Quantitative Economics 14 (2023), 1021–1058 1759-7331/20231021

Estimating large-dimensional connectedness tables: The great moderation through the lens of sectoral spillovers

Felix Brunner
Nova School of Business and Economics

Ruben Hipp
Financial Stability Department, Bank of Canada

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.

Keywords. VAR models, shrinkage, networks, industrial production.

JEL classification. C32, C52, E23, E27.

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 Yılmaz (2014)). Namely, FEVDs measure how much the variation of one sector can explain the variation of an-

Felix Brunner: brunner.felix@gmail.com
Ruben Hipp: rhipp@bankofcanada.ca

We are grateful for comments from Jason Allen, Tatjana Dahlhaus, Thibaut Duprey, Soojin Jo, Paulo M.M. Rodrigues, various seminar participants, and two anonymous referees. This work was supported by the Fundação para a Ciência e Tecnologia, Portugal. A special thanks goes to Matteo Barigozzi and Christian Brownlees, without whom this work would not exist.

 https://doi.org/10.3982/QE1947

other. 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 (e.g., see Demirer, Diebold, Liu, and Yilmaz (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, Carvalho, Ozdaglar, and Tahbaz-Salehi (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, Hastie, and Tibshirani (2008), optimal shrinkage estimators, as in Ledoit and Wolf (2004), and sample covariance thresholding, as in Bickel and Levina (2008), Rothman, Levina, and Zhu (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, and Watson (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, and Watson (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, McConnell, and Perez-Quiros (2002), Summers et al. (2005), and Davis and Kahn (2008).

The rest of the paper is organized as follows. In Section 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 3. Section 4 applies the regularization of FEVDs to the IP setup to answer the question of sectoral spillovers. Finally, Section 5 concludes. The Supplementary Material (Brunner and Hipp (2023)), such as mathematical details, practical illustrations, and complementary empirical graphs can be found in the Online Appendix.

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,

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

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 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 regu-

[1]Note 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, p. 15)).

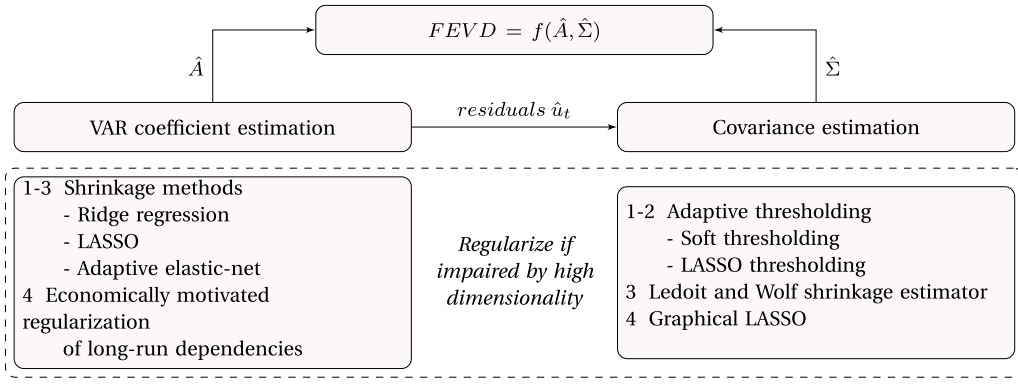

FIGURE 1. Estimation overview for high-dimensional forecast error variance decompositions (FEVDs). The estimates $\hat{A}$ and $\hat{\Sigma}$ are required for the estimation of FEVDs. The lists describe the considered regularization methods in this paper. Note that although the VAR residuals are fed into the estimation of the covariance matrix, properties such as the degree of shrinkage and sparsity structure are determined independently by the estimators.

larization approaches for large VARs.[2] Figure 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 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 2.2.

[2]We 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).

2.1 *Generalized forecast error variance decompositions*

Due to the stability assumption, the VAR process in (1) can be written in moving average (MA) representation

$$y_t = \mu + \sum_{k=0}^{\infty} \Phi_k u_{t-k}, \tag{2}$$

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, Hashem Pesaran, and Potter (1996) and Pesaran and Shin (1998) define the (unscaled) generalized impulse response (IR) function at horizon h to an impulse Δ_j on the jth 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:

$$\mathsf{IR}(h, \Delta_j, j) = E[y_{t+h}|u_{j,t} = \Delta_j] - E[y_{t+h}|u_{j,t} = 0]. \tag{3}$$

Under a Gaussian assumption, we can use

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

with σ_{ij} being the ijth entry in Σ and e_j as the jth column of the identity matrix. Substituting this expression in (3), the impulse response function can be rewritten as

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

It is customary to set $\Delta_j = \sqrt{\sigma_{jj}}$, which yields the scaled generalized impulse response functions $\mathsf{IR}(h, \sqrt{\sigma_{jj}}, j)$. We assemble the scaled IRs in the $(N \times N)$ matrix

$$\Psi^g(h) = \left[\psi_{ij}^g(h)\right] = \left[\mathsf{IR}(h, \sqrt{\sigma_{11}}, 1), \ldots, \mathsf{IR}(h, \sqrt{\sigma_{NN}}, N)\right] = \Phi_h \Sigma \operatorname{diag}(\Sigma)^{-\frac{1}{2}}, \tag{4}$$

where $\operatorname{diag}(M)$ 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 $\mathrm{P} = \Sigma \operatorname{diag}(\Sigma)^{-\frac{1}{2}}$ and define the generalized shock u_t^g as

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

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

$$y_t = \mu + \sum_{k=0}^{\infty} \Psi^g(k) u_{t-k}^g. \tag{5}$$

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

$$\mathrm{MSE}\big[y_{i,t+H-1}|\{u_{t+h}^g = e_j\}_{h=0}^{H-1}\big] = \sum_{h=0}^{H-1}\big(\psi_{ij}^g(h)\big)^2. \tag{6}$$

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

$$\mathrm{MSE}[y_{i,t+H-1}] = \left(\sum_{h=0}^{H-1}\big(\Phi_h \Sigma \Phi_h'\big)\right)_{ii}. \tag{7}$$

Pesaran and Shin (1998) divide (6) by (7) and get a table showing the contributions from innovations in variable j to the H-step-ahead forecast error variance of variable i. Like Diebold and Yılmaz (2014), we denote the H-step-ahead generalized FEVD as $D^{gH} = [d_{ij}^{gH}]$ with entries

$$d_{ij}^{gH} = \frac{\mathrm{MSE}\big[y_{i,t+H-1}|\{u_{t+h}^g = e_j\}_{h=0}^{H-1}\big]}{\mathrm{MSE}[y_{i,t+H-1}]}. \tag{8}$$

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 Yılmaz (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) 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.

[3]For a discussion of Ω, see Raveh (1985).

2.2 *Estimating large forecast error variance decompositions*

2.2.1 *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,

$$y = \big(Z' \otimes I_N\big)\beta + u, \tag{9}$$

where $y = \text{vec}([y_1, \ldots, y_T])$, $Z_{t-1} = (1, y'_{t-1})'$, $Z = [Z_0, \ldots, Z_{T-1}]$, $\beta = \text{vec}([\nu, A])$ and $u = \text{vec}([u_1, \ldots, u_T])$. 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

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

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 Zou and Zhang's (2009) adaptive elastic-net. This penalized estimator is a compound of the general concepts of elastic-net and adaptive LASSO. In particular, it simultaneously shrinks and selects entries in the coefficient matrices and, moreover, has the oracle property, which ensures optimal large-sample performance. A comprising definition of the adaptive elastic-net estimator class is

$$\hat{\beta}_{\text{AEnet}} = \underset{\beta}{\text{argmin}} \left[\|y - X\beta\|^2 + \lambda^{\text{NET}} \sum_{i=1}^{N^2+N} w_i \left(\alpha|\beta_i| + (1-\alpha)\frac{1}{2}\beta_i^2 \right) \right], \tag{10}$$

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, Hastie, and Tibshirani (2010).[4]

The regression in (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)).

[4]The `glmnet` routine is available for many programming languages. We employ the implementation in MATLAB by Qian, Hastie, Friedman, Tibshirani, and Simon (2013).

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 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 $w_i = |\hat{\beta}_{i,\mathrm{OLS}}|^{-1}$ for the regularization and hyperparameter selection.

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

$$\begin{aligned} \mathrm{FE}(H) &= \sum_{h=0}^{H-1} \Phi_h = \sum_{h=0}^{H-1} A^h, \\ \lim_{H\to\infty} \mathrm{FE}(H) &= \sum_{h=0}^{\infty} A^h = (I_N - A)^{-1}, \end{aligned} \tag{11}$$

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 $\mathrm{FE}(\infty) = (I_N - A)^{-1}$ is zero.

It is evident that zeros in the forecast error most likely imply that the respective forecast error variance—that is, the elementwise squared version—is also zero. Thus, to impose sparsity on the spillover network, this matrix is a potential regularization target. Take (11) and plug it into the model (1):

$$\begin{aligned} y_t &= \nu + \big(I_N - \mathrm{FE}(\infty)^{-1}\big) y_{t-1} + u_t, \\ \Delta y_t &= y_t - y_{t-1} = \nu - \mathrm{FE}(\infty)^{-1} y_{t-1} + u_t, \\ y_{t-1} &= \mathrm{FE}(\infty)\nu + \mathrm{FE}(\infty)(-\Delta y_t) + \mathrm{FE}(\infty) u_t \end{aligned} \tag{12}$$

Estimating (12) as a penalized regression permits us to regularize $\mathrm{FE}(\infty)$ and with $A = I_N - \mathrm{FE}(\infty)^{-1}$ we can back out the autoregressive coefficients. To the best of our

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

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.

2.2.2 *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 u_t corresponding to the regression residuals $\hat{u}_t$ 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 $\mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_N)'$ be a N-variate random vector with covariance matrix $\Sigma = [\sigma_{ij}]_{N\times N}$, and assume an *i.i.d.* random sample $\{u_1, \ldots, u_T\}$ 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

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

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

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

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

$$\lambda_{ij} := \lambda_{ij}\big(\delta^{\mathrm{AT}}\big) = \delta^{\mathrm{AT}}\sqrt{\frac{\hat{\theta}_{ij}\log N}{T}}, \tag{13}$$

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

$$\hat{\Sigma}^{\mathrm{AT}} = \left[\hat{\sigma}_{ij}^{\mathrm{AT}}\right]_{N\times N} = \left[s_{\lambda_{ij}}(\hat{\sigma}_{ij})\right]_{N\times N} \tag{14}$$

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 (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 A 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

$$\hat{\Sigma}^{\mathrm{LW}} = \delta^{\mathrm{LW}} m I_N + \left(1 - \delta^{\mathrm{LW}}\right)\hat{\Sigma}, \tag{15}$$

where $\hat{\Sigma}$ denotes the sample covariance matrix, $m = N^{-1}\operatorname{tr}(\hat{\Sigma})$ 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.

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

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

[6]This rule incorporates soft thresholding for $\eta = 1$: $s_\lambda(\hat{\sigma}_{ij}) = \operatorname{sgn}(\hat{\sigma}_{ij})(|\hat{\sigma}_{ij}| - \lambda)_+$. 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.

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 (13) in the adaptive threshold case, the term $\delta^{\mathrm{LW}} = b^2/d^2$ automatically assigns more weight to mI_N if the sample's second-order variance, measured by b^2, is large.

GLASSO Friedman, Hastie, and Tibshirani (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

$$\hat{\Theta}^{\mathrm{GLASSO}} = \underset{\Theta \geq 0}{\operatorname{argmax}} \left\{ \operatorname{logdet} \Theta - \operatorname{tr}(\hat{\Sigma}\Theta) - \delta^{\mathrm{GL}} \|\Theta\|_1 \right\} \tag{16}$$

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.

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's (2004) 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 param-

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

[8] The 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.

eters 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 B.1 in Appendix B. The validity of this approach is shown in Bergmeir, Hyndman, and Koo (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 kth 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 on the squared observations $u_t u_t'$ in the kth test sample κ,

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

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 4.

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

[10]Friedman, Hastie, and Tibshirani (2010) provide formulas to set the minimum and maximum hyperparameters for general elastic-nets.

3. Simulation study

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)
$A = 0.5I_N$.

$\Sigma = I_N$.

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.

$\Sigma = SDS'$. Here, $S = (I_N - B)^{-1}$ where B is an upper triangular matrix with entries $b_{ij} = \tau$ if $a_{ij} \neq 0, \forall i > j$. $D = \text{diag}(1, \ldots, 2)$ 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.

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

Note that *DGPs 1* and *4* always produce a sparse FEVD while *DGPs 2* and *3* do not. Moreover, all but *DGP 3* are deterministic and *DGPs 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.

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 *DGPs* and calculate the Frobenius norm of the regularized versus the nonregularized version: $\|\widehat{D}^{gH}_{\text{reg}} - D^{gH}\| / \|\widehat{D}^{gH}_{\text{OLS}} - D^{gH}\|$, where $H = 10$.

TABLE 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 $\|\widehat{D}^{gH}_{\text{reg}} - D^{gH}\|/\|\widehat{D}^{gH}_{\text{OLS}} - D^{gH}\|$. 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 Ns and Ts 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.

	Ridge			LASSO			AENET			Geometric		
$N\backslash T$	75	175	500	75	175	500	75	175	500	75	175	500
DGP1												
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%
DGP2												
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%
DGP3												
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%
DGP4												
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%

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

Table 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

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

TABLE 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 $\|\widehat{D}^{gH}_{\text{reg}} - D^{gH}\| / \|\widehat{D}^{gH}_{\text{OLS}} - D^{gH}\|$. 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 Ns and Ts with 500 Monte Carlo repetitions. *DGP 3* has 25 different random realizations of A and Σ.

	Sample-Cov			Threshold			Ledoit–Wolf			GLASSO		
$N\backslash T$	75	175	500	75	175	500	75	175	500	75	175	500
DGP1												
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%
DGP2												
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%
DGP3												
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%**
DGP4												
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%**

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: $\lambda^{\text{NET}*} = \text{argmin}_\lambda \|\hat{A}(\lambda) - A\|$. Additionally, we set the respective penalty parameter for the covariance regularization such that $\delta^* = \text{argmin}_\delta \|\widehat{D}^{gH}_{\text{reg}}(\delta) - D^{gH}\|$.

Table 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 autocorre-

lation, 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.

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 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 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.

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 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 (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 Ts. 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.

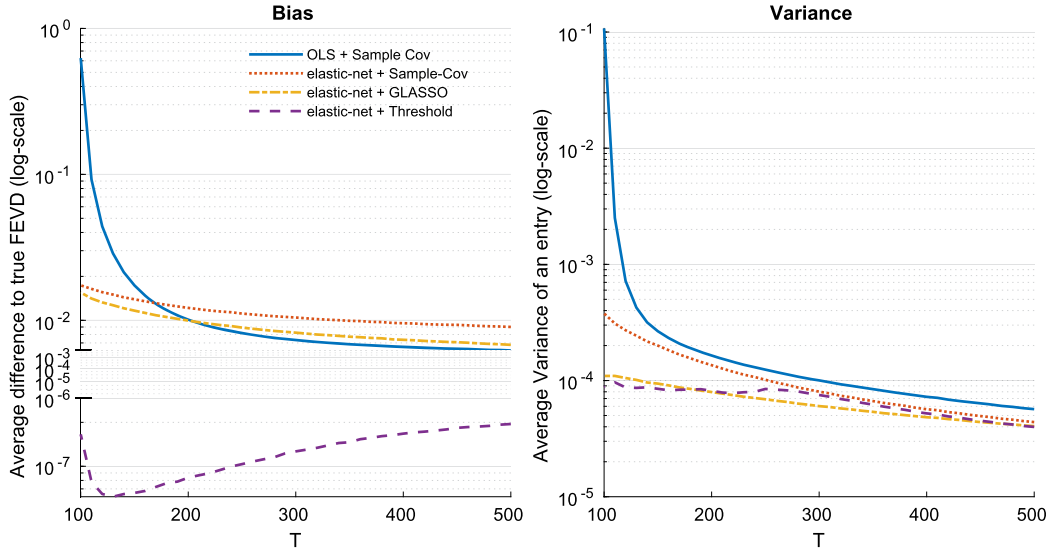

FIGURE 2. Simulation results for 500 Monte Carlo repetitions of *DGP 4* and $N = 100$. The left-hand panel shows the average mean difference of the estimates to the true values $N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\text{mean}(\widehat{D}_{ij}^{gH} - D_{ij}^{gH})$. The right-hand panel shows the respective 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.

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 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: $\text{accuracy} = (\text{TP} + \text{TN})/N^2$. We compare this metric of the estimators for increasing T in Figure 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 $\text{FPR} = \text{FP}/(\text{FP} + \text{TN})$ against the true-positive rate $\text{TPR} = \text{TP}/(\text{TP} + \text{FN})$ 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

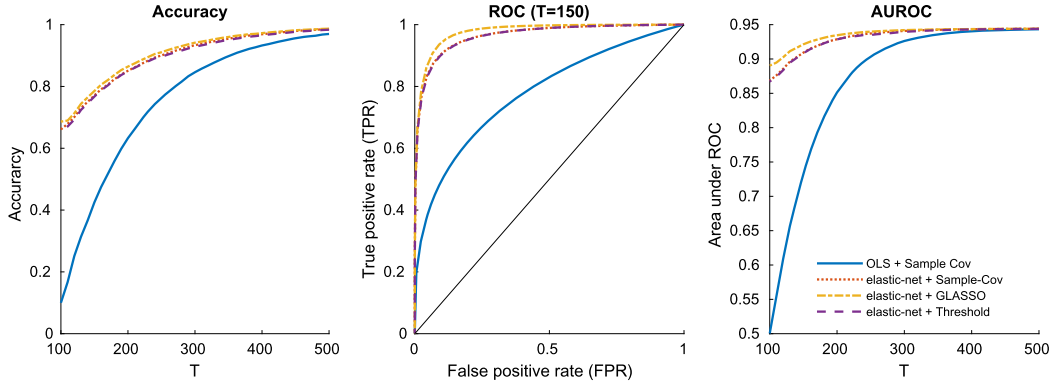

FIGURE 3. Simulation results for 500 Monte Carlo repetitions of *DGP 4* and $N = 100$. The first panel shows the accuracy for increasing sample size T on the x-axis. The center panel shows the receiver operating characteristic (ROC) for $T = 200$, which has FPR on the x-axis and TPR on the y-axis. The 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.

$(x, y) = (1, 1)$. 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 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 Ts. 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.

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 tradeoffs 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.

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, and Watson (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, and Watson (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 (e.g., Leduc and Sill (2007), Justiniano and Primiceri (2008)), factor productivity (e.g., Arias, Hansen, and Ohanian (2007)), or oil supply (e.g., 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, and Watson (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, and Watson (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, Sarte, and Watson (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, McConnell, and Perez-Quiros (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

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

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, and Watson (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.

4.1 *Data*

Similar to Foerster, Sarte, and Watson (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 $\mathrm{IP}_{i,t}$ denote the value of IP of sector i at date t. We take monthly growth rates and annualize the respective percentage points, $g_{i,t} = 1200 \times \ln(\mathrm{IP}_{i,t}/\mathrm{IP}_{i,t-1})$. 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) 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.

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, Foerster, Sarte, and Watson's (2011) average pairwise correlations and aggregate shocks may be better understood if we break up quarterly volatilities into three serially correlated monthly volatilities. Hence, the central regression specification is as follows:

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

While Foerster, Sarte, and Watson (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]

[13] See Figure D.1 in Appendix D for a visual display of the growth rates of IP on an aggregate level.

[14] Note 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.

TABLE 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.

	MSE				AIC
	1972–1983	1984–1995	1996–2007	2008–2019	1972–2019
Panel A: Model selection for regression					
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**
Panel B: Model selection for covariance matrix					
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

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 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 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 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 extend 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

[15]As 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.

TABLE 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.

		A	
AIC		Static	Dynamic
Σ	static	4.6900	4.6417
	dynamic	4.6444	**4.5920**

throughout. For all subsequent analysis, we focus only on the selected 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 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 (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 Yılmaz (2014): in-, out-, and average connectedness. These measures are defined as the row sum, column sum, and the average row sum without the diagonal entries, respectively,[17]

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

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

[17] To 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.

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

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

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.

4.3 *Results*

Table 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 3).

Moreover, we want to emphasize the decreased usage of the degrees of freedom of the regularized estimators in Table 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 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 4. From eye-balling, it is evident that the network significantly changed after 1984. Whereas the pre-Great Moderation subsample shows

TABLE 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 Ledoit and Wolf's (2002) test statistic (see Appendix C).

	Standard				Regularized			
	1972–1983	1984–1995	1996–2007	2008–2019	1972–1983	1984–1995	1996–2007	2008–2019
Three-digit (88 sectors)								
$C(D^{gH})$	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%
Two-digit (26 sectors)								
$C(D^{gH})$	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%
Four-digit (117 sectors)								
$C(D^{gH})$	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%
Five-digit (138 sectors)								
$C(D^{gH})$	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%

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 be-

[18]In Appendix D, the corresponding nonregularized estimates (Figure D.4) fail to provide similar insight.

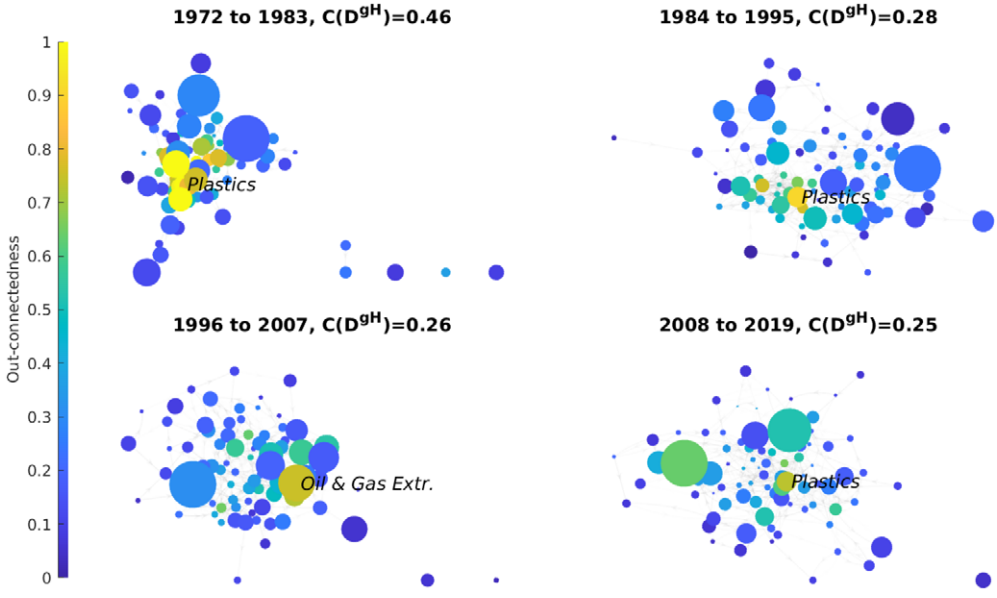

FIGURE 4. Connectedness networks for the respective periods. The force-directed graph drawing algorithm arranges the nodes. That is, two nodes appear closer in the graph if they have stronger connections to each other. Although we initialize the subgraphs on the same scale, the algorithm cannot guarantee 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.

fore the Great Moderation and decreased contributions afterwards. As a first observation from Figure 4, large values for out-connectedness (lighter 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 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

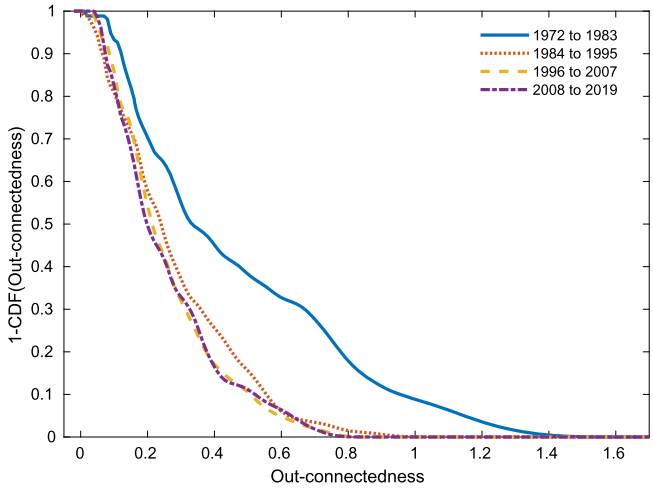

FIGURE 5. Estimated complementary cumulative distribution functions (CCDFs) of the out–connectedness of the 88 three-digit-level sectors. This function is the equivalent to survival functions in the time domain.

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 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), whereas for the majority of sectors the out-connectedness stays relatively constant (close to the diagonal). Remarkably, it seems that only a few sectors increased in out-connectedness (below the diagonal). 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 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.

These findings connect to a strand of literature on inventory-based explanations of the Great Moderation (see Kahn, McConnell, and Perez-Quiros (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

[19]Supporting histograms of the in- and out-connectedness measures (Figures D.5 and D.7) are available in Appendix D. Here, the 25% strongest sectors by pre-Great Moderation out-connectedness increasingly blend into the distribution in subsequent subsamples.

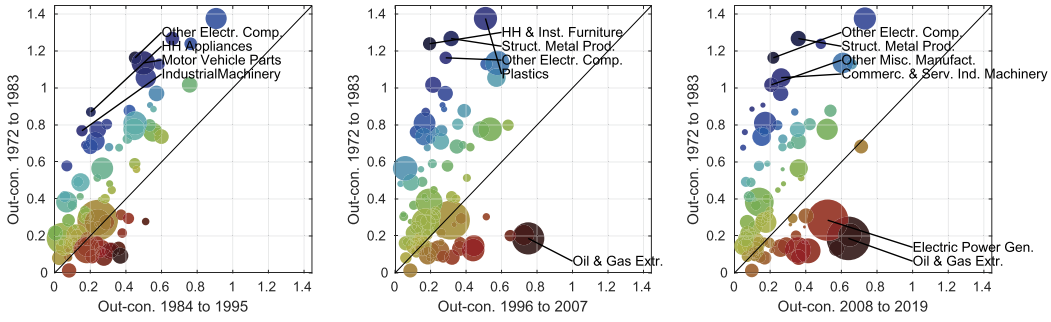

FIGURE 6. Out-connectedness scatter plot of the 88 three-digit-level sectors pre- versus post-Great Moderation. Each dot in the plot corresponds to a sector, with the position along the y- and x-axes according to their out-connectedness before and after the Great Moderation, respectively. The sizes of 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.

(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, Padmanabhan, and Whang (1997)). The amplification of demand shocks may contribute to excessively low demand in input-supplying sectors, and thus, intensify input–output dependencies.

Empirically, Chen, Frank, and Wu (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 D, Figure D.8 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 D.5 and D.6 in Appendix D, 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 $\hat{\sigma}^2_{\text{IP}} = \bar{w}' \hat{\Sigma}_g \bar{w}$, 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, $\text{IRV}(H) = \sum_{h=0}^{H-1} \Phi_h \Sigma = \sum_{h=0}^{H-1} \text{cov}(y_t, u_{t-h})$. We denote this term as the innovation response variance (IRV), and define the vector $v^{\text{gran}}_{\cdot \to \text{IP}}$ as the sectoral variance contributions to the model-implied aggregate IP index:

$$v^{\text{gran}}_{\cdot \to \text{IP}} = \bar{w}^{\cdot 2'} \, \text{IRV}(H) \quad \text{and} \quad v^{\text{gran}} := v^{H}_{\cdot \to \text{IP}} \mathbf{1}, \tag{17}$$

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 $\text{FRV}(H) = \sum_{h=0}^{H-1} \Phi_h B \Sigma_f B' = \sum_{h=0}^{H-1} \text{cov}(y_t, Bf_{t-h})$. Analogously, the vector of sectoral contributions to the index variance induced by reactions to factor innovations is $v^f = \bar{w}^{\cdot 2'} \, \text{FRV}(H) \mathbf{1}$.

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 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

TABLE 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.

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

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, McConnell, and Perez-Quiros (2002), Summers et al. (2005), and Davis and Kahn (2008), the decline in extrinsic and lagged contributions 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 (17). To do so, Figure 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 6:

$$v^{\text{gran}}_{\cdot\to\text{IP}} = \underbrace{\bar{w}^{\cdot 2'} \,\text{diag}\big(\text{IRV}(H)\big)}_{\text{intrinsic component}} + \underbrace{\bar{w}^{\cdot 2'} \big[\text{IRV}(H) - \text{diag}\big(\text{IRV}(H)\big)\big]}_{\text{extrinsic component}}. \tag{18}$$

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 rel-

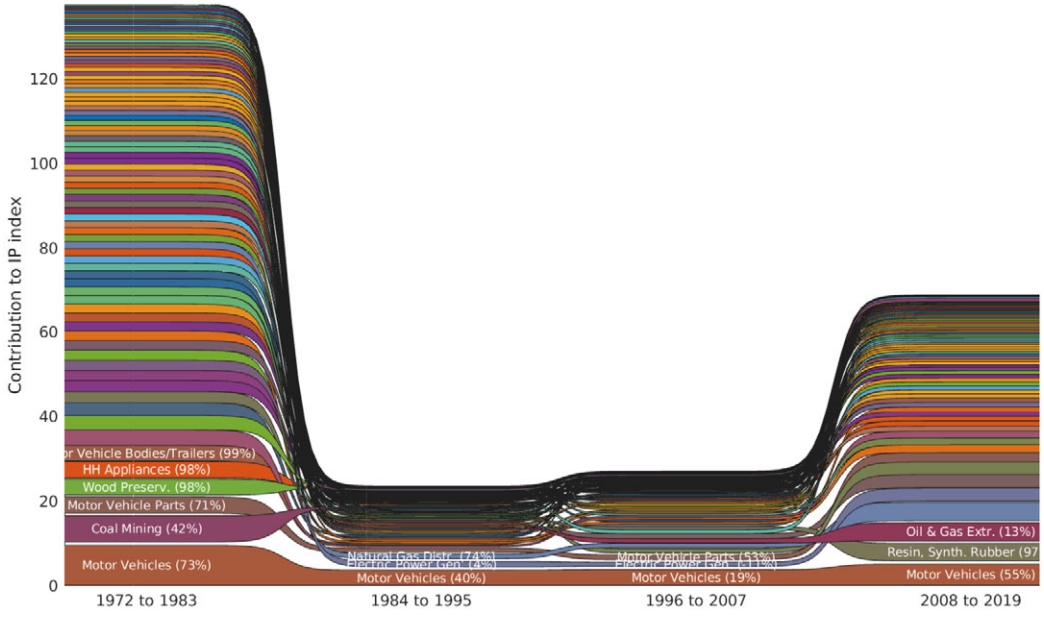

FIGURE 7. Bump chart of sectoral contributions to the variation of the aggregate IP index. Contributions are calculated as in (17). Data includes IP indices for 88 three-digit-level sectors. Adjacent sectors are in similar colors. The top contributors per subsample are labeled. Values in brackets correspond to the extrinsic component of the contribution. For example, the value of the Motor Vehicles sector in the first subsample states that 73% of this sector's contribution is due to spillovers (extrinsic) to other sectors and 27% is due to its intrinsic component.

evance in the aggregate index. Complementarily, Appendix D presents figures for the intrinsic (Figure D.10) and extrinsic (Figure D.9) contributions separately.

First, note that with the Great Moderation, total contributions decrease sharply and relative contributions of individual sectors vary in importance. While the Motor Vehicles sector is the strongest contributor due to its size and outgoing linkages, the Coal Mining sector's effect on the aggregate is relatively small due to lower spillovers of its large intrinsic variation documented in Figure D.10.[20] In contrast to Foerster, Sarte, and Watson's (2011) narrative that metal-related industries are well explained by common factors, our analysis shows that downstream sectors related to the metal industry exhibit strong contributions in the form of spillovers. This finding is consistent with the idea of shocks propagating upstream, for example, the amplification of demand shocks through excessive inventory holdings. Considering that heavy industries are still among the most influential sectors after 1984 (compare Figure 7) with nearly unchanged intrinsic contributions, we reckon that macroeconomic shocks and sectoral innovations alone do not suffice to explain the decline in aggregate volatility. Instead, 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

[20]High volatility before the Great Moderation in the Coal Mining sector mainly stems from the strikes in 1974, 1977–1978, and 1981.

partial increase in spillover contributions in the 2008 to 2019 subsample (compare Figure D.9 in Appendix D), 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 5, which also shows in many visualizations such as Figures 4, 6, and 7. As illustrated in Figure 5, our evidence suggests that outgoing links have weakened considerably. A decomposition of aggregate variance in Table 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 6 and 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 D.8). 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, and Watson (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, and Watson (2011) and Gabaix (2011) by connecting their narratives to the inventory-based explanations of Kahn, McConnell, and Perez-Quiros (2002), Summers et al. (2005), and Davis and Kahn (2008).

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.

References

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi (2012), "The network origins of aggregate fluctuations." *Econometrica*, 80 (5), 1977–2016. [1022, 1023, 1040, 1041, 1053]

Arias, Andres, Gary D. Hansen, and Lee E. Ohanian (2007), "Why have business cycle fluctuations become less volatile?" *Economic theory*, 32 (1), 43–58. [1040]

Barigozzi, Matteo and Christian Brownlees (2013), "Nets: Network estimation for time series." *Journal of Applied Econometrics*. [1032]

Bauer, Michael D. and Eric T. Swanson (2022), "A reassessment of monetary policy surprises and high-frequency identification." Technical report. National Bureau of Economic Research. [1042]

Bedrick, Edward J. and Chih-Ling Tsai (1994), "Model selection for multivariate regression in small samples." *Biometrics*, 226–231. [1043]

Bergmeir, Christoph, Rob J. Hyndman, and Bonsoo Koo (2018), "A note on the validity of cross-validation for evaluating autoregressive time series prediction." *Computational Statistics & Data Analysis*, 120, 70–83. [1033]

Bernanke, Ben (2004), "The great moderation." Washington, DC. [1040]

Bickel, Peter J. and Elizaveta Levina (2008), "Regularized estimation of large covariance matrices." *The Annals of Statistics*, 199–227. [1023]

Bray, Robert L. and Haim Mendelson (2012), "Information transmission and the bullwhip effect: An empirical investigation." *Management Science*, 58 (5), 860–875. [1049]

Brunner, Felix, and Ruben Hipp (2023), "Supplement to 'Estimating large-dimensional connectedness tables: The great moderation through the lens of sectoral spillovers'."

Quantitative Economics Supplemental Material, 14, https://doi.org/10.3982/QE1947. [1024]

Cai, Tony and Weidong Liu (2011), "Adaptive thresholding for sparse covariance matrix estimation." *Journal of the American Statistical Association*, 106 (494), 672–684. [1023, 1030, 1031, 1033]

Carvalho, Vasco and Xavier Gabaix (2013), "The great diversification and its undoing." *American Economic Review*, 103 (5), 1697–1727. [1023, 1040]

Carvalho, Vasco M. (2014), "From micro to macro via production networks." *Journal of Economic Perspectives*, 28 (4), 23–48. [1041]

Carvalho, Vasco M. and Alireza Tahbaz-Salehi (2019), "Production networks: A primer." *Annual Review of Economics*, 11, 635–663. [1023, 1041]

Chen, Hong, Murray Z. Frank, and Owen Q. Wu (2005), "What actually happened to the inventories of American companies between 1981 and 2000?" *Management science*, 51 (7), 1015–1031. [1049]

Davis, Steven J. and James A. Kahn (2008), "Interpreting the great moderation: Changes in the volatility of economic activity at the macro and micro levels." *Journal of Economic perspectives*, 22 (4), 155–180. [1024, 1040, 1048, 1049, 1051, 1053]

Demirer, Mert, Francis X. Diebold, Laura Liu, and Kamil Yilmaz (2017), "Estimating global bank network connectedness." *Journal of Applied Econometrics*. [1022, 1023, 1028]

Diebold, Francis X. and Kamil Yılmaz (2014), "On the network topology of variance decompositions: Measuring the connectedness of financial firms." *Journal of Econometrics*, 182 (1), 119–134. [1021, 1027, 1044]

Fernald, John (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. [1042]

Foerster, Andrew T., Andreas Hornstein, Pierre-Daniel G. Sarte, and Mark W. Watson (2022), "Aggregate implications of changing sectoral trends." *Journal of Political Economy*, 130 (12), 3286–3333. [1040]

Foerster, Andrew T., Pierre-Daniel G. Sarte, and Mark W. Watson (2011), "Sectoral versus aggregate shocks: A structural factor analysis of industrial production." *Journal of Political Economy*, 119 (1), 1–38. [1023, 1040, 1041, 1042, 1052, 1053]

Forrester, Jay Wright (1961), *Industrial Dynamics*. MIT Press. [1049]

Friedman, Jerome, Trevor Hastie, and Rob Tibshirani (2010), "Regularization paths for generalized linear models via coordinate descent." *Journal of statistical software*, 33 (1), 1. [1028, 1029, 1033]

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani (2008), "Sparse inverse covariance estimation with the graphical lasso." *Biostatistics*, 9 (3), 432–441. [1023, 1032]

Gabaix, Xavier (2011), "The granular origins of aggregate fluctuations." *Econometrica*, 79 (3), 733–772. [1023, 1040, 1041, 1050, 1053]

Galí, Jordi and Luca Gambetti (2009), "On the sources of the great moderation." *American Economic Journal: Macroeconomics*, 1 (1), 26–57. [1024]

Hastie, Trevor, Robert Tibshirani, Jerome H. Friedman, and Jerome H. Friedman (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Vol. 2. Springer. [1033]

Hipp, Ruben (2020), "On causal networks of financial firms: Structural identification via non-parametric heteroskedasticity." Staff Working Papers 20-42, Bank of Canada. [1027]

Hoerl, Arthur E. and Robert W. Kennard (1970), "Ridge regression: Biased estimation for nonorthogonal problems." *Technometrics*, 12 (1), 55–67. [1023]

Hurvich, Clifford M. and Chih-Ling Tsai (1989), "Regression and time series model selection in small samples." *Biometrika*, 76 (2), 297–307. [1043]

Justiniano, Alejandro and Giorgio E. Primiceri (2008), "The time-varying volatility of macroeconomic fluctuations." *American Economic Review*, 98 (3), 604–641. [1040]

Kahn, James A., Margaret Mary McConnell, and Gabriel Perez-Quiros (2002), "On the causes of the increased stability of the us economy." *Economic Policy Review*, 8 (1). [1024, 1040, 1048, 1051, 1053]

Kappen, Hilbert J. and Vicenç Gómez (2014), "The variational garrote." *Machine Learning*, 96 (3), 269–294. [1025]

Kascha, Christian and Carsten Trenkler (2015), "Forecasting vars, model selection, and shrinkage." Technical report, Working Paper Series. Department of Economics, University of Mannheim. [1023, 1028, 1033]

Koop, Gary, M. Hashem Pesaran, and Simon M. Potter (1996), "Impulse response analysis in nonlinear multivariate models." *Journal of Econometrics*, 74 (1), 119–147. [1026]

Lam, Clifford et al. (2016), "Nonparametric eigenvalue-regularized precision or covariance matrix estimator." *The Annals of Statistics*, 44 (3), 928–953. [1025]

Ledoit, Olivier and Michael Wolf (2002), "Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size." *The Annals of Statistics*, 30 (4), 1081–1102. [1046]

Ledoit, Olivier and Michael Wolf (2004), "A well-conditioned estimator for large-dimensional covariance matrices." *Journal of Multivariate Analysis*, 88 (2), 365–411. [1023, 1031, 1032]

Leduc, Sylvain and Keith Sill (2007), "Monetary policy, oil shocks, and tfp: Accounting for the decline in us volatility." *Review of Economic Dynamics*, 10 (4), 595–614. [1040]

Lee, Hau L., Venkata Padmanabhan, and Seungjin Whang (1997), "Information distortion in a supply chain: The bullwhip effect." *Management science*, 43 (4), 546–558. [1049]

Lütkepohl, Helmut (2005), *New Introduction to Multiple Time Series Analysis*. Springer Science & Business Media. [1024]

Nakov, Anton and Andrea Pescatori (2010), "Oil and the great moderation." *The Economic Journal*, 120 (543), 131–156. [1040]

Pesaran, H. Hashem and Yongcheol Shin (1998), "Generalized impulse response analysis in linear multivariate models." *Economics letters*, 58 (1), 17–29. [1024, 1026, 1027]

Qian, Junyang, Trevor Hastie, Jerome Friedman, Rob Tibshirani, and Noah Simon (2013), "Glmnet for Matlab." http://www.stanford.edu/~hastie/glmnet_matlab/. [1028]

Raveh, Adi (1985), "On the use of the inverse of the correlation matrix in multivariate data analysis." *The American Statistician*, 39 (1), 39–42. [1027]

Rigobon, Roberto and Brian Sack (2003), "Measuring the reaction of monetary policy to the stock market." *The quarterly journal of Economics*, 118 (2), 639–669. [1027]

Romer, Christina D. and David H. Romer (2004), "A new measure of monetary shocks: Derivation and implications." *American Economic Review*, 94 (4), 1055–1084. [1042]

Rothman, Adam J., Elizaveta Levina, and Ji Zhu (2009), "Generalized thresholding of large covariance matrices." *Journal of the American Statistical Association*, 104 (485), 177–186. [1023]

Sims, Christopher A. (1980), "Macroeconomics and reality." *Econometrica: Journal of the Econometric Society*, 1–48. [1027]

Sims, Christopher A. and Tao Zha (2006), "Were there regime switches in us monetary policy?" *American Economic Review*, 96 (1), 54–81. [1043, 1044]

Stock, James H. and Mark W. Watson (2003), "Has the business cycle changed? Evidence and explanations." *Monetary Policy and Uncertainty: Adapting to a Changing Economy*, 9–56. [1040]

Stone, Mervyn (1977), "An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion." *Journal of the Royal Statistical Society: Series B (Methodological)*, 39 (1), 44–47. [1044]

Summers, Peter M. et al. (2005), "What caused the great moderation? Some cross-country evidence." *Economic Review-Federal Reserve Bank of Kansas City*, 90 (3), 5. [1024, 1040, 1048, 1051, 1053]

Tibshirani, Robert (1996), "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288. [1023]

Wieland, Johannes F. and Mu-Jeung Yang (2020), "Financial dampening." *Journal of Money, Credit and Banking*, 52 (1), 79–113. [1042]

Zou, Hui (2006), "The adaptive lasso and its oracle properties." *Journal of the American Statistical Association*, 101 (476), 1418–1429. [1029]

Zou, Hui and Hao Helen Zhang (2009), "On the adaptive elastic-net with a diverging number of parameters." *The Annals of Statistics*, 37 (4), 1733. [1023, 1028, 1029]

Co-editor Tao Zha handled this manuscript.

Manuscript received 8 July, 2021; final version accepted 23 March, 2023; available online 17 April, 2023.