granular</i>									
common	1	0.0818	0.0856	0.0926	0.0119 (1.0885)	1	0.0918	0.0797	0.0785 -0.0157 (-1.5934)
	2	0.1008	0.1030	0.1259	0.0250** (2.0138)	2	0.0813	0.0803	0.0828 0.0015 (0.1600)
	3	0.1318	0.1434	0.1774	0.0512*** (3.3791)	3	0.0832	0.0642	0.0952 0.0172 (1.3302)
	3-1	0.0506** (2.5339)	0.0579*** (2.9010)	0.0890*** (3.2031)		3-1	-0.0099 (-0.6565)	-0.0155 (-1.1535)	0.0246 (1.2112)
<i>Panel B: Sequential sort on common first and granular second</i>									
common	1	0.0857	0.0848	0.0869	0.0012 (0.1483)	1	0.0899	0.0828	0.0809 -0.0090 (-1.0326)
	2	0.1003	0.1009	0.1255	0.0253** (2.0915)	2	0.0784	0.0788	0.0825 0.0041 (0.4415)
	3	0.1339	0.1541	0.1786	0.0447** (2.2014)	3	0.0637	0.0786	0.0983 0.0346** (2.1919)
	3-1	0.0481** (2.2817)	0.0693*** (2.7312)	0.0917*** (3.1209)		3-1	-0.0261* (-1.7216)	-0.0042 (-0.2811)	0.0175 (0.7925)
<i>Panel C: Sequential sort on granular first and common second</i>									
common	1	0.0842	0.0857	0.0995	0.0153 (1.2985)	1	0.0904	0.0823	0.0760 -0.0143 (-1.3421)
	2	0.0946	0.0978	0.1297	0.0351** (1.9723)	2	0.0821	0.0765	0.0859 0.0038 (0.2759)
	3	0.1217	0.1421	0.1957	0.0740*** (2.7450)	3	0.0840	0.0693	0.0875 0.0035 (0.2131)
	3-1	0.0375*** (2.5874)	0.0565*** (3.2948)	0.0962*** (3.0122)		3-1	-0.0064 (-0.5000)	-0.0130 (-1.0021)	0.0114 (0.5824)

Table 3.4: Two-way cross-sectional portfolio sorts. The 100 assets contained in each sample are sorted into nine equally weighted portfolios according to the model's expected common and granular return predictions as in (3.11). Observations stem from 1-year end-of-month samples spanning from Dec-1973 to Dec-2022. Measured outcomes of the portfolios are one-year in-sample and out-of-sample returns ending and starting at the respective sampling date. t-statistics are calculated from Newey-West standard errors with $T^{1/4}$ lags.

Next, given the bisection of the expected return in (3.11), I consider two-way portfolio sorts on the common and granular component, which allows to isolate the predictions of one component while factoring out the potentially heavy correlation with the other component. Table 3.4 contains the results of three variations of such sorts, in three-by-three portfolios, again in-sample and out-of-sample in the subsequent year. First, the simultaneous sort in Panel A confirms the results of the one-way sorts in that it finds significant differences in both directions. However, the difference on the granular return component is particularly strong for assets that also load heavily on

the common risk, and insignificant for low-risk assets. Moving to sequential sorts in Panels B and C, common expected returns consistently predict in-sample returns in all granular groups. Conversely, the effect of granular sorts appears mostly for the high common expected return group. When fixing common return expectations first in Panel B, return predictions on the high group appear to extend also to the out-of-sample period, while all other differences are insignificant or even invert for the common return expectations.

Overall, there are some hints to suggest granular contributions to expected returns could be a relevant part of asset pricing models. While the majority of findings does not persist out-of-sample in the time series direction, this is also true for the common component which is widely accepted in the asset pricing literature. The portfolio sort analysis results suggest that the relation of granular expected returns to actual returns holds at least in-sample. Hence, it appears that systemically relevant granular sources of risk are priced into the cross-section of assets in the respective sample. Although this analysis has issues in terms of sampling bias and the limited size of the sampled cross-sections, granular risk premia remain a promising hypothesis to explain cross-sectional variation in asset prices.

Conjecture 6 *Granular risks explain time series variation in aggregate returns.*

The expected return formulation in (3.11) has to hold for any asset in the economy, including value-weighted portfolios such as the market aggregate. The question of what drives expected aggregate returns is closely related to inquiries about the equity risk premium. Therefore, I want to understand how expectations under the model relate to returns at different horizons, and if granular risk premia contribute to mean aggregate returns. To derive return expectations for the aggregate portfolio, I use the calculated premia from Section 3.4 and combine them with the estimated exposures. Specifically, consider the structural expression for the market return in (3.5). Estimates for the loadings on both components are given by a weighted average of individual assets' loadings as $\hat{\beta}_M = w'\hat{\Gamma}\hat{L}$ and $\hat{\gamma}_M = w'\hat{\Gamma}$. With each sample comprising of 100 stocks, I confine the exposure and risk premia calculations to the same sample of assets. Then, I use time series regressions to check how calculated expected aggregate returns relate to the returns of the value-weighted portfolio of assets included in the sample, and to the returns of the value-weighted CRSP index as a broader aggregate.

Table 3.5 presents the time series regression results in four panels.¹⁵ First, Panel A testifies the overall ability of the model to match aggregate returns, where a large

¹⁵Figure 3.13 in Appendix 3.A.5 shows the evolution of the included variables over time. The common and granular return expectations are correlated with a coefficient of 81.6%. The two value-weighted aggregates are 91.4% correlated.

r^M : Value-weighted portfolio of 100 sampled assets														
	in-sample		out-of-sample				in-sample		out-of-sample					
	$r^M_{t-1Y:t}$	r^M_{t+1M}	r^M_{t+3M}	r^M_{t+6M}	r^M_{t+9M}	r^M_{t+12M}	r^M_{t+24M}	$r^M_{t-1Y:t}$	r^M_{t+1M}	r^M_{t+3M}	r^M_{t+6M}	r^M_{t+9M}	r^M_{t+12M}	r^M_{t+24M}
<i>Panel A: $r^M_t = \theta_0 + \theta_1(\lambda_t\beta_M + \xi_t\gamma_M)$</i>														
constant	0.1711*** (10.2891)	0.0085*** (3.7441)	0.0245*** (4.2041)	0.0505*** (5.1163)	0.0807*** (6.1044)	0.1180*** (7.3141)	0.2933*** (9.8355)	0.1444*** (10.5971)	0.0060** (2.4680)	0.0167*** (2.7063)	0.0339*** (3.3754)	0.0547*** (4.1932)	0.0824*** (5.2985)	0.1956*** (7.6926)
total	-2.5171* (-1.9276)	2.6741 (1.0917)	3.2336** (1.6961)	3.2336** (2.4941)	2.9114** (2.5464)	2.1511** (1.9785)	0.5256 (0.5641)	-5.6705*** (-5.5608)	1.5643 (0.5442)	2.2151 (1.0179)	2.3644* (1.6480)	2.1043* (1.8169)	1.4108 (1.3551)	0.3074 (0.3036)
R-squared	0.0293	0.0033	0.0151	0.0291	0.0345	0.0254	0.0025	0.1647	0.0011	0.0066	0.0146	0.0176	0.0111	0.0012
T	589	588	586	583	580	577	565	589	588	586	583	580	577	565
<i>Panel B: $r^M_t = \theta_0 + \theta_1(\lambda_t\beta_M)$</i>														
constant	0.1815*** (10.7908)	0.0075*** (3.3196)	0.0225*** (3.8527)	0.0468*** (4.6773)	0.0757*** (5.5283)	0.1099*** (6.4487)	0.2788*** (9.1019)	0.1523*** (11.0460)	0.0049** (2.0330)	0.0142** (2.3236)	0.0290*** (2.8967)	0.0480*** (3.6360)	0.0717*** (4.4956)	0.1806*** (7.0684)
common	-6.4087*** (-2.8670)	6.8394 (1.5927)	7.2364** (2.2584)	7.2045*** (3.2590)	6.5037*** (3.0940)	5.5624*** (2.7722)	2.2961 (1.6410)	-11.6972*** (-5.9612)	5.0853 (1.0041)	5.8498 (1.5815)	6.0947*** (2.5955)	5.4751*** (2.7735)	4.5449** (2.5693)	1.9542 (1.3603)
R-squared	0.0549	0.0063	0.0229	0.0417	0.0493	0.0482	0.0132	0.1989	0.0034	0.0134	0.0280	0.0340	0.0326	0.0133
T	589	588	586	583	580	577	565	589	588	586	583	580	577	565
<i>Panel C: $r^M_t = \theta_0 + \theta_1(\xi_t\gamma_M)$</i>														
constant	0.1577*** (9.5750)	0.0097*** (4.5626)	0.0278*** (4.9737)	0.0566*** (5.7760)	0.0890*** (6.8680)	0.1304*** (8.2827)	0.3078*** (10.6581)	0.1275*** (9.0844)	0.0071*** (3.1863)	0.0199*** (3.3890)	0.0402*** (3.9930)	0.0633*** (4.8172)	0.0940*** (6.0594)	0.2100*** (8.4376)
granular	-2.5537 (-0.8419)	2.5888 (0.5437)	4.0482 (1.0775)	4.3773 (1.4756)	3.9596 (1.5625)	2.1483 (0.9049)	-0.5187 (-0.2216)	-8.8114*** (-3.8005)	0.3717 (0.0680)	2.0012 (0.4647)	2.3367 (0.7205)	2.0471 (0.7592)	0.4569 (0.1884)	-1.0033 (-0.3965)
R-squared	0.0080	0.0008	0.0067	0.0145	0.0172	0.0068	0.0006	0.1061	0.0000	0.0015	0.0039	0.0045	0.0003	0.0032
T	589	588	586	583	580	577	565	589	588	586	583	580	577	565
<i>Panel D: $r^M_t = \theta_0 + \theta_1(\lambda_t\beta_M) + \theta_2(\xi_t\gamma_M)$</i>														
constant	0.1808*** (10.7315)	0.0076*** (3.3631)	0.0226*** (3.8677)	0.0470*** (4.6586)	0.0760*** (5.5140)	0.1105*** (6.4495)	0.2788*** (8.8589)	0.1521*** (11.0437)	0.0050** (2.0789)	0.0144** (2.3492)	0.0293*** (2.8944)	0.0484*** (3.6287)	0.0725*** (4.5041)	0.1806*** (6.7630)
common	-13.3449*** (-2.9113)	14.3450** (2.1899)	12.0587** (2.5762)	11.1760*** (2.7775)	10.0872*** (2.7507)	11.5723*** (3.8031)	8.3613*** (3.5027)	-14.1662*** (-3.4826)	14.3454** (2.0301)	12.7550*** (2.6400)	12.6864*** (3.1301)	11.5444*** (3.2557)	12.5395*** (4.3948)	8.4816*** (3.7102)
granular	8.6843* (1.6854)	-9.4854 (-1.3745)	-6.0946 (-1.2754)	-5.0173 (-1.2622)	-4.5235 (-1.3237)	-7.5855*** (-2.6490)	-7.7204*** (-2.9055)	3.1183 (0.6809)	-11.7029 (-1.6218)	-8.7273* (-1.7758)	-8.3275** (-2.1439)	-7.6613** (-2.2851)	-10.0904*** (-3.6806)	-8.3087*** (-3.1076)
R-squared	0.0859	0.0101	0.0280	0.0481	0.0567	0.0763	0.0570	0.2033	0.0092	0.0228	0.0445	0.0549	0.0831	0.0843
T	589	588	586	583	580	577	565	589	588	586	583	580	577	565

Table 3.5: Regressions of rolling return on the sample and CRSP value-weighted portfolios on return expectations, contemporaneously (in-sample), and for subsequent windows of 1, 3, 6, 9, 12, and 24 months (out-of-sample). Observations stem from 1-year end-of-month samples spanning from Dec-1973 to Dec-2022. The independent variables are the expected market returns under the spillover network model as in (3.7), and risk exposures of the aggregate portfolio are according to (3.5). t-statistics are calculated from Newey-West standard errors with $T^{1/4}$ lags.

expected return has a negative impact on in-sample returns followed by larger returns out-of-sample at horizons around 9 months. Return forecasts are strongest for the sample portfolio, but there is evidence that they also contribute to the forecasts of the CRSP index. Second, Panels B and C show that most of the model's power comes from common risk premia, while granular return expectations share a highly significant relation only with in-sample returns of the broad aggregate. Finally, Panel D confirms that common premia dominate granular premia when it comes to explaining aggregates, whereas the share of explained variation benefits from having a granular component in the regression.

The fact that aggregate expected returns are primarily driven by common return premia is unsurprising as the aggregate primarily loads on common shocks.¹⁶ Nevertheless, adding granular return expectations often improves forecasts such that they should not be dismissed in modelling expected returns. Instead, the weak relation with aggregate portfolio returns hints that it may be difficult to reduce the complex interactions of the spillover network model into sparse representations via aggregated factors. The propagation mechanism in my framework implies a multitude of influences being at play at the same time which makes it hard to isolate their effects. Therefore, despite granular risks not appearing to be major drivers of the equity risk premium itself, the spillover framework's descriptive power may lie primarily in additional return differentials induced by direct network exposures between individual assets. Lastly, it is striking that the coefficient values in Table 3.5 are considerably larger than one, consistent with the Equity Premium Puzzle.

Finally, it is worth mentioning some possible extensions of the analysis to investigate the framework's capabilities out-of-sample also cross-sectionally. Here, it is tempting to consider a two-factor approximation of the SDF in (3.6), where the second factor is an index of granular innovations.¹⁷ One could argue that such a 'granular factor' provides a way to probe the explanatory power of granular risk premia even on assets that are not included in the samples. However, such an approach contradicts the notion of the spillover network as proposed in this paper. If the spillover network framework holds, the existence of the network implies exposures between individual assets rather than towards a shared factor that is common to all assets. Even if exposure to the granular drivers of aggregate wealth could be measured through exposure to a summarizing factor, the short-lived importance of individual granular risks would necessitate a highly dynamic approach with time-varying parameters to adequately capture elusive

¹⁶Compare Figure 3.12 in Appendix 3.A.5. On average, the portfolio of included assets has an exposure to common shocks of 1.06, whereas the period-wise maximum granular exposure is 0.15.

¹⁷Malevergne et al. (2009) propose a similar two-factor model based on firm-size concentration, that is, without considering a spillover network.

exposures. Furthermore, it is questionable if my empirical approach is capable to convincingly identify the granular shocks η_t , which are necessary to construct the second addend in (3.6). Hence, any attempt at a two-factor model to represent the spillover network framework would come at the risk of oversimplification.

3.6 Conclusions

In this paper, I propose an extension to asset pricing models that considers structural interactions between assets instead of only common factors. The resulting framework makes predictions about compensation for asset-level risks, which I term ‘granular risk premia’. I then use estimated FEVD spillover networks to approximate structural linkages in the cross-section of equity returns and calculate corresponding values for the granular risk premia. Such premia only occur temporarily, but sometimes have economically meaningful magnitude. As a complication, the short-lived nature of risk premia and exposures necessitates a highly dynamic approach to investigate connected hypotheses. However, in-sample results suggest granular risk premia could be an overlooked determinant of the returns of individual assets.

Despite the successes of the spillover network framework in offering a new perspective on asset markets, there are some limitations of the conducted empirical analysis. On the one hand, my structural estimates are imprecise due to the large-dimensional nature of the estimations, and incomplete due to the restriction of the cross-section to 100 assets in each period, where the latter step additionally introduces a sampling bias. On the other hand, approximation errors may lead to the limited external validity of my empirical results. An additional limitation arises theoretically, as the presented framework only includes common and granular asset-level shocks, but not on other aggregation levels such as sectoral shocks. Nonetheless, this analysis generalizes existing frameworks and captures novel dynamics in the analyzed samples.

My findings are important to investors and policymakers alike: If assets are connected through directional links, investors should consider inter-asset dependencies in their asset selection and in portfolio choice. That is, if reduced-form alpha arises due to compensated granular risk, strategies that aim at extracting alpha may unintentionally load on sources of systemic risk. Second, the mean-variance efficient portfolio changes when replacing the variance-covariance matrix with a directed spillover matrix, and therefore alters portfolio optimization. Similarly, the hazard of granular events turning into systemic distress mandates regulators to monitor the most influential assets when implementing prudential policy in the financial system. Further, the connection of granular risk premia to excessive concentration of market value lends justification to

antitrust bodies being wary of individual firms getting too large, even outside the heavily connected financial sector.

Granular risk premia also offer new insights for researchers in asset pricing. My results show that the regnant paradigm of the linear factor pricing model has limitations due to the negligence of spillovers that cause reduced-form residuals to be correlated. While the literature proposes to eliminate such correlations by adding to the list of candidate factors, the presence of dynamically changing granular effects suggests that finding a universal factor pricing model could be an unsolvable problem. Instead, finding structural approximations for asset markets could be a promising venue for future research. In any case, I hope my results encourage researchers to think more carefully about granular shocks in the context of asset prices.

3.A Appendices

3.A.1 Derivation of beta representation

Combining the fundamental asset pricing equation $1 = \mathbb{E}[m_t R_{i,t}] \forall i$ with the covariance identity $Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$, I obtain

$$1 = Cov(m_t, R_{i,t}) + \mathbb{E}[m_t]\mathbb{E}[R_{i,t}].$$

With $R^f = 1/\mathbb{E}[m_t]$, the equation can be transformed into

$$\mathbb{E}[R_{i,t}] - R^f = -R^f Cov(m_t, R_{i,t}) \quad (3.12)$$

That is, any excess return in the economy is defined by its covariance with the stochastic discount factor. Given the definition of the SDF in (3.6), terms for each element in the innovation vectors can be separated into distinct terms:

$$\mathbb{E}[R_{i,t}] - R^f = -R^f Cov(a + bw'\Gamma L\delta_t + bw'\Gamma\eta_t, R_{i,t}) \quad (3.13)$$

$$= -R^f bw'\Gamma L Cov(\delta_t, R_{i,t}) - R^f bw'\Gamma Cov(\eta_t, R_{i,t}) \quad (3.14)$$

Expanding this with $\Sigma_x \Sigma_x^{-1}$, where Σ_x is a matrix with the variances of x on its diagonal, I obtain the well-known beta representation of the excess return:

$$\mathbb{E}[R_{i,t}] - R^f = -R^f bw'\Gamma L \Sigma_{\delta_t}^{-1} Cov(\delta_t, R_{i,t}) - R^f bw'\Gamma \Sigma_{\eta_t}^{-1} Cov(\eta_t, R_{i,t}) \quad (3.15)$$

$$= \lambda_t \beta_i + \xi_t \Gamma'_i \quad (3.16)$$

The last step uses the fact that $\Sigma_{\delta_t}^{-1} Cov(\delta_t, R_{i,t})$ and $\Sigma_{\eta_t}^{-1} Cov(\eta_t, R_{i,t})$ are coefficients in the structural equation (3.4), where $\beta_i = [\Gamma L]_i$.

3.A.2 Network construction details

I start by modeling daily log-volatilities s_t as a VAR(1) process

$$s_t = \mu + A_1 s_{t-1} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma_u), \quad (3.17)$$

and rewrite it in its VMA(∞)-form

$$s_t = \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i}, \quad \Phi_i = A_1^i. \quad (3.18)$$

Here, the components of u_t include the contemporaneous dependencies of s_t , such that u_t is not an orthogonal shock vector. Hence, as in Brunner and Hipp (2023), I use the multivariate generalized framework based on Pesaran and Shin (1998) to construct impulse responses $\Psi_h = \mathbb{E}[s_{t+h} | \{u_{t,j} = \sigma_{jj}\}_{j=1}^N]$ at horizon h to one standard deviation multivariate generalized impulses

$$\Psi_h = \Phi_h \Sigma_u^{1/2}, \quad (3.19)$$

where $\Sigma_u^{1/2}$ is the matrix square root of Σ_u .¹⁸ Next, the forecast error variances until horizon H follow as the sum element-wise squared impulse responses:

$$FEV(H) = \sum_{h=0}^{H-1} \Psi_h^2. \quad (3.20)$$

Finally, the forecast error variance decomposition FEVD contains relative contributions to the row-wise total forecast error variances:

$$FEVD(H) = \frac{FEV(H)}{FEV(H) \mathbf{1}_N}. \quad (3.21)$$

I denote the FEVD at the horizon of 21 days (the average number of business days in a month) as $\hat{D} = FEVD(21)$. Entries in this table show the fraction of forecasting uncertainty attributable to each innovation element in u_t . Hence, the tables reflect the relative strengths of the directed information flow for a given asset pair, and are suitable proxies for the spillover tables.

¹⁸The use of the square root decomposition $\Sigma_u^{1/2}$ instead of of a generalized impulse $\Sigma_u \text{diag}(\Sigma_u)^{-1/2}$ has shown advantageous in simulations of settings with large-dimensionality.

3.A.3 Large-dimensional VAR estimation details

First, I calculate daily log-volatilities following the extreme value method of Parkinson (1980):

$$s_{i,t} = \ln \left(\sqrt{0.3607} (p_{i,t}^{high} - p_{i,t}^{low}) \right). \quad (3.22)$$

$p_{i,t}^{high}$ and $p_{i,t}^{low}$ are the natural logarithms of the intraday high and low prices, respectively. Next, I define a VAR(1) process with multivariate Gaussian error term for this data while controlling for common innovations through an equally weighted index of the whole CRSP universe f_t :

$$s_t = \mu + b f_t + A_1 s_{t-1} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma). \quad (3.23)$$

I estimate regression coefficients by means of the adaptive elastic net estimator of Zou and Zhang (2009):

$$(\hat{\mu}, \hat{b}, \hat{A}) = \underset{(\mu, b, A)}{\operatorname{argmin}} \left[\frac{1}{2NT} \sum_{i=1}^N \sum_{t=1}^T \left(s_{i,t} - \left(\mu_i + \sum_{k=1}^K b_{i,k} f_{k,t} + \sum_{j=1}^N a_{i,j} s_{j,t-1} \right) \right)^2 + \lambda \sum_{i=1}^N \sum_{j=1}^N \hat{m}_{i,j} (\kappa |a_{i,j}| + (1 - \kappa) \frac{1}{2} a_{i,j}^2) \right]. \quad (3.24)$$

Here, the hyperparameters λ and κ are chosen via 12-fold cross-validation, and $\hat{m}_{i,j} = |\hat{d}_{i,j}^{ENet} + \frac{1}{NT}|^{-1}$ is the absolute inverse of the first-stage estimates via an Elastic Net with hyperparameters $\kappa = 0.0001$ and with a cross-validated λ . Note that the estimator needs to be applied to standardized data, such that reversal of the standardization is necessary after the estimation.

Now, given the estimates of the model parameters, I calculate the model residuals u_t and estimate their covariance matrix $\Sigma_u = \Theta_u^{-1}$ via the graphical LASSO estimator of Banerjee et al. (2008) and Friedman et al. (2008):

$$\hat{\Theta}_u = \underset{\Theta_u > 0 \in \mathbb{R}^{N \times N}}{\operatorname{argmax}} \left[\ln |\Theta_u| - \operatorname{trace} \left(\left(\frac{1}{T-2} \sum_{t=2}^T u_t u_t' \right) \Theta_u \right) - \rho \sum_{i \neq j} |\theta_{i,j}| \right] \quad (3.25)$$

Here, Θ_u is optimized over the set of all non-negative definite matrices in $\mathbb{R}^{N \times N}$. Again, I employ 12-fold cross-validation to choose the hyperparameter ρ .

3.A.4 Maximum likelihood estimation details

Required are estimates for the parameters $\theta = \{\mu, L, c, \Sigma_\eta\}$ in the following model:

$$r_t = \mu + L\tilde{\delta}_t + c\tilde{D}^- r_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \Sigma_\eta), \quad \forall t = 1, \dots, T \quad (3.26)$$

Provided observations for the return vector r_t and some observed proxies $\tilde{\delta}_t$ and $\tilde{D}^-$ for the common shocks and for the spillover matrix, assume an uncorrelated and normally distributed *i.i.d.* disturbance term $\eta_t \sim \mathcal{N}(0, \Sigma_\eta)$, where Σ_η is a diagonal matrix. Collecting terms in (3.26):

$$(I_N - c\tilde{D}^-)r_t = \mu + L\tilde{\delta}_t + \eta_t \quad (3.27)$$

Now let $A = I_N - c\tilde{D}^-$ and collect the return observations in the $N \times T$ matrix $R := [r_1, \dots, r_T]$. Similarly, let $\Delta := [\tilde{\delta}_1, \dots, \tilde{\delta}_T]$ and $\mathbf{1}_T$ be a vector of ones with length T . Given some candidate values for θ , calculate the errors $E := [\eta_1, \dots, \eta_T]$ and the diagonal error covariance Σ_η :

$$E = AR - \mu\mathbf{1}_T' - L\Delta \quad (3.28)$$

$$\Sigma_\eta = T^{-1} \text{diag}(\text{diag}(EE')) \quad (3.29)$$

With $\Sigma_\delta = T^{-1}\Delta\Delta'$, the model has the following variance for an observation r_t :

$$\Sigma_r = A^{-1}L\Sigma_\delta L'A'^{-1} + A^{-1}\Sigma_\eta A'^{-1} \quad (3.30)$$

The log-likelihood of r_t is then given by the log-density of the multivariate normal distribution:

$$\ell(r_t|\theta, \tilde{\delta}_t, \tilde{D}^-) = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln |A^{-1}\Sigma_\eta A'^{-1}| - \frac{1}{2} \eta_t' \Sigma_\eta^{-1} \eta_t \quad (3.31)$$

Following Lütkepohl (2005), the likelihood function for a sample $r_1, \dots, r_T$ is given by:

$$\mathcal{L}(\theta|r_t, \tilde{\delta}_t, \tilde{D}^-) = \sum_{t=1}^T \ell(r_t|\theta, \tilde{\delta}_t, \tilde{D}^-) \quad (3.32)$$

$$= -\frac{NT}{2} \ln(2\pi) - \frac{T}{2} \ln |A^{-1}\Sigma_\eta A'^{-1}| - \frac{1}{2} \text{trace}\{E'\Sigma_\eta^{-1}E\} \quad (3.33)$$

To ensure stability of the estimated process, the eigenvalues of the companion matrix $c\tilde{D}^-$ need to lay within the unit circle. Equivalently, valid solutions for the coefficient c are strictly smaller than $\bar{c} = 1/\max|\lambda_{\tilde{D}^-}|$, where $\lambda_{\tilde{D}^-}$ denotes the eigenvalues of $\tilde{D}^-$. To achieve a stable solution I therefore impose the condition $c \leq 0.9999\bar{c}$ during the optimization. However, throughout this analysis, the constraint does not bind (compare

Figure 3.7 in Appendix 3.A.5). Finally, the maximum likelihood estimates can be obtained as:

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmin}} \mathcal{L}(\theta | r_t, \tilde{\delta}_t, \tilde{D}^-) \quad \text{s.t.} \quad c < \bar{c}. \quad (3.34)$$

A likelihood-ratio test of the form

$$\lambda_{LR} = T(\ln |\hat{A}^{-1} \hat{\Sigma}_\eta \hat{A}'^{-1}| - \ln |\hat{\Sigma}_\varepsilon|) \quad (3.35)$$

can be used to test for overidentifying restrictions, where $\hat{\Sigma}_\varepsilon$ is the sample covariance matrix of the reduced-form errors $\hat{\varepsilon}_t = r_t - \hat{\mu} - \hat{L}\tilde{\delta}_t$. The null hypothesis that the restrictions are valid is rejected if λ_{LR} exceeds a threshold based on its asymptotic χ^2 -distribution with $N(N+1)/2 - 2N - NK - 1$ degrees of freedom. Depending on the specific context, however, the asymptotic theory underlying this test may fail due to the dimensionality of N .

3.A.5 Auxiliary plots

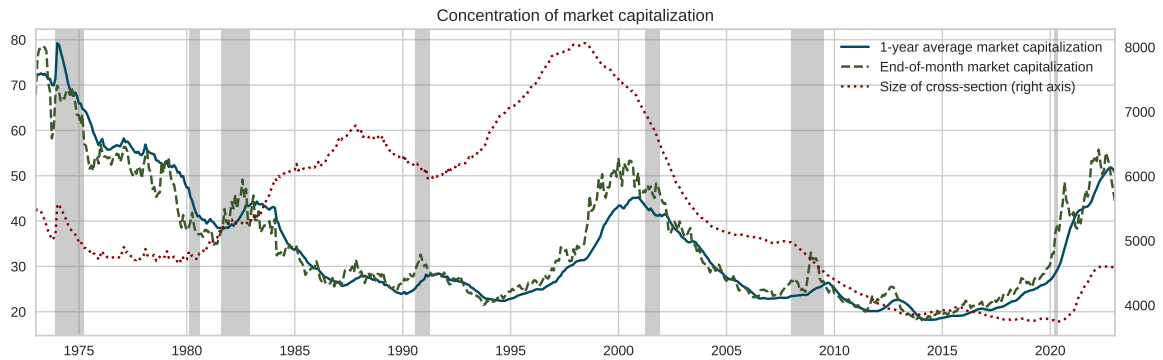


Figure 3.6: Concentration of market capitalizations in the CRSP universe. To account for the varying size of the cross-section, concentration values are calculated as the ratio of the Herfindahl index of the observed market capitalizations divided by a Herfindahl index with equal weighting. The shaded intervals indicate NBER recessions.

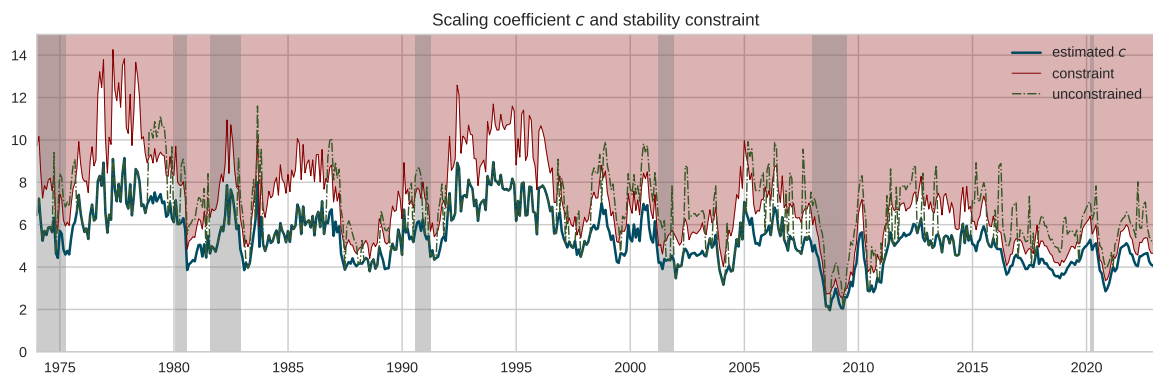


Figure 3.7: Estimate of the scaling coefficient c for each sample. The constraint is the inverse of the largest eigenvalue of the proxy matrix $\tilde{D}$ in modulus. The shaded intervals indicate NBER recessions.

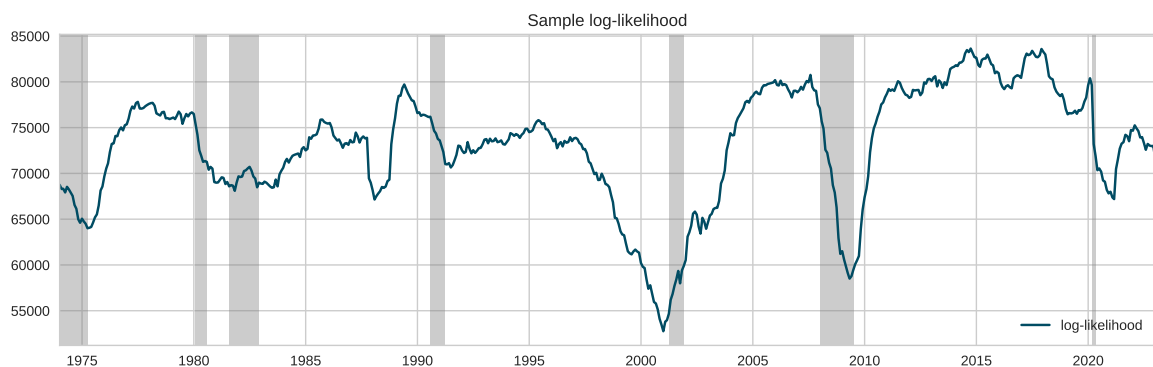


Figure 3.8: Log-likelihood for each sample. The shaded intervals indicate NBER recessions.

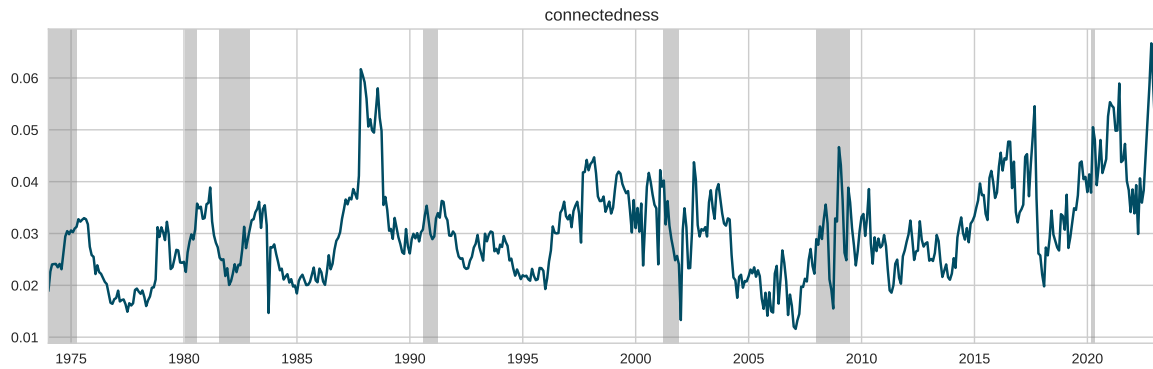


Figure 3.9: Γ connectedness for each sample. Connectedness is the average of all off-diagonal elements in Γ . The shaded intervals indicate NBER recessions.

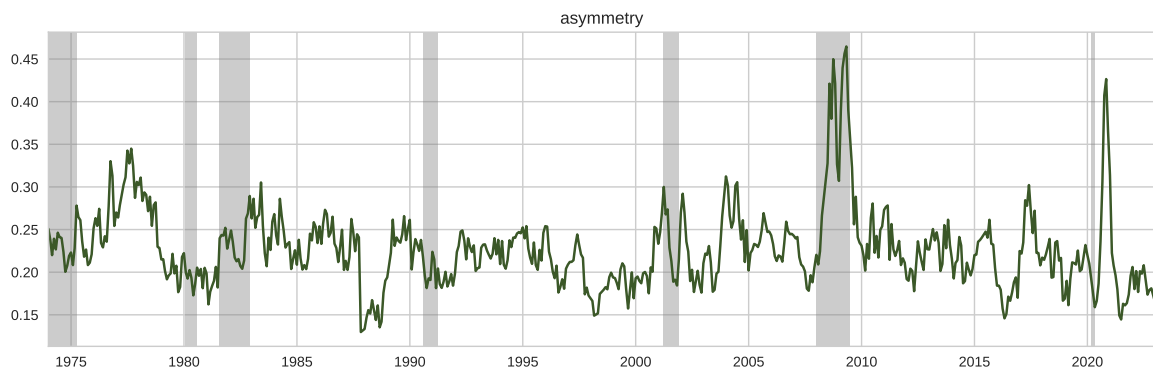


Figure 3.10: Γ asymmetry for each sample. Asymmetry is the fraction of all off-diagonal elements in Γ that differs from the symmetric average $(\Gamma + \Gamma')/2$. The shaded intervals indicate NBER recessions.

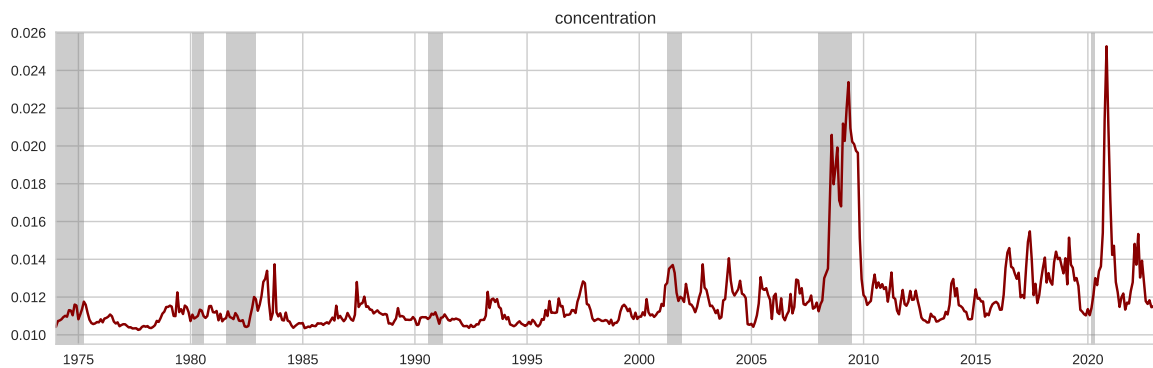


Figure 3.11: Γ concentration for each sample. Concentration is the Herfindahl index over the column sums of Γ . The shaded intervals indicate NBER recessions.

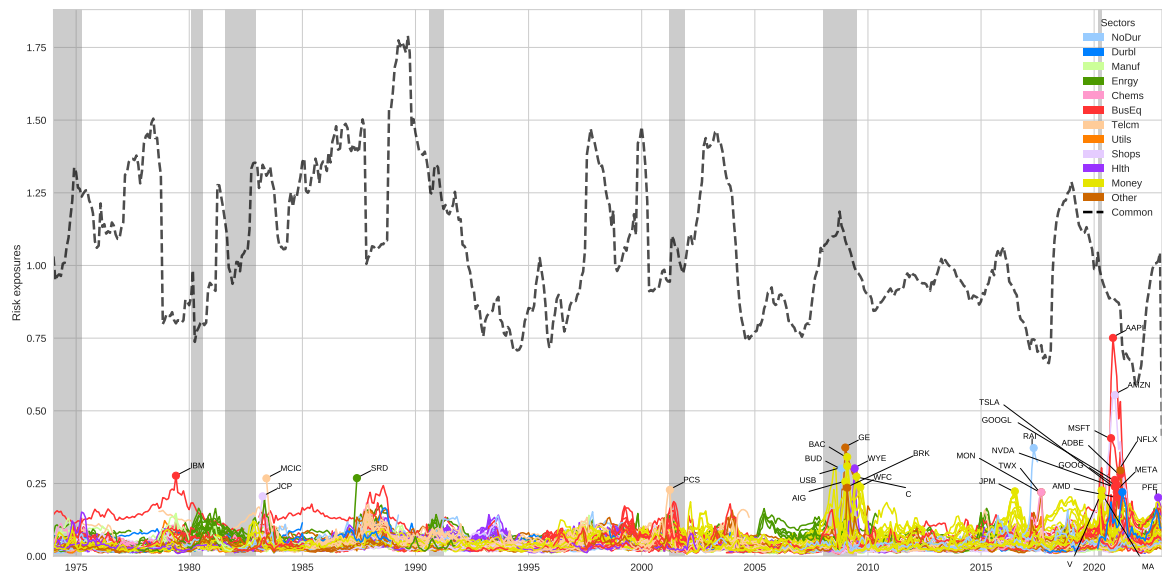


Figure 3.12: Risk exposures of the aggregate portfolio. Exposures are calculated from (3.5), where the cross-section consists of the 100 included assets in the respective sample. Granular exposures are colored according to the Fama and French (1997) industry definitions. The shaded intervals indicate NBER recessions.

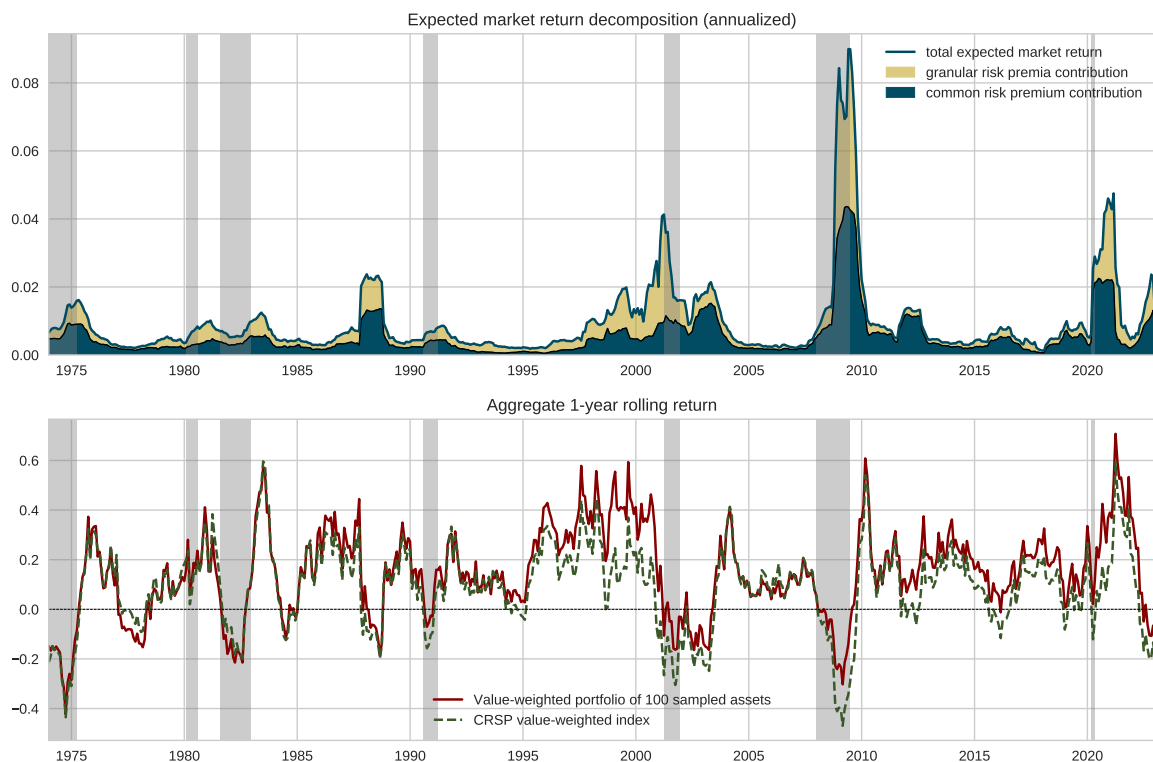


Figure 3.13: Decomposition of the expected aggregate return into common and granular risk premia according to (3.7), and rolling return of the CRSP value-weighted index over the horizon of 1 year. The shaded intervals indicate NBER recessions.

3.B Bibliography

- Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). “The network origins of aggregate fluctuations.” In: *Econometrica* 80.5, pp. 1977–2016.
- Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015). “Systemic risk and stability in financial networks.” In: *American Economic Review* 105.2, pp. 564–608.
- Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2017). “Microeconomic origins of macroeconomic tail risks.” In: *American Economic Review* 107.1, pp. 54–108.
- Ahern, K. R. (2013). “Network centrality and the cross section of stock returns.” In: *Available at SSRN 2197370*.
- Andrei, D., H. Friedman, and N. B. Ozel (2023). “Economic uncertainty and investor attention.” In: *Journal of Financial Economics* 149.2, pp. 179–217.
- Andrei, D. and M. Hasler (2015). “Investor attention and stock market volatility.” In: *The review of financial studies* 28.1, pp. 33–72.
- Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). “The cross-section of volatility and expected returns.” In: *The Journal of Finance* 61.1, pp. 259–299.
- Autor, D. et al. (2020). “The fall of the labor share and the rise of superstar firms.” In: *The Quarterly Journal of Economics* 135.2, pp. 645–709.
- Azar, J., M. C. Schmalz, and I. Tecu (2018). “Anticompetitive effects of common ownership.” In: *The Journal of Finance* 73.4, pp. 1513–1565.
- Baba Yara, F., B. H. Boyer, and C. Davis (2021). “The factor model failure puzzle.” In: *Working Paper*.
- Bae, K.-H., W. Bailey, and J. Kang (2021). “Why is stock market concentration bad for the economy?” In: *Journal of Financial Economics* 140.2, pp. 436–459.

- Banerjee, O., L. E. Ghaoui, and A. d'Aspremont (2008). "Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data." In: *Journal of Machine learning research* 9.Mar, pp. 485–516.
- Barigozzi, M. and C. Brownlees (2019). "Nets: Network estimation for time series." In: *Journal of Applied Econometrics* 34.3, pp. 347–364.
- Barillas, F. and J. Shanken (2018). "Comparing asset pricing models." In: *The Journal of Finance* 73.2, pp. 715–754.
- Bianchi, D., M. Billio, R. Casarin, and M. Guidolin (2015). "Modeling contagion and systemic risk." In: *Available at SSRN*.
- Billio, M., M. Getmansky, A. W. Lo, and L. Pelizzon (2012). "Econometric measures of connectedness and systemic risk in the finance and insurance sectors." In: *Journal of financial economics* 104.3, pp. 535–559.
- Blank, S., C. M. Buch, and K. Neugebauer (2009). "Shocks at large banks and banking sector distress: The Banking Granular Residual." In: *Journal of Financial Stability* 5.4, pp. 353–373.
- Boehm, C. E., A. Flaaen, and N. Pandalai-Nayar (2019). "Input linkages and the transmission of shocks: Firm-level evidence from the 2011 Tōhoku earthquake." In: *Review of Economics and Statistics* 101.1, pp. 60–75.
- Brunner, F. and R. Hipp (2023). "Echo-nomics: Does the stock market resonate with granular innovations?" In: *Working Paper*.
- Bryzgalova, S., J. Huang, and C. Julliard (2022). "Bayesian solutions for the factor zoo: We just ran two quadrillion models." In: *The Journal of Finance*.
- Carvalho, V. M. (2014). "From micro to macro via production networks." In: *Journal of Economic Perspectives* 28.4, pp. 23–48.
- Carvalho, V. M. and A. Tahbaz-Salehi (2019). "Production networks: A primer." In: *Annual Review of Economics* 11, pp. 635–663.
- Chuluun, T., A. Prevost, and A. Upadhyay (2017). "Firm network structure and innovation." In: *Journal of Corporate Finance* 44, pp. 193–214.
- Demirer, M., F. X. Diebold, L. Liu, and K. Yilmaz (2018). "Estimating global bank network connectedness." In: *Journal of Applied Econometrics* 33.1, pp. 1–15.

- Diebold, F. X. and K. Yilmaz (2014). “On the network topology of variance decompositions: Measuring the connectedness of financial firms.” In: *Journal of Econometrics* 182.1, pp. 119–134.
- Elliott, M., B. Golub, and M. O. Jackson (2014). “Financial networks and contagion.” In: *American Economic Review* 104.10, pp. 3115–3153.
- Fama, E. F. and K. R. French (1997). “Industry costs of equity.” In: *Journal of financial economics* 43.2, pp. 153–193.
- Feng, G., S. Giglio, and D. Xiu (2020). “Taming the factor zoo: A test of new factors.” In: *The Journal of Finance* 75.3, pp. 1327–1370.
- Foerster, A. T., P.-D. G. Sarte, and M. W. Watson (2011). “Sectoral versus aggregate shocks: A structural factor analysis of industrial production.” In: *Journal of Political Economy* 119.1, pp. 1–38.
- Freyberger, J., A. Neuhierl, and M. Weber (2020). “Dissecting characteristics nonparametrically.” In: *The Review of Financial Studies* 33.5, pp. 2326–2377.
- Friedman, J., T. Hastie, and R. Tibshirani (2008). “Sparse inverse covariance estimation with the graphical lasso.” In: *Biostatistics* 9.3, pp. 432–441.
- Gabaix, X. (2011). “The granular origins of aggregate fluctuations.” In: *Econometrica* 79.3, pp. 733–772.
- Giannone, D., M. Lenza, and G. E. Primiceri (2021). “Economic predictions with big data: The illusion of sparsity.” In: *Econometrica* 89.5, pp. 2409–2437.
- Giglio, S. and D. Xiu (2021). “Asset pricing with omitted factors.” In: *Journal of Political Economy* 129.7, pp. 1947–1990.
- Goyal, A. and P. Santa-Clara (2003). “Idiosyncratic risk matters!” In: *The journal of finance* 58.3, pp. 975–1007.
- Gu, S., B. Kelly, and D. Xiu (2021). “Autoencoder asset pricing models.” In: *Journal of Econometrics* 222.1, pp. 429–450.
- Harvey, C. R. and Y. Liu (2021). “Lucky factors.” In: *Journal of Financial Economics* 141.2, pp. 413–435.
- Harvey, C. R., Y. Liu, and H. Zhu (2016). “... and the cross-section of expected returns.” In: *The Review of Financial Studies* 29.1, pp. 5–68.

- Hasler, M. and C. Ornathanalai (2018). “Fluctuating attention and financial contagion.” In: *Journal of Monetary Economics* 99, pp. 106–123.
- Herskovic, B. (2018). “Networks in production: Asset pricing implications.” In: *The Journal of Finance* 73.4, pp. 1785–1818.
- Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh (2016). “The common factor in idiosyncratic volatility: Quantitative asset pricing implications.” In: *Journal of Financial Economics* 119.2, pp. 249–283.
- Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh (2020). “Firm volatility in granular networks.” In: *Journal of Political Economy* 128.11, pp. 4097–4162.
- Hou, K. and D. T. Robinson (2006). “Industry concentration and average stock returns.” In: *The journal of finance* 61.4, pp. 1927–1956.
- Kelly, B. and H. Jiang (2014). “Tail risk and asset prices.” In: *The Review of Financial Studies* 27.10, pp. 2841–2871.
- Kelly, B. T., S. Pruitt, and Y. Su (2019). “Characteristics are covariances: A unified model of risk and return.” In: *Journal of Financial Economics* 134.3, pp. 501–524.
- Kozak, S., S. Nagel, and S. Santosh (2020). “Shrinking the cross-section.” In: *Journal of Financial Economics* 135.2, pp. 271–292.
- Lintner, J. (1965). “Security prices, risk, and maximal gains from diversification.” In: *The Journal of Finance* 20.4, pp. 587–615.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.
- Malevergne, Y., P. Santa-Clara, and D. Sornette (2009). *Professor zipf goes to wall street*. Tech. rep. National Bureau of Economic Research.
- Mehra, R. and E. C. Prescott (1985). “The equity premium: A puzzle.” In: *Journal of monetary Economics* 15.2, pp. 145–161.
- Parkinson, M. (1980). “The extreme value method for estimating the variance of the rate of return.” In: *Journal of business*, pp. 61–65.
- Pesaran, H. H. and Y. Shin (1998). “Generalized impulse response analysis in linear multivariate models.” In: *Economics letters* 58.1, pp. 17–29.

- Rothenberg, T. J. (1971). "Identification in parametric models." In: *Econometrica: Journal of the Econometric Society*, pp. 577–591.
- Savor, P. and M. Wilson (2016). "Earnings announcements and systematic risk." In: *The Journal of Finance* 71.1, pp. 83–138.
- Sharpe, W. F. (1964). "Capital asset prices: A theory of market equilibrium under conditions of risk." In: *The Journal of Finance* 19.3, pp. 425–442.
- White, H. (1982). "Maximum likelihood estimation of misspecified models." In: *Econometrica: Journal of the econometric society*, pp. 1–25.
- Zou, H. and H. H. Zhang (2009). "On the adaptive elastic-net with a diverging number of parameters." In: *The Annals of Statistics* 37.4, p. 1733.

Conclusions

In this thesis, I have explored the manifold consequences of spillover networks in economics and finance. The overarching question is whether aggregate variation can be traced through dynamic networks of spillovers all the way to their granular sources. Using tools from time series econometrics, network analysis, statistical learning, and empirical finance, I have investigated the existence and implications of such granular effects. I have approached this theme from three different angles. First, Chapter 1 probes several methods to infer spillover networks from large panels, and applies the methodology to industrial production data to gain insight into the decline in aggregate volatility. Second, Chapter 2 meticulously documents volatility spillover networks in financial markets and makes a connection of asset-level impulses to market-level outcomes. Lastly, Chapter 3 proposes an asset pricing framework that incorporates inter-asset spillovers, and collects evidence for the empirical adequacy of its predictions.

The work in this thesis has yielded a multitude of findings and results that shed light on spillover networks in economics and finance. In particular, it has produced the following methodological, descriptive, theoretical, and empirical contributions:

- Bias and estimation uncertainty of standard estimators of large forecast error variance decomposition connectedness tables can be mitigated by regularization techniques. (Chapter 1)
- The spillover network of sectoral industrial production has considerably weakened, which gives new support to structural change driving the decline in aggregate volatility known as the Great Moderation. (Chapter 1)
- The degree of volatility spillovers in asset markets varies rapidly and has repeatedly provided a structure that permits the market-wide propagation of granular innovations. (Chapter 2)
- Asset-level impulses influence aggregate market outcomes through a combination of directed spillovers and a concentrated size distribution. (Chapter 2)

- A formulation of the stochastic discount factor that includes direct asset-to-asset return spillovers implies risk premia arise for systemically relevant granular innovations. (Chapter 3)
- Various empirical observations suggest that granular risk premia exist and that they could be a reason for the continuously emerging anomalies relative to reduced-form factor pricing models. (Chapter 3)

In the larger view on the fields of economics and finance, these insights hopefully inspire more researchers to revisit the way we think about the relation of micro-economic narratives and macro-economic outcomes. If spillover networks between low-level entities play a significant role in determining aggregate quantities, they bring into focus the need for the careful consideration of granular effects in economic frameworks. The findings in this thesis show that such mechanisms are at play in several particular circumstances. Although some of the empirical results may need additional support from future studies to overcome their limitations, the evidence collected in this thesis points towards the undeniable importance of granular effects and spillover networks in economics and finance.