University of Pretoria
Department of Economics Working Paper Series

Unveiling True Connectedness in US State-Level Stock Markets: The Role of Common Factors

Massimiliano Caporin
University of Padova
Oguzhan Cepni
Copenhagen Business School and University of Edinburgh Business School and Ostim Technical University
Rangan Gupta
University of Pretoria
Working Paper: 2025-09
February 2025

Department of Economics
University of Pretoria
0002, Pretoria
South Africa
Tel: +27 12 420 2413

Unveiling True Connectedness in US State-Level Stock Markets: The Role of Common Factors

Massimiliano Caporin[*], Oguzhan Cepni[**] and Rangan Gupta[****]

Abstract

The objective of this paper is to analyze the time-varying degree of interconnectedness of 50 state-level stock returns and their volatility of the United States (US) while filtering out common factors and insignificant coefficients using Least Absolute Shrinkage and Selection Operator (Lasso) regularization. Based on monthly data from February 1994 to November 2024, we find that not accounting for common factors is likely to result in relatively higher spillover indexes. Our findings, beyond their academic value, have important implications for investors and policymakers.

Keywords: US state-level stock indexes; returns and volatility; common factors; Lasso; spillover indexes

JEL Codes: C32, G10

1. Introduction

The literature on stock market interdependence, both in terms of returns and its volatilities, which originated during the Asian Financial Crisis in 1997, has grown exponentially since the Global Financial Crisis of 2007-2009, and more recently, the COVID-19 pandemic (Yilmaz, 2010; Diebold and Yilmaz, 2023). This is not surprising since understanding connectedness in moments of equity markets is central to risk measurement and management for both investors and policymakers (Bouri et al., 2021). In this regard, recent studies have primarily investigated connectedness of returns and volatilities of stock equity markets across countries by relying on the so-called static (full-sample) and dynamic (rolling-window) spillover indexes approaches developed by Diebold and Yilmaz (2009, 2012, 2014), which are based on forecast error variance decompositions (FEVDs) from vector autoregressions (VARs).

Our objective in this paper is to extend this line of research by focusing, for the first-time, within the United States (US), wherein we estimate and analyze total and net dynamic spillover indexes involving stock returns and volatilities of its 50 states, both at the aggregate- and individual-level, respectively, over the monthly period of February 1994 to November 2024. The underlying motivation for investigating regional equity markets emanates from the fact that core business activities of firms tend to take place close to their headquarters (Pirinsky and Wang, 2006; Chaney et al., 2012). Given this, stock prices, besides having common drivers, should contain a significant regional component, so much so that portfolios of investors put more weight on local firms (Coval and Moskowitz, 1999, 2001; Korniotis and Kumar, 2013).

With this in mind, we filter out common factors that are likely to impact the state-level stock returns to reduce residual correlations in the underlying VAR. At the same time, we also incorporate Least

[*] Corresponding author. Department of Statistical Sciences, University of Padova, Via Cesare Battisti 241, 35121 Padova, Italy. Email: massimiliano.caporin@unipd.it.
[**] Ostim Technical University, Ankara, Turkiye; University of Edinburgh Business School, Centre for Business, Climate Change, and Sustainability; Department of Economics, Copenhagen Business School, Denmark. Email: oce.eco@cbs.dk.
[****] Department of Economics, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa; Email: rangan.gupta@up.ac.za.

Absolute Shrinkage and Selection Operator (Lasso)-based regularization (Tibshirani, 1996) to eliminate non-significant parameters in the VAR model. These two data treatments ensure that the spillover indexes are not, as we show, overestimated, thereby, preventing the transmission of inaccurate information to portfolio managers and policy authorities.

The remainder of the paper is organized as follows: Section 2 presents the data and basics of the methodology, while Section 3 discusses the empirical findings, with Section 4 concluding the paper.

2. Data and Methodology

The daily state-level stock market indexes, from which we compute log-returns, are derived from the Bloomberg terminal, which in turn creates these indexes by taking the capitalization-weighted index of equities domiciled in each state. As discussed above, we filter out the impact of common factors from the stock log-returns of the 50 states by utilizing the following variables: the market returns, the size factor, the value factor, the momentum factor, and two factors obtained by crossing size with operating profitability and investment, with the data downloaded from the Data Library of Professor Kenneth R. French.[1] Based on data availability of the variables under consideration, our daily sample period covers 1st February 1994 to 30th November 2024.

Following Diebold and Yilmaz (2009, 2012 and 2014), we start from a VAR model of order p for the n-dimensional vector time series $\boldsymbol{Y}_t$ that might refer to either returns or their volatilities:

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \boldsymbol{Y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{Y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{1}$$

where the error term has zero-mean and covariance equal to Σ. Then, building on Pesaran and Shin (1998), we recover the Generalized FEVD (GFEVD) over a horizon of h-periods:

$$\theta_{ij}(h) = \frac{\sigma_{jj}^{-1} \sum_{l=0}^{h} \left(\boldsymbol{e}_i' \boldsymbol{\Gamma}_l \boldsymbol{\Sigma} \boldsymbol{e}_j\right)^2}{\sum_{l=0}^{h} \left(\boldsymbol{e}_i' \boldsymbol{\Gamma}_l \boldsymbol{\Sigma} \boldsymbol{\Gamma}_l' \boldsymbol{e}_j\right)} \tag{2}$$

where σ_{jj} is the j-the diagonal element of the innovation covariance, $\boldsymbol{\Gamma}_l$ is the coefficient matrix at lag l in the Vector Moving Average (VMA) representation of the VAR(p) model, and $\boldsymbol{e}_j$ is a vector of dimension n composed of zeros, but at position j it takes value of 1. The GFEVD represents the contribution of variable j to the forecast error variance of variable i, over a horizon of h periods. The Diebold and Yilmaz (2009, 2012) Total Spillover Index for the 50 states builds on the row normalized GFEVD, denoted as $\tilde{\theta}_{ij}(h)$, and is equal to:

$$S_h = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \tilde{\theta}_{ij}(h)\,. \tag{3}$$

A second commonly used indicator derived from the GFEVD is the Net Spillover (or Directional Spillover), that allows us to evaluate individual state-level spillover, and is defined as:

$$N_{i,h} = \sum_{j=1, j \neq i}^{n} \tilde{\theta}_{ij}(h) - \sum_{j=1, j \neq i}^{n} \tilde{\theta}_{ji}(h). \tag{4}$$

[1] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Bonaccolto et al. (2024) show that the spillover indexes might be over-estimated due to the presence of non-significant parameters in the estimated VAR(p) model. They advocate for the use of General-to-Specific approaches (Sucarrat and Escribano, 2012) or the introduction of regularization (Demirer et al., 2017), both of which are being effective in reducing the index over-estimation. Moreover, they highlight that the size of the index is significantly affected by high levels of correlation in the model's innovations.

This observation led us to consider an issue that has not yet been accounted for in the stock market spillover literature, namely, the presence of common factors that could inflate the spillover indexes, especially when analyzing state-level stock indexes within the US. The technical intuition stems from the common use of the spillover indexes in the analysis of interdependence across variance proxies, as introduced in the seminal work of Diebold and Yilmaz (2009). However, if we analyze (state-level) equity data, we know that common factors exist, and that the variance of the stocks can be decomposed into a systematic part and an idiosyncratic part. Therefore, by focusing on the total variance for the estimation of the VAR model, the common factors will result in the inflating the residual correlation and, consequently, leading to the overestimation of spillover indexes. To address this issue, we propose pre-filtering the data to remove the impact of common factors. While our proposed solution is specifically tailored to the data described in the previous section, the procedure can be adapted to other datasets.

Let us denote $\boldsymbol{R}_t$ as the *50*-dimensional vector of US states equity returns, and $\boldsymbol{F}_t$ captures the set of common factors outlined above. Then, we specify the following linear model:

$$\boldsymbol{R}_t = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{F}_t + \boldsymbol{\eta}_t. \tag{5}$$

As the spillover indexes are usually evaluated on a moving window basis, we adopt the same strategy for the estimation of the linear model in (5) by using a window with size $\tau = 126$ daily observations, i.e., roughly covering 6-months, and a step of 1-day. Then, we compute the sequence of residuals as:

$$\widehat{\boldsymbol{\eta}}_t = \boldsymbol{R}_t - \widehat{\boldsymbol{\alpha}}_{t-1:t-\tau} - \widehat{\mathbf{B}}_{t-1:t-\tau}\boldsymbol{F}_t, \quad t = \tau+1, \tau+2, \dots, T. \tag{6}$$

By using a moving window approach, understandably, we lose the first six months in our dataset. As our interest is in contrasting the estimation of the spillover indexes computed over the mean and a proxy for variance that captures state-level financial risks, we compute the monthly innovations, and the monthly realized variances as follows:

$$\widehat{\boldsymbol{\eta}}_m = \sum\nolimits_{t\in m} \widehat{\boldsymbol{\eta}}_t, \quad \widehat{\boldsymbol{V}}_m = \sum\nolimits_{t\in m} \widehat{\boldsymbol{\eta}}_t^2, \quad m = 1,2,\dots,M, \tag{7}$$

where, by $t \in m$, we indicate the open market days belonging to calendar month m and M is the total number of available months. Similar quantities can also be defined for the original (un-filtered) returns.

Then, we proceed with the evaluation of the Total Spillover and Net Spillover indexes and estimate a VAR(p) model on the time series of $\widehat{\boldsymbol{\eta}}_m$ and $\widehat{\boldsymbol{V}}_m$ under regularization, as discussed above, to ensure that only statistically significant coefficients are utilized and to mitigate overestimation of these indexes. To simplify model estimation, we first determine the optimal lag order using the Bayesian Information Criterion (BIC) without including regularization. The optimal order equals $p = 1$. Then, we estimate the parameters of each VAR model by using an equation-by-equation approach resorting to a least square method with a Lasso penalty. To determine the optimal shrinkage parameter, we use a standard 10-fold cross-validation to select the optimal value by minimizing the Mean Squared Error

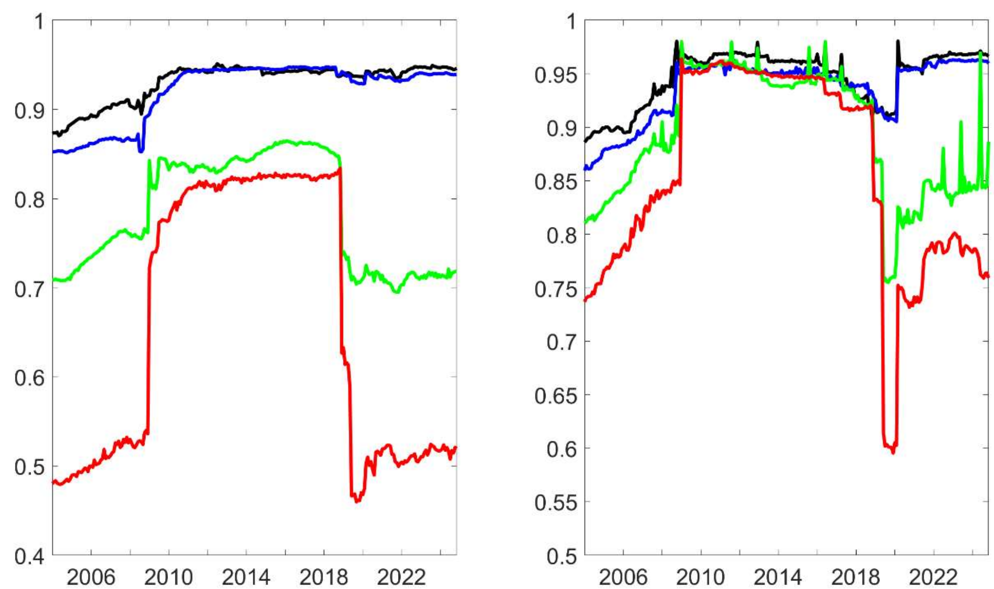
(MSE). The innovation covariance is then evaluated using the sample estimator, but we do not include regularization on the correlation across the innovations. For the VAR model, we adopt a rolling evaluation scheme with a window of size 120 months (i.e., 10 years).

In the following we focus on the evolution over time of the Total Spillover Indexes, comparing two different cases: with and without the filtering for the common factors, while applying Lasso regularization in both cases. The state-level results are presented in the Appendix and discussed briefly to save space. Intermediate results (for example, parameter estimation, shrinkage parameters) and the spillover without the Lasso regularization are available upon request from the authors.

3. Results

We first refer to the Total Spillover Index computed under four different designs, which involves combining the absence or presence of Lasso regularization and common factor filtering, as presented in Figure 1. Due to the use of a rolling method for the VAR(p) model estimation, the spillover indexes are available starting from February 2004.

Figure 1: Total Spillover Indexes for returns and realized variances



Note: Total Spillover Index for returns is plotted on the left and same for realized variances is in the plot on the right. Four different cases are considered: baseline index (black line), i.e., without Lasso regularization and without common factor filtering; Lasso-regularized index without common factor filtering (blue line); index with common factor filtering and without Lasso regularization (green line); index with common factor filtering and Lasso regularization (red line).

Notably, for both monthly returns and their realized variances, the baseline indexes (black lines) assume very high values, being, for most of the sample, above 0.9. Introducing regularization, even though we notice some reduction, during the early 2000s, the indexes do not change in a significant way. This is in line with the findings of Bonaccolto et al. (2024) and indicates the strong impact of correlation across the VAR model residuals. In fact, the correlations are, on average, 0.54 across returns and 0.64 across realized variances (averaged cross-sectionally and over time), a value that, according to the simulations in Bonaccolto et al. (2024), might lead to spillover levels even above 0.9. Nevertheless, this is also coherent with the existence of common factors, not accounted for in the VAR(p) model, thus leading to residual correlation. Therefore, we proceed to filter common factors. Notably, this step leads to spillover indexes to be lower in general, especially when regularization is added. This evidence suggests two relevant aspects. First, common factors inflate the spillover index. Second, filtering out common factors enhances the effectiveness of regularization, making the estimated spillover dynamics more accurate and reflective of the true interconnectedness of state-level markets. A peculiar case is given by the realized variances, where all the spillover indexes are close to one another over 2009 to 2018, possibly due to the fluctuations (uncertainty) in the US stock market being quite calm in general post the global financial crisis and pre-COVID-19.[2] The absence of major shocks during this time likely contributed to the convergence of spillover measures, reinforcing the importance of considering economic conditions when interpreting spillover estimates.

If we move to the evaluation of the level of the indexes, we note jumps and drops of the indexes. These are not an artifact of the factor filtering or of regularization but depend on the use of a 10-year rolling approach. When relevant turmoil periods enter the evaluation sample, spillovers tend to jump-up, as can be observed in 2009, which corresponds to the most volatile period of the great financial crisis being included in the estimation sample. Similarly, when the stock market turmoil exists from the estimation sample, the spillover indexes drop, as we note when the subprime mortgage crisis period gets excluded from the sample (i.e., 10 years after the sudden increase in spillovers). The jump in the last part of the sample which we note on the realized variance, is, understandably, due to the COVID-19 outbreak.

Overall, our findings emphasize that failing to account for common factors can lead to systematically inflated spillover indexes, potentially misrepresenting the actual degree of interconnectedness in state-level stock markets. This has critical implications for interpreting econometric results. Besides the academic perspective of this finding, higher artificial connectedness of returns and risks could mistakenly suggest to investors that there is reduced chances of portfolio reallocation across industries domiciled in the states. At the same time, it could end-up signaling higher possibility of contagion and heightened risk in the overall financial system to policymakers, when this is indeed not the case, and thus leading to possible miscalculation of policy intervention.

From a policymaking standpoint, overestimated spillover indexes could incorrectly signal heightened financial contagion risks, prompting unnecessary or mis-calibrated policy interventions. Regulatory authorities might perceive an exaggerated threat to financial stability, leading to excessive tightening of financial regulations or misallocation of resources toward mitigating systemic risk that, in reality, is less severe than indicated by the unadjusted spillover measures.

Finally, we turn our attention to the net spillover indexes, which we compute only in cases where regularization is applied, both with and without common factor filtering. These results, depicted in Figure A1 in the Appendix, for selected states, with the largest number of quoted companies (cumulating to a total of over 80%; see Table A1), further reinforce the importance of filtering common factors. In addition to expected changes in the level of net spillovers, we also observe shifts in their sign. This is particularly problematic from an empirical standpoint, as it implies a transition from being a net receiver of risk to being a net contributor to financial system risk. Such a shift could

[2] See, for example, in this regard the plot of the Chicago Board Options Exchange (CBOE)'s Volatility Index (VIX) at: https://fred.stlouisfed.org/series/VIXCLS.

fundamentally alter risk assessments and policy responses, underscoring the critical need to correctly account for common factors in spillover estimation.

4. Conclusion

Against the backdrop of a large existing literature on connectedness of stock returns and its volatilities across countries, in this paper we aim to extend this area of research by taking a regional perspective, which involves looking at such spillovers involving 50 states of the US. By filtering out common factors, and insignificant coefficients via Lasso, over the monthly period of February 1994 to November 2024, we find that that not accounting for common factors, might lead to the spillover indexes to be higher than what they should actually be in reality. This, in turn, would send incorrect information to the policymakers associated with the overestimation of the possibility of contagion and extent of risk in the financial system, and suggest to investors reduced opportunity for portfolio reallocation, when this is not necessarily the case. Understandably, filtering to obtain interdependence via spillover indexes has important implications beyond its academic value.

Our results highlight the importance of properly filtering interdependence through spillover indexes to ensure more accurate assessments of financial market risks and opportunities. Beyond its academic contributions, our study underscores the practical relevance of methodological refinements in financial modeling, offering valuable insights for policymakers, investors, and market analysts seeking to navigate the complexities of interconnected stock markets.

As part of future research, to generalize our findings, it would be interesting to investigate other within-asset-class spillovers, given the likelihood of common driving factors.

References

Bonaccolto, G., Caporin, M., and Shahzad, S.J.H. (2024). (Quantile) Spillover Indexes: simulation-based evidence, confidence intervals and decomposition. Available at SSRN: https://ssrn.com/abstract=4629224.

Bouri, E., Cepni, O., Gabauer, D., and Gupta, R. (2023). Return connectedness across asset classes around the COVID-19 outbreak. International Review of Financial Analysis, 73(C), 101646.

Chaney, T., Sraer, D., and Thesmar, D. (2012). The collateral channel: how real estate shocks affect corporate investment. American Economic Review, 102(6), 2381-2409.

Coval, J.D., and Moskowitz, T.J. (1999). Home bias at home: local equity preference in domestic portfolios. Journal of Finance, 54(6), 2045-2073.

Coval, J.D., and Moskowitz, T.J. (2001). The geography of investment: Informed trading and asset prices. Journal of Political Economy, 199(4), 811-841.

Demirer, M., Diebold, F.X., Liu, L., and Yilmaz, K. (2017). Estimating global bank network connectedness. Journal of Applied Econometrics, 33(1), 1-15.

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

Diebold, F.X., and Yilmaz, K. (2012). Better to give than to receive: predictive directional measurement of volatility spillovers. International Journal of Forecasting, 28(1), 57-66.

Diebold, F.X., and Yilmaz, K. (2014). On the network topology of variance decompositions: measuring the connectedness of financial firms. Journal of Econometrics 182(1), 119-134.

Diebold, F.X., and Yilmaz, K. (2023). On the past, present, and future of the Diebold–Yilmaz approach to dynamic network connectedness. Journal of Econometrics, 234(C), 115-120.

Korniotis, G.M., and Kumar, A. (2013). State-level business cycles and local return predictability. Journal of Finance, 68(3), 1037-1096.

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

Pirinsky, C., and Wang, Q. (2006). Does corporate headquarters location matter for stock returns? Journal of Finance, 61(4), 1991-2015.

Sucarrat, G., and Escribano, A. (2012). Automated model selection in finance: General-to-specific modelling of the mean and volatility specifications. Oxford Bulletin of Economics and Statistics, 74(5), 716-735.

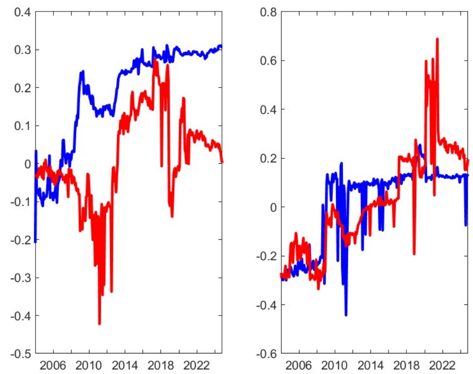
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B (Methodological), 58(1), 267-288.

Yilmaz, K. (2021). Return and volatility spillovers among the East Asian equity markets. Journal of Asian Economics, 21(3), 304-313.

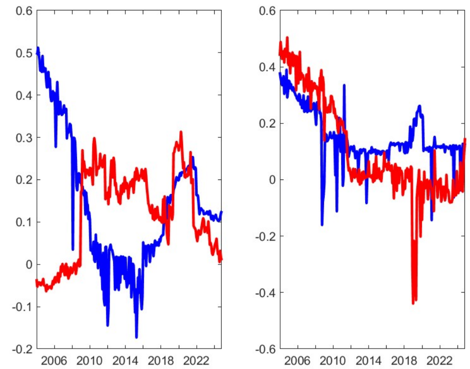
APPENDIX:

Figure A1: Net Spillover Index for returns and realized variances of selected states

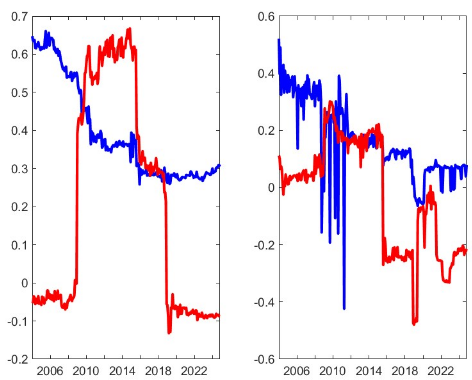
California



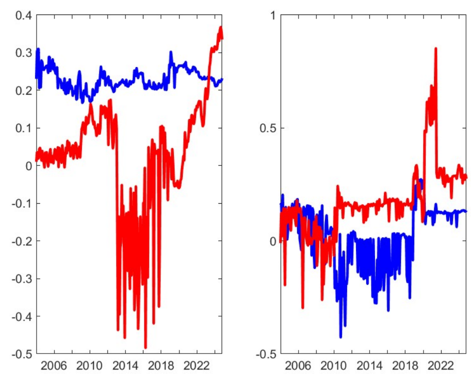
Colorado



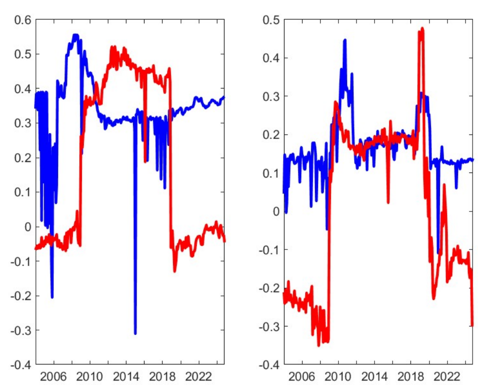
Florida



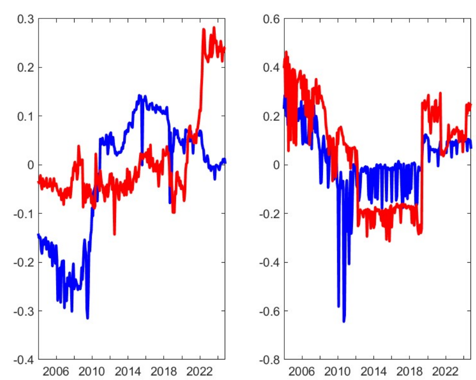
Georgia



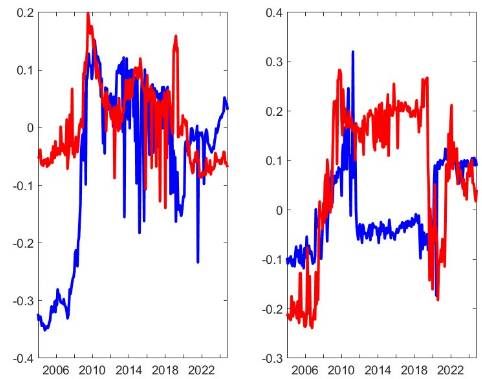
Illinois



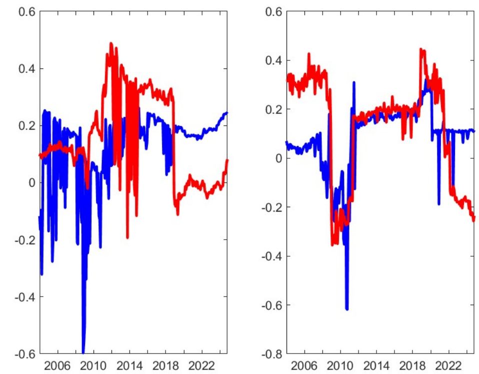
Maryland



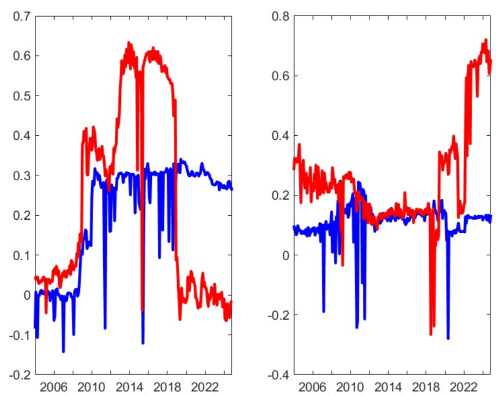
Massachusetts



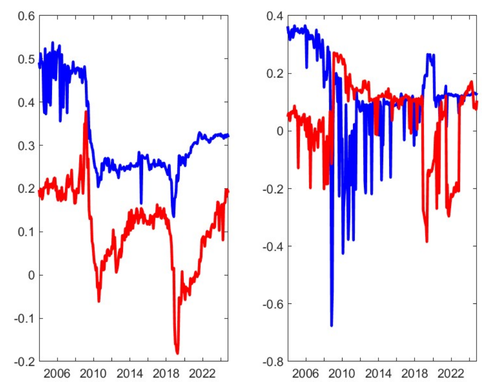
Nevada



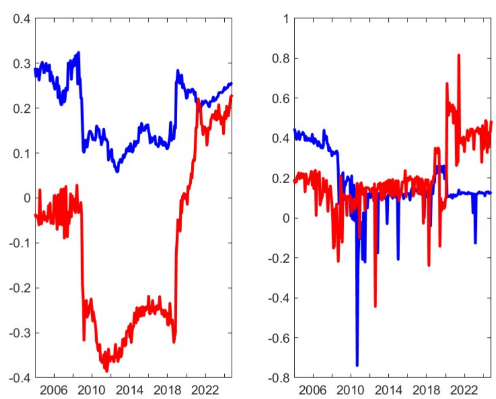
New Jersey



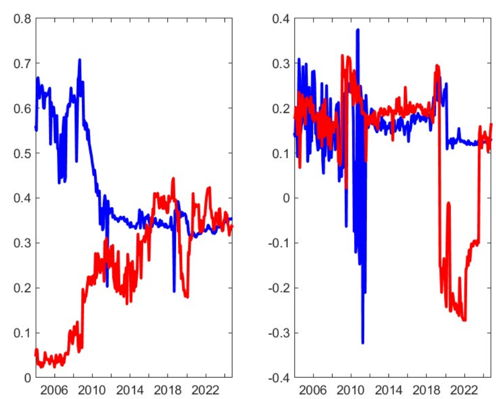
New York



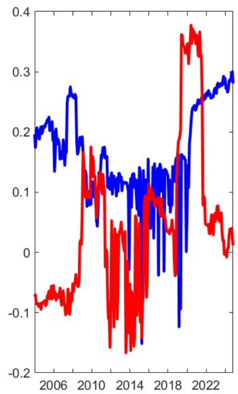
Ohio



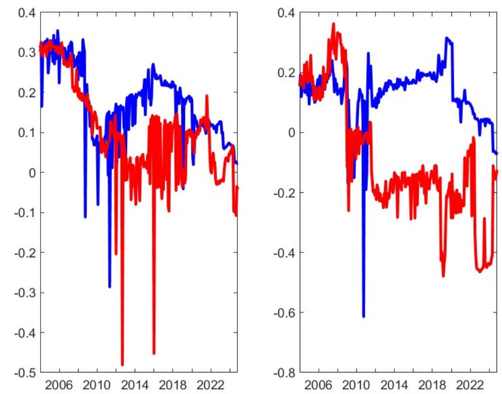
Pennsylvania



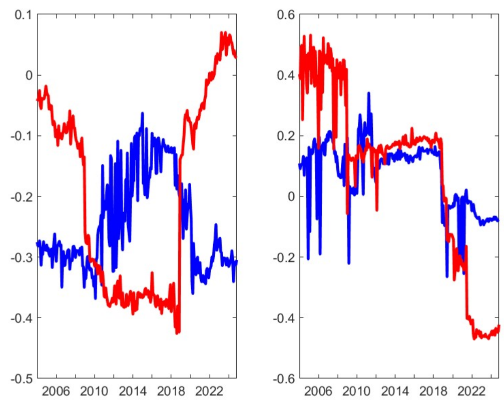
Texas



Virginia



Washington



Note: Net Spillover Index for returns is plotted on the left and same for realized variances is in the plot on the right, involving the cases of without common factor filtering (blue line) and with common factor filtering (red line). In all cases, Lasso regularization is included.

Table A1: Number of companies included in state stock index

State	Number	%	Cumulative %
Maryland	90	1.7%	80.1%
Washington	103	1.9%	78.5%
Nevada	104	1.9%	76.5%
Ohio	132	2.5%	74.6%
Virginia	132	2.5%	72.1%
Georgia	135	2.5%	69.7%
Colorado	164	3.1%	67.2%
New Jersey	184	3.4%	64.1%
Illinois	187	3.5%	60.7%
Pennsylvania	201	3.7%	57.2%
Florida	315	5.9%	53.5%
Massachusetts	321	6.0%	47.6%
Texas	505	9.4%	41.6%
New York	716	13.3%	32.2%
California	1012	18.9%	18.9%

Source: Bloomberg.