Hipp, Ruben

Working Paper

On causal networks of financial firms: Structural identification via non-parametric heteroskedasticity

Bank of Canada Staff Working Paper, No. 2020-42

Provided in Cooperation with:
Bank of Canada, Ottawa

Suggested Citation: Hipp, Ruben (2020) : On causal networks of financial firms: Structural identification via non-parametric heteroskedasticity, Bank of Canada Staff Working Paper, No. 2020-42, Bank of Canada, Ottawa, https://doi.org/10.34989/swp-2020-42

BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel — 2020-42

Last updated: October 15, 2020

On Causal Networks of Financial Firms: Structural Identification via Non-parametric Heteroskedasticity

by Ruben Hipp

Financial Stability Department
Bank of Canada, Ottawa, Ontario, Canada K1A 0G9

rhipp@bankofcanada.ca

ISSN 1701-9397

Acknowledgements

I am particularly indebted to Florian Böser, Christian Brownlees, Felix Brunner, Franco Esteban Cattaneo, Niklas Garnadt, David Hipp, Carsten Jentsch, Matthias Meier, Xiangjin Shen, André Stenzel, Carsten Trenkler, Kerem Tuzcuoglu, Luis Henrique Uzeda Garcia and various seminar participants for helpful comments.

Abstract

We investigate the causal structure of financial systems by accounting for contemporaneous relationships. To identify structural parameters, we introduce a novel non-parametric approach that exploits the fact that most financial data empirically exhibit heteroskedasticity. The identification works locally and, thus, allows structural matrices to vary smoothly with time. With this causality in hand, we derive a new measure for systemic relevance. An application on volatility spillovers in the US financial market demonstrates the importance of structural parameters in spillover analyses. Finally, we highlight that the COVID-19 period is mostly an aggregate crisis, with financial firms' spillovers edging slightly higher.

Bank topics: Econometric and statistical methods; Financial markets; Financial stability
JEL codes: C1, C3, C32, C58, L14

1 Introduction

Most econometricians base their studies on questions concerning the causality of economic associations. When interpreting the results, though, they shy away from the term "causality" because correlations, significant or not, do not imply causation. Econometricians do so because if it is clear what is the cause and what is the effect, causality is implicitly evident. This particular idea of quantifying causality was first mentioned by Wright (1921) with the usage of path diagrams, which illustrate the chain of causation. However, directions of causality are not always apparent. For example, if we are interested in both directions, and all variables are cause and effect simultaneously, identification without prior information proves to be difficult.

In this study, we want to answer the question concerning two-way causality to unveil the Systemic Relevance of financial firms. Take the case of two banks that connect via their asset and liability structures. Namely, bank i owes money to bank j, such that a default of i will cause j to default. Further, assume that j is debt-free and that the reverse is not true. For example, for default probabilities, we observe that when i's risk increases, j's does as well. If this response occurs in the same period, we measure only an undirected correlation. The absence of the opposite effect, then, only lowers this correlation, and we remain in the dark about the direction of causality. However, for policymakers, it is crucial knowledge that i is the systemically more critical bank. Thus, they require methods to disentangle the two effects.

While pairwise associations between a single bank and the market are a useful beginning to understand systemic issues, they are of limited value because the network architecture remains latent (see Acemoglu et al., 2015). More precisely, the question of a system's financial stability is closely related to the estimation of a unified casual network. In principle, multivariate time series analysis estimates this full set of dependencies. However, responses can be captured only when they occur lagged. Less-frequent data almost always miss out on prompt responses and, thus, we measure responses only as undirected co-movements or correlations. Employing correlations neglects the causality not only of the first response but also of all following ones. Consequently, long-term dependencies, e.g., forecast error variance decompositions (FEVDs), are also not fully understood. Overall, a key challenge to measuring a unified causal network is the identification of contempora-

neous responses.

In this paper, we introduce a novel approach that identifies contemporaneous dependencies. This approach exploits the fact that most financial data exhibit heteroskedasticity. That is, the covariance matrix of the error terms varies over time. The evolution is free of any functional form and is only parameterized by fitting a local time trend. To decompose the covariance matrix into its structural components, we impose two separate assumptions on the time variation of the connectedness parameter and the idiosyncratic volatility of structural shocks. This way, we can identify structural parameters if the connectedness of shocks alternates reasonably more slowly than their idiosyncratic volatility. While the assumptions are slack, the necessary local approximations add to uncertainty and do require frequent data. Unlike previous work, this local identification improves on the assumption of connectedness by allowing it to be dynamic. Eventually, we introduce a new intuitive centrality measure, which is compatible with the systemic importance of financial firms.

The estimation of structural parameters and return volatility spillovers offers new insights into US financial firms. The results suggest that average idiosyncratic risks of banks shrank over the sample, with connectedness edging a bit higher in the end. Moreover, peaks in spillovers coincide with financially stressful events. That is, all peaks but the global financial crisis (GFC) of 2007-09 also have increased spillovers from the market to financial firms. Factoring in contemporaneous responses leads to more accurate and reasonable results, in particular, for the bailout of the American International Group (AIG) in September 2008. The snapshot analysis reveals that all firms received strong spillovers from AIG. Finally, we highlight the Systemic Relevance of institutions over the last two decades. As of June 2020, we find that JPMorgan Chase is the systemically most relevant institution, suggesting that it warrants further monitoring by the supervisory authorities.

Our study is related to two strands of the literature: the empirical studies of connectedness, and the identification of structural shocks.

While earlier literature measures dependencies as pairwise associations, e.g., Adrian and Brunnermeier (2011) and Brownlees and Engle (2012), the whole concept of connectedness as a network was first addressed by Diebold and Yılmaz (2014). Their main claim is the interpretation of FEVDs as networks. However, this framework considers only contemporaneous relations as undirected correlations. If most significant reactions occur contemporaneously, this method is prone to misspecifications since it estimates only slower

reactions. For daily market-based data, we expect a majority of severe responses on the same day, i.e., contemporaneously. Similarly, causality has been tackled by Billio et al. (2012), who provide a binary network with entries based on positive Granger causality tests. These tests are carried out for multiple lags and, hence, also ignore contemporaneous directions. In contrast, Barigozzi and Brownlees (2019) use equation-wise LASSO-type techniques to estimate a sparse causal contemporaneous directed network. Yet, a non-sparse version remains undetected. De Santis and Zimic (2017) directly estimate a network by employing a structural vector autoregression (SVAR), which is, however, only set-identified.

With the growing literature on economic networks and the resurrection of graph theory, it is straightforward to think about causality again. Causal relationships go back to Sewall Wright's (1921) paper on correlation and causation, in which he uses path diagrams to explain causality. His pioneering work demonstrates that correlation approaches mostly miss out on the knowledge we possess. That is, a priori ideas about the chain of causation support that correlation measures causality. In economics, the question of contemporaneous causal relations is equivalent to the identification of structural shocks. Such shocks are generally interpreted as unexpected, uncorrelated, and exogenous innovations with economic interpretation. They are identified when we possess the right knowledge, e.g., when we impose some restrictions. For example, in the case of *unambiguous directions* or one-way causality, causal assumptions are not about the links but rather about the missing links. In the SVAR literature, triangularization as in Sims (1980) and long-run restrictions as in Blanchard Olivier and Quah (1989) are the most prominent examples of the exclusion of effects. These equality constraints, however, are too restrictive for many economic applications. Therefore, set-identification of structural shocks by inequality restrictions, e.g., Uhlig (2005), are popular in applications.[1]

Another strand of the literature uses *counterfactual* reasoning. Typically, economists use knowledge about causality to get an idea of what an alternative (counterfactual) world would have been like. While this reasoning is deductive, the literature on heteroskedasticity identification uses induction to quantify causation. Precisely, observing two states of the world, "factual" and "counterfactual" in the form of changing volatility, can identify

[1]Fry and Pagan (2011) point out the advantages and disadvantages of identification via inequality restrictions and highlight explicitly that the estimations are not interpretable with probabilistic language.

structural shocks (see Rigobon and Sack, 2003). This reasoning generated attention since the presence of heteroskedasticity in the form of time-varying volatility is uncontroversial in most applications. Prominent extensions such as Markov-switching processes by Lanne et al. (2010) and generalized autoregressive conditional heteroskedasticity (GARCH) by Milunovich and Yang (2013) model heteroskedasticity parametrically absent external information. To avoid restrictive functional form assumptions from Markov-switching and GARCH models, Lewis (2017) uses non-parametric heteroskedasticity to identify a finite set of possible solutions. In contrast, the non-parametric approach in our paper point-identifies structural parameters and, additionally, allows for time variation in the response matrix.

The remainder of the paper is organized as follows. Section 2 first sets up the mathematical framework, states identifying assumptions, and offers a likelihood-based extremum estimator. In Section 3, we introduce a causal network model and show how we claim to quantify causality. In Section 4, we give further insights into causal financial connectedness in the US and focus on the spillovers between financial institutions and market components. Section 5 concludes.

2 Theoretical considerations

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

$$A_{0,t}u_t = B_t\epsilon_t, \quad \forall t = 1, ..., T, \tag{1}$$

where the structural matrix $A_{0,t}$ is a real valued parameter matrix of size $(N \times N)$ with full rank and unit diagonal. B_t is a diagonal matrix with real valued positive entries. The structural shocks ϵ_t have a multivariate distribution with mean zero and unit variance. The unit diagonal of $A_{0,t}$ and the diagonal structure of B_t ensure that all connections of the N variables are in the off-diagonal of $A_{0,t}$. Heteroskedasticity is now included by $B_t\epsilon_t$ and is, without further restrictions, unconditional. Moreover, since $A_{0,t}$ is also assumed

to be unconditional, we can use multiple estimation techniques (e.g., general methods of moments, likelihood methods, or general least squares).

Note that the basic structural equation reads $u_t = S_t\epsilon_t$ with S_t as the structural matrix. Both problems can be regarded as equivalent since diagonality of B_t and the unit diagonal of $A_{0,t}$ ensure that there exists a unique decomposition of $S_t = A_{0,t}^{-1}B_t$.[2] The decomposition into $A_{0,t}$ and B_t allows us to impose different assumptions and will be essential for heteroskedasticity identification. The restrictions we consider focus on the time evolution of matrix entries and are generally in line with recent literature on time-varying parameters.[3]

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

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

Further,

***(A1) locally linear volatility**: B_t^{-1} has a time derivative or time trend different from zero.*

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

The smoothness condition (A0) implies that parameters drift slowly with time, which enables consistent estimation and allows for local approximations. Moreover, (A1) and (A2) ensure identification of parameters. In fact, the identifying assumption is the difference in the convergence (alteration) rate. By (A0), the heteroskedastic volatility parameter B_t is assumed to have an asymptotic derivative, which (A1) ensures is different from zero. In contrast, (A2) states that $A_{0,t}$ changes more slowly, such that its derivative or time trend is negligible. In a nutshell, we expect volatility depicted by B_tB_t' to evolve faster than connectedness in $A_{0,t}$.

Assumption (A0) is a generalization of the standard assumptions for locally stationary processes. Namely, in the work of Dahlhaus et al. (2006), the parameter $\theta_t = \theta(t/T)$ is assumed to be smooth, deterministic and piecewise differentiable. In contrast, (A0) also

[2]Note that this decomposition will later be problematic due to the normalization. Without further restrictions, the identification will only be up to column reordering of $A_{0,t}$.

[3]See, e.g., Giraitis et al. (2016).

allows for stochastic parameter processes but ensures the necessary degree of persistence in the entries.[4] Assumptions (A1) and (A2) are further specifications under the smoothness condition. While B_t^{-1} is linear with a non-zero gradient for sufficiently small segments, $A_{0,t}$ is constant on the same segment. Intuitively, variations in B_t dominate variations in $A_{0,t}$, and thus we can neglect the variations of the latter. In sections 2.1 and 2.2, we make clear how this preponderance comes into play.

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

2.1 Identification via non-parametric heteroskedasticity

Since this approach is intended to work under parsimonious conditions, identification and estimation of parameters have two conceptional tasks. First, identification of structural parameters should proceed absent any equality restrictions and parametric assumptions, and second, time-varying parameter estimation should be prior-free. To overcome these challenges, we double use the locally constant and locally linear assumptions for both tasks. That is, identification works under assumptions (A1) and (A2), and estimation requires (A0) and (A2). In doing so, we take the idea of Rigobon and Sack (2003), who point out that structural parameters are identifiable when models exhibit heteroskedasticity. More precisely, since B_t must have at least two distinct values in (1), it is time-varying.

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

[4] Since B_t is diagonal, its inverse is a diagonal matrix with the entries inverted.

The missing parametric assumptions from the non-parametric approach challenge structural identification. To gain additional information in parametric form, we exploit the assumption of asymptotic differentiability and, thereby, we remain in a non-parametric environment. Precisely, this assumption allows us to apply a derivative process to model heteroskedasticity and fit a locally weighted kernel estimator. The local optimization of an objective function with respect to the functional value and its derivative gives us additional knowledge about the drift of the parameters. Such drifts can be estimated and, therefore, parametrized.

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

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

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

We target the inverse covariance matrix (also called the concentration or precision matrix) in the Taylor expansion since it obtains a more elegant representation than the covariance matrix. Its entries relate to the contemporaneous correlations between two variables conditional on others. More importantly, $A_{0,t/T}$ is non-inverted at the cost of an inverse $B(t/T)$, which is easier to take derivative since it is diagonal. Thus, we find it easier to comply with the assumptions. Lastly, targeting the concentration matrix allows for a more appealing objective function but does not affect estimation and identification.

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

$$\begin{aligned}
\Sigma_\tau^{-1}(t/T) &\approx \Sigma_\tau^{-1} + (\frac{t}{T} - \tau)\dot{\Sigma}_\tau^{-1}, && (2)\\
\Sigma_\tau^{-1} &= A'_{0,\tau}B_\tau^{-2}A_{0,\tau} && (3)\\
\dot{\Sigma}_\tau^{-1} &= \frac{\partial(A'_{0,\tau}B(\tau)^{-2}A_{0,\tau})}{\partial\tau}(\tau)\\
&= \dot{A}'_{0,\tau}B_\tau^{-2}A_{0,\tau} - 2A'_{0,\tau}B_\tau^{-3}\dot{B}_\tau A_{0,\tau} + A'_{0,\tau}B_\tau^{-2}\dot{A}_{0,\tau}, && (4)
\end{aligned}$$

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

$$\lim_{T\to\infty}\frac{\theta(\tau + {}^t\!/_T) - \theta(\tau)}{{}^t\!/_T} = \frac{\partial\theta}{\partial\tau}(\tau).$$

By assumption (A0), this limit exists, and therefore the derivative is asymptotically defined. Additionally, (A2) ensures that the derivative of $A_{0,\tau}$ goes to zero with convergence rate T. For the sake of clarity, derivatives of $A_{0,\tau}$ will be dropped since assumption (A2) ensures asymptotical negligibility. Equation (4) becomes

$$\dot{\Sigma}_\tau^{-1} = -2A'_{0,\tau}B_\tau^{-3}\dot{B}_\tau A_{0,\tau}. \quad (5)$$

Intuitively, the assumptions allow us to attribute all variation in Σ_τ to variations in B_τ; hence, they provide additional information about $A_{0,t}$. The mapping from the structural parameters to the reduced-form ones is consequently given by (3) and (5), where there are $N(N+1)$ reduced-form parameters in $(\Sigma_\tau^{-1}, \dot{\Sigma}_\tau^{-1})$ and $N(N+1)$ structural parameters in $(A_{0,\tau}, B_\tau, \dot{B}_\tau)$.

In the Appendix, Proposition 1 proves identification of the mappings up to sign reversal and column reordering. The sign reversals can be thought of as stemming from the square component B_τ^{-2}. We can easily solve this issue by restricting its elements to be positive, which in turn imposes sign restrictions on $\dot{B}_\tau$. The labeling of shocks can then, if needed, be checked by the narrative of the story. Also, note that the identification, which is only up to column reordering of $A_{0,\tau}$, parallels the result in Lewis (2017). However, in a

network context, we can impose magnitude restrictions to make it uniquely identified. In Proposition 2 we show that these restrictions are sufficient for global identification.

In a nutshell, identification results from the mappings in (3) and (5) and, therefore, estimation of the structural parameters is possible if we can find estimates for Σ_τ^{-1} and $\dot{\Sigma}_\tau^{-1}$. Note that, in contrast to previous approaches, identification works for one observation τ and thus is independent of others. Not only does this peculiarity allow for time variation in both parameters, but it also leads to a local estimation function, which is well tractable.

The identification works under the condition $\dot{b}_i/b_i \neq \dot{b}_j/b_j$ for all $i \neq j$, which assumes that the relative derivatives of the structural variances are never the same for two variables. Considering that it is highly unlikely that relative marginal changes of structural variances match the same value, we assume that it holds for most heteroskedasticity applications with (A1) and (A2) fulfilled. Nevertheless, we are aware that weak identification might become an issue if two relative derivatives come close. Weak identification can stem from two sources. First, if assumption (A2) is not fulfilled, we erroneously ascribe time variation of $A_{0,\tau}$ to B_τ. This case in particular is problematic because we are unable to make clear predictions about the resulting bias. Second, if heteroskedasticity in the form of B_τ is insufficiently fulfilled, global identification in the context of Proposition 2 is invalid.[5]

2.2 Extremum estimator

In this section, we propose an estimator for the locally constant, locally linear assumptions. Recall that, in contrast to previous heteroskedasticity literature, we only require parameters at one observation for identification. This idiosyncrasy allows us to estimate the structural parameters point-wise. In doing so, we tackle a weakness of previous papers that need to assume that $A_{0,t}$ is constant over multiple periods.

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

By plugging in the structural parameters in the log likelihood for normally distributed

[5]For a more detailed discussion about weak identification in the setup of heteroskedasticity, we refer to Lewis (2018).

errors, we get

$$\begin{aligned} l(u_t|\theta_\tau) &= -\frac{N}{2}\ln|2\pi| - \frac{1}{2}\ln|\Sigma_\tau(\frac{t}{T})| - \frac{1}{2}u_t'\Sigma_\tau^{-1}(\frac{t}{T})u_t, \\ &= -\frac{N}{2}\ln|2\pi| + \ln|A_{0,\tau}| + \frac{1}{2}\ln|B_\tau^{-2}(I_N - 2(\frac{t}{T}-\tau)\dot{B}_\tau B_\tau^{-1})| \\ &\quad -\frac{1}{2}u_t'A_{0,\tau}'B_\tau^{-2}A_0u_t - \frac{1}{2}u_t'A_{0,\tau}'(-2)(\frac{t}{T}-\tau)\dot{B}_\tau B_\tau^{-3}A_{0,\tau}u_t, \end{aligned}$$

with $|\cdot|$ denoting the matrix determinant. Note that we take the log likelihood of any time point t with parameters at τ. This step is necessary for local estimation techniques and requires us to have approximations for $t \neq \tau$. In particular, we use (2) to get an idea of other time points. Although the setup implies unconditional covariance matrices, we plug in the conditional moments in the likelihood.

Due to the nature of Taylor approximations and local smoothing estimators, we choose an extremum estimator as in Giraitis et al. (2016). This estimator is similar to Fan et al.'s (1995) locally weighted quasi maximum likelihood (QML) estimator. In contrast to a standard QML, the extremum estimator estimates locally due to different weights for log-likelihood realizations. Precisely, it weights residual log likelihoods with a pre-specified kernel $K_h(x) = K(x/h)/h$ and bandwidth h. The kernel is a symmetric continuous bounded function with compact support and normalized to one. Moreover, the bandwidth h satisfies $h \to \infty$ and $h = o(T^{1/2})$. The log likelihood at t has weight $K_h(t/T - \tau)$ for the estimate at τ such that the extremum estimator reads

$$L_\tau(\theta_\tau) = \sum_{t=1}^{T} K_h(\frac{t}{T}-\tau)l(u_t|\theta_\tau). \tag{6}$$

We reformulate (6) with the properties of the Kronecker product, determinant, and inner product to obtain

$$\begin{aligned} L_\tau(\theta_\tau) &= c + \ln|A_{0,\tau}| - \ln|B_\tau| + \sum_{t=1}^{T} K_h(\frac{t}{T}-\tau)\frac{1}{2}\ln|I_N - 2(\frac{t}{T}-\tau)\dot{B}_\tau B_\tau^{-1}| \\ &\quad -\frac{1}{2}\text{trace}(\tilde{\Sigma}_\tau A_{0,\tau}'B_\tau^{-2}A_{0,\tau}) + \text{trace}(\tilde{\dot{\Sigma}}_\tau A_{0,\tau}'\dot{B}_\tau B_\tau^{-3}A_{0,\tau}), \end{aligned} \tag{7}$$

with

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

where c is a constant term encompassing $-\frac{N}{2}\ln|2\pi|$, and $U = [u_1, ..., u_T]$ is the matrix of realizations of u_t. Note that $\tilde{\Sigma}_\tau$ and $\tilde{\dot{\Sigma}}_\tau$ represent the local (least square) estimates with a kernel weighting. Finally, optimizing (7) with respect to $(A_{0,\tau}, B_\tau, \dot{B}_\tau)$ obtains the estimates for the structural parameters.

To build intuition, take, for example, the case of an equally weighted kernel, i.e., full sample or rolling windows. The estimator then assumes that the connectedness matrix A_0 is constant over the sample. Similarly, it assumes that the volatility matrix B is a linear function in this sample. In contrast to first heteroskedasticity identification schemes, the approach does not need any specification about regime changes; rather, it works with the slope coefficient of the volatility parameter. Re-scaling to an incremental basis with the infill asymptotics, then, allows us to approximate this procedure at all time points. The Taylor expansion of $\Sigma_\tau(t/T)$ around τ yields an error at observations different from τ. In the local polynomial estimation, this error is asymptotically negligible by the smoothness assumption (A0) but adds up to the estimation uncertainty.[6]

3 Causal networks and measures of connectedness

3.1 The causality of forecast error variance decompositions

We model a group of financial firms and stack respective observables with all other potential causes in an N dimensional time series vector y_t. Then, business between firms in the form of contractual obligations creates dependencies. Further, and maybe most important, intangible factors, such as trust in the context of lending and liquidity, add to the relationship. These dependencies show, for example, how gains, losses, and risks of firms are dependent on the financial success of others. The full set of financial dependencies spans a structure that we want to estimate to deduce systemic issues. We interpret this

[6]The term $\sum_{t=1}^{T} K_h(\frac{t}{T}-\tau)(l(u_t|\theta_t) - l(u_t|\theta_\tau))$ is asymptotically negligible due to similar arguments as in Giraitis et al. (2016). Clearly, the kernel automatically gives less weight to τ^* distant from τ and comprises the approximation errors for Σ_{τ^*}. Note that by construction, Σ_t also fulfills (A0).

structure as a network and want to label its entries with causality. More precisely, we aim to obtain various key figures resulting from the estimation of a unified set of dependencies.

This section aims to overcome the fear of using the label "causal" in entries of a network. In the econometrics literature, the term "causality" is tightly connected to the identification of quasi-experiments through exogenous variations (see Angrist and Pischke, 2009). While experiment-like treatments can be found in many micro-econometric applications, it proves to be impossible to find policy interventions that comply with time variation and simultaneity of causal relationships. Thus, we deviate from this approach and instead opt for a more basic establishment of causation. Specifically, we say a causal effect occurs if variation in one variable is followed by variation in another, everything else equal. Thus, we aim to answer the question of how much variation of i can be explained by variations in j.

Plainly speaking, we quantify contributions of one variable to another. To impart a network character to this concept, we analyze all effects by decomposing the total variation in single contributions. The resulting matrix shows the percentage contribution by each variable. In addition, we allow for different response times by including lagged observations. That is, the variation can be seen in the forecast variances. Mathematically, we express the forecast error variance decomposition as[7]

$$\text{FEVD}_t(y_{i,t+H}|y_{j,t}) = \frac{\text{FEV}_t(y_{i,t+H}|\epsilon_{j,t}=1)}{\text{FEV}_t(y_{i,t+H}|\epsilon_t = 1_{N\times 1})}, \tag{8}$$

where $\text{FEV}_t(\cdot)$ is the forecast error variance conditioned on the knowledge at t. $\text{FEV}_t(y_{i,t+H}|\epsilon_{j,t} = 1, \epsilon_{-j,t} = 0)$ is the forecast error variance of i conditioned on a one-standard-deviation shock to j. $\text{FEV}_t(y_{i,t+H}|\epsilon_t = 1_{N\times 1})$ is the total forecast error variance of j, i.e., the mean squared error of i-th H-step-ahead forecast. This ratio, then, explains the contribution of j on the H-step forecast error variance of i. And with that we can answer the question of how much variation in i is explained by variations in j.

Diebold and Yılmaz (2014) argue that arranging a new matrix $D_t^H = [d_{ij,t}^H]$ with entries $d_{ij,t}^H = \text{FEVD}_t(y_{i,t+H}|y_{j,t})$ gives us a network. This network is essential for connectedness analyses to understand how much variation of a single firm is externally driven. Conceptually, it relates closely to the problem of systemic risk and, with the interpretation of

[7]From now on, we drop the notation of τ and denote all time-varying parameters with subscript t.

causality, allows for more sophisticated policy decisions by the regulators. To label its entries with causality, we need to be sure to account for all significant causes. Namely, leaving out effects translates into misspecifications and false results.

3.2 A causal network model of financial firms

Assuming y_t includes all relevant causes, we have to account for only two types of responses, contemporaneous and lagged. For both response types, we have to argue for their causal nature as we aim to have a full network allowing simultaneous causation. Again, we explain the idea of causality with the example of two banks. Assume that bank i causes j but not vice versa. If the reaction to bank i's default happens lagged, we can estimate the responses in a vector autoregression (VAR). The irreversibility of consecutive events implicitly rules out effects from the future to the past. Even though auto-correlations are rather a statistical construct, they are sufficient for the interpretation of causality if they are not confounded.

Similar to Diebold and Yılmaz (2014), we select a VAR(3) with daily observations of volatility as a baseline model. The common practice is to use the generalized version of the FEVD, which approximates contemporaneous effects with undirected correlations. This approach is a good enough approximation if we expect negligible contemporaneous responses. In most financial setups, it is highly conceivable that the strongest reactions to adverse shocks occur on the same day. By contrast, a structural VAR obtains contemporaneous coefficients and therefore implies a decomposition of the covariance matrix. Further, the structural representation theoretically labels exogenous shocks so that we can see them as emanating from the respective source. Thus, our model leverages the methodology of Section 2.1 and incorporates contemporaneous two-way causality via heteroskedasticity.

Retake the example of two banks, but assume that the defaults happen within the same observation period. Heteroskedasticity translates to observing two states of the world. For example, in one state i defaults first and j as a result of that. In the other state, we observe only j's default. From the first state, we infer that there exists a relationship since both banks default. From the second, we only see j defaulting, and we infer that there is no effect from j to i. Thus, the effect must be from i to j. Statistically, if we observe enough responses for every state, we can estimate them. In other words, we have two worlds with, by assumption, the same dependency structure. Thus, to reason for causal relationships,

we see one world as the "counterfactual" of sorts to a factual world. Heteroskedasticity identification identifies linear causal coefficients.

Additionally to the VAR(3), we nest a network to obtain a structural model. That is, we extend the VAR(3) of Diebold and Yılmaz (2014) by a contemporaneous directed network to give it a cleaner causal interpretation. The structural model reads

$$\begin{aligned} y_t &= \alpha_t + G_t y_t + A_{1,t} y_{t-1} + A_{2,t} y_{t-2} + A_{3,t} y_{t-3} + B_t \epsilon_t, \quad \forall t = 4, ..., T \quad (9) \\ (I_N - G_t) y_t &= \alpha_t + A_{1,t} y_{t-1} + A_{2,t} y_{t-2} + A_{3,t} y_{t-3} + B_t \epsilon_t, \end{aligned}$$

where G_t is an $(N \times N)$ adjacency matrix with zero on the diagonal. The network G_t contains nodes and links representing firms and dependencies, respectively. We assume it to be directed and weighted, i.e., G_t is non-symmetric and has non-negative values, to ensure global identification. $A_{1:3,t}$ are real valued time-varying autoregressive matrices within the boundaries of stationarity. B_t is an $(N \times N)$ diagonal matrix representing the idiosyncratic risks, and ϵ_t is the normalized structural shock vector. α_t is a vector of intercepts.

Within this structural model, we ensure global identification since G_t is assumed to have values between 0 and 1. Recall that it would be sufficient to restrict its entries to be in (-1,1). However, to ensure a network interpretation, we restrict G_t to be non-negative. Since $A_{0,t}$ has ones on the diagonal by assumption, we can write the structural equation

$$\begin{aligned} u_t &= A_{0,t}^{-1} B_t \epsilon_t, \\ A_{0,t} &= (I_N - G_t). \end{aligned}$$

Clearly, the existence of Σ_t implies that $A_{0,t}$ is invertible. Hence, the geometric series $(I_N - G_t)^{-1} = \sum_{k=0}^{\infty} G^k$ exists and is finite. Accordingly, the maximum magnitude of G_t's eigenvalues, i.e., the spectral radius $\rho(G_t)$, is smaller than 1. With all entries being non-negative, G_t only has values in $[0, 1)$. Then G_t is an adjacency matrix. Estimating equation (9) allows us to determine (8). Appendix B shows the calculation of structural FEVDs and gives insights into their mathematical interpretations.

Networks appear in three forms since various responses help to understand different kinds of dependencies. First, the "causal" network G_t depicts immediate reactions or

direct dependencies. It is inexact but perhaps helpful to see this matrix as a within-seconds reaction type. Second, $A_{0,t}^{-1} = (I_N - G_t)^{-1}$ is the "impact" network, which appears contemporaneously. It quantifies the contagion result at the one-step forecast errors u_t. Loosely speaking, this matrix shows the end-of-the-day outcome of chain reactions. Third, the "spillover" network shows the dynamic propagation of shocks after H days. In the upcoming analysis, we focus mainly on the latter as it combines the reactions of the others.

3.3 Measures of connectedness

Visualization of the results proves to be difficult for this analysis since we estimate $T - p$ matrices for each network. To provide useful insights into systemic issues of the network, we define connectedness measures in this section. The measures generally apply to all network matrices: the causal network matrix G_t, the contemporaneous impact matrix $A_{0,t}^{-1} B_t$, and the H-step FEVD D_t^H outlined in Appendix B.

In analogy to Diebold and Yılmaz (2014), we add "average connectedness" to analyze spillover effects over time. For a network matrix $M = [m_{ij}]$, we define

$$C(M) = \frac{100}{N} \sum_i \sum_{j \neq i} m_{ij}. \qquad \text{(average connectedness)}$$

The average effect a firm receives multiplied by 100 represents the average connectedness and provides information about the overall spillover. A value of 50 depicts that 50% of a firm's variation is attributable to shocks at other firms. In the empirical section, this measure applies only to the FEVD and is not further presented for the other two network matrices.

Moreover, we are concerned with measuring the centrality of firms within the network. We want to account for characteristics of the network architecture, since considering only the outgoing or ingoing effects neglects the overall shape of the network. We add a new model-derived centrality measure in the context of financial risk, which allows for an economic interpretation. First, note that the impact matrix is a result of contagion through the causal network G_t. In one period, the reactions through the causal network G_t mathematically occur infinitely often. Precisely, the contemporaneous reactions of y_t to a shock are depicted by the inverse structural matrix $A_{0,t}^{-1} = (I_N - G_t)^{-1}$. As mentioned before, the inverse is the geometric sum $\sum_{k=0}^{\infty} G_t^k$.

Now, for risk mitigation, regulators are primarily interested in the contagion result. While regulatory interventions relate to changes in the causal network G_t (and potentially delayed responses), the contagion result is in the impact matrix. Therefore, we account for this fact by constructing an artificial control experiment.[8] That is, we treat one variable i as the placebo, i.e., all outgoing causal connections of i are set to zero. In algebra, the placebo variable has a zero column in the causal network G_t. The counterfactual causal and impact networks are denoted by $G^*_{t,-i}$ and $A^{-1*}_{0,t,-i}\, B_t = (I - G^*_{t,-i})^{-1} B_t$, respectively. The difference in the average impact is defined as the "Systemic Relevance" of i:

$$\Delta \bar{C}^s_{i,t} = \bar{C}\left((I_N - G_t)^{-1} B_t\right) - \bar{C}\left((I_N - G^*_{t,-i})^{-1} B_t\right), \qquad \text{(Systemic Relevance of } i\text{)}$$
$$\text{with} \qquad \bar{C}(M) = \frac{100}{N} \sum_i \sum_j m_{ij}.$$

Note that this artificial control experiment corresponds to a regulator's concessions about the safety of an institution's obligations. In particular, because the institution's obligations are secured, other institutions do not receive any more risk from it. Then $\Delta \bar{C}^s_{i,t}$ measures the reduction in average risk exposure if i was a non-emanating variable. Amplifications in the impact matrix decrease more if institution i is systemically more relevant, and hence bigger values of this measure relate to a higher Systemic Relevance. Note that it is possible to extend this idea to the lagged components analogously.

4 Connectedness in the US financial system

4.1 Data

For an analysis of financial dependencies and spillovers, high-frequency balance sheets and other frequently updated obligations are ideal. However, the lack of such data forces us to estimate dependencies with publicly observable data. We use market-based data since they reflect the expectations of many strategically acting investors. As in most studies about spillovers, we study volatility to track investors' fear. The resulting connectedness represents the investor-anticipated dependencies of risk, gains, and losses of firms. Moreover, we expect that investors react more severely according to the dependency structure

[8] Note that we can construct this experiment because we can construct counterfactuals due to the causality of the matrix.

of firms if they see those as critical. We follow Parkinson (1980) and measure volatility with the extreme value estimation,

$$\tilde{\sigma}_{i,t}^2 = 0.3607(p_{i,t}^{high} - p_{i,t}^{low})^2,$$

where $p_{i,t}^{high/low}$ denotes the daily maximum and minimum log prices, respectively. The extreme value estimation is particularly appealing due to its simplicity, and we believe it to be more robust to recording errors.[9] Martens and Van Dijk (2007) point out that realized ranges like this are more efficient and more robust to microstructure noise.

To identify the contemporaneous network, assumptions (A1) and (A2) must hold.[10] First, (A1) implies that the exogenous shock vector $B(t)\epsilon_t \sim (0, B(t)B(t)')$ has covariance matrix $B(t)B(t)'$ with a trend component. We expect shocks on realized volatility to have time-varying variance in the form of stable processes. It is conceivable that this component comes from an investor sentiment cycle and, thus, we assume (A1) to be satisfied. Second, (A2) implies that entries in G_t evolve smoothly and have slower alterations than B_t. We argue that this network is a result of dependencies caused by contractual obligations between financial firms. Stacks of such obligations constantly grow and shrink as old contracts mature and new ones are signed. Since one contract is just a small increment of the stack, the network is assumed to be constant for sufficiently short periods. Thus, we see both assumptions fulfilled.

To provide an economically meaningful interpretation of the orthogonal shocks, we aim to include a full market in its unity. Therefore, we focus on the US financial market, which is arguably the financial market taking the fewest foreign shocks. However, because numerical optimizations and estimation uncertainty suffer in a high-dimensional setup, we limit the analysis to the most important institutions and shocks. We include the six largest banks in the US: JPMorgan Chase, Citigroup, Bank of America, Wells Fargo, Goldman Sachs, and Morgan Stanley. We also include one smaller bank: Bank of New York Mellon. Note that smaller banks might still have significant effects on bigger banks if they are non-primary dealers since they refinance solely via the inter-bank market. Moreover, we include non-bank financial institutions: American Express, a financial services institute;

[9] The extreme value estimation requires only daily high and low. Thus, we do not require high-frequency data.

[10] (A0) follows from (A1) and (A2).

and AIG, which played a crucial role in 2008.

As previously argued, the statistical method only orthogonalizes ϵ_t. However, we also want to bolster the shocks with some economic meaning. More precisely, we want to account for shocks that are common to multiple financial institutions. We include market-specific effects with the SPDR S&P 500 Trust (SPY), which tracks a broad market index. Because the index will diversify shocks to individual entities, the exchange-traded fund's (ETF) volatility will mainly spike if common shocks happen to all stocks. Additionally, we include the Energy Select Sector SPDR (XLE) to control for common shocks to the energy sector. We think that this one is important because our sample includes multiple shocks to the oil price: the GFC, the 2014 oil glut, and the COVID-19 price decline. With these two ETFs, we hope to address the orthogonality of shocks in a more meaningful way. So, if we assume that common shocks exclusively occur to the markets, we can interpret all other shocks as economically meaningful. Nevertheless, there is, most likely, no helpful interpretation of shocks to the market since the identified shocks could be a mix of various macroeconomic factors. Figure 1 shows the average log volatility over time.

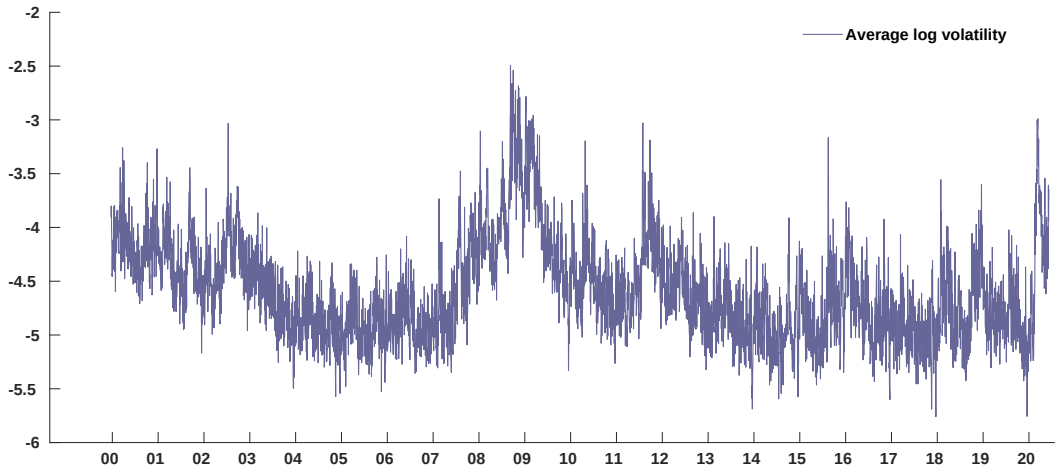

Figure 1: Average volatility over time of the nine financial firms as in Table 1, the Energy Select Sector SPDR (XLE), and the SPDR S&P 500 Trust (SPY). Volatility is measured by the extreme value estimator as in (4.1).

Table 1 shows descriptive statistics of the nine financial institutions. In general, all institutions increased their total assets. Excluding the non-banks AIG and American Express, we observe that all banks had a notable increase in their total assets, ranging from two times to nearly ten times more assets than in 2000.

Institution	Ticker	Market Cap.		Total Assets	
		2000Q1	2019Q4	2000Q1	2019Q4
American Int. Group	AIG	168.8	44.66	279.3	525.06
American Express	AXP	66.2	100.84	150.7	198.32
Bank of America	BAC	86.9	311.21	656.1	2434.08
Bank of NY Mellon	BK	30.6	45.33	76.0	381.51
Citigroup	C	201.8	168.9	738.2	1951.16
Goldman Sachs	GS	44.8	83.19	276.9	992.97
JPMorgan Chase	JPM	71.9	429.91	391.5	2687.38
Morgan Stanley	MS	79.9	81.48	408.1	895.43
Wells Fargo	WFC	66.4	222.43	222.3	1927.56

Table 1: US financial institutions key figures in billions of US dollars.

Our sample period starts on January 3, 2000, and ends June 12, 2020; hence, it includes three major crisis: the 9/11 crash in 2001, the GFC with Lehman Brothers defaulting and the AIG bailout in September 2008, and at the beginning of 2020 the COVID-19 market crash. The data source for all the aforementioned is Reuters. We estimate the model in a two-step procedure (see Appendix C). Specifically, we use kernel-weighted least-square estimation for the coefficients and the extremum estimator from Section 2.2 for the structural matrices. The bandwidths assign 95% of the weight to 150 trading days for the VAR coefficients and the structural parameters.

In the following, we first visualize the whole network for specific dates, and the sample average, and then we show the evolution of connectedness and other key figures related to systemic risk.

4.2 Idiosyncratic risk

Figure 2 shows the entries of B_t over time, and it contains two main findings. First, we see that AIG faced the highest peaks for the idiosyncratic risks. Mainly, we see its curve peaking around the GFC in 2008-09. This result is expected since AIG faced significant problems around the Lehman default in September 2008, followed by AIG's bailout. Later, we will highlight the Lehman default further. Interestingly, we see that the idiosyncratic risks peaked roughly a year later in August 2009. This is because the bailout did not mitigate risks altogether but instead transferred them to AIG. More precisely, potential risks are now more rooted in the firm than received from other firms.

Second, the solid black line shows that the average idiosyncratic risk for the banks

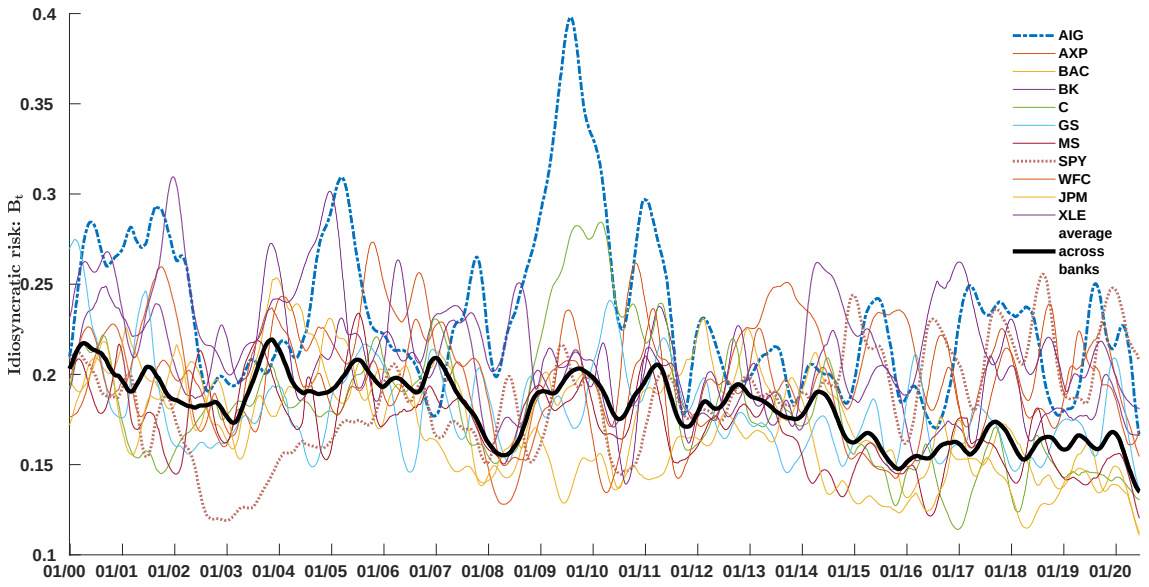

Figure 2: Average idiosyncratic risks over time. The lines show the estimation results for entries in B_t, which correspond to the idiosyncratic risk of a single firm, i.e., the standard deviation of an orthogonal shock. The fat blue dashed line highlights American International Group (AIG), and the fat red dotted line highlights the stock market exchange-traded fund SPY. The bold black solid line shows the average across all banks in the sample. The following stocks and exchange traded funds are included: American International Group (AIG), American Express (AXP), Bank of America (BAC), Bank of NY Mellon (BK), Citigroup (C), Goldman Sachs (GS), JPMorgan Chase (JPM), Morgan Stanley (MS), Wells Fargo (WFC), SPDR S&P 500 Trust (SPY), Energy Select Sector SPDR (XLE).

decreased over the sample. This finding highlights that even though banks massively increased their total assets, they could lower their risks stemming from their business models. A potential explanation is that the Basel accords forced banks to diversify their portfolios further, such that they face fewer risks idiosyncratically. Note, however, that this does not imply that banks face fewer risks overall. Instead, it hints that bank-specific risks edged lower. Dependent risks relate to the concept of connectedness, which we will investigate further.

4.3 Network representation

To show the effect of having an idiosyncratic and connectedness component, we compare the generalized forecast error variance decomposition with its structural counterpart in Figure 3.

In the liquidity crisis of 2008, AIG, in particular, stood in focus when it became public that major banks and trading partners, such as Goldman Sachs, Morgan Stanley, and Bank of America, depended on AIG's liquidity. Credit default swaps and collateralized debt obligations were mainly responsible for the firms' tight entanglement. On September 14, AIG sought $40 billion in Fed aid to survive the uprising liquidity crisis. It followed the default of Lehman Brothers on September 15 and the bailout of AIG on September 16. For this sample, September 15, 2008, marks a key date since it moved investor attention towards AIG. Figure 3 shows the five-day forecast error variance decompositions for this date. We compare the generalized version as in Diebold and Yılmaz (2009) to the structural version introduced in Section 3.1.[11]

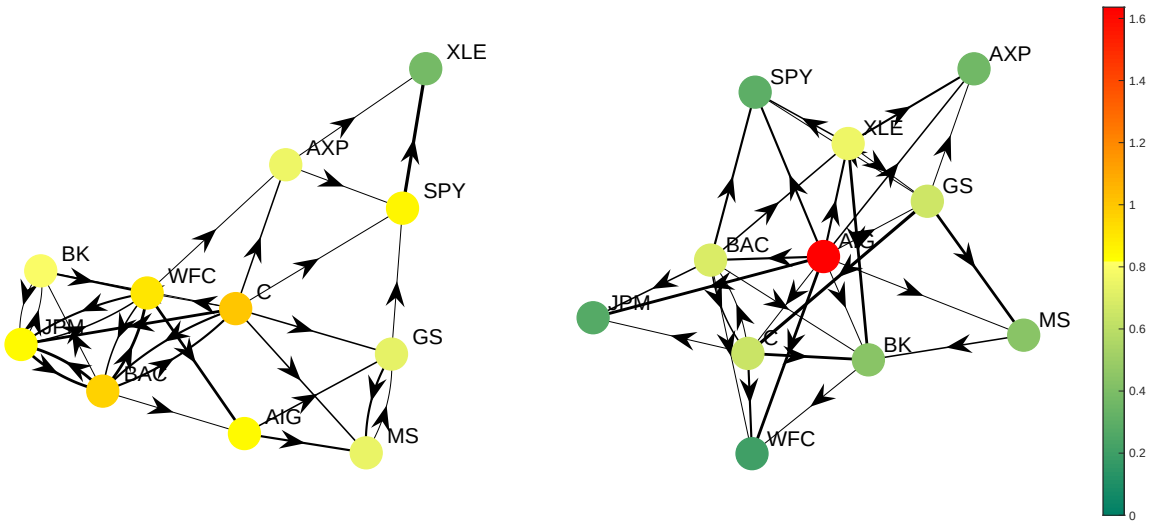

Figure 3: One day after Lehman Brothers defaulted on September 15, 2008: Comparison of forecast error variance decompositions with the generalized version as in Diebold and Yılmaz (2014) on the left and the structural version on the right. Arrows show the 25% strongest connections, with the thickness indicating the quantile of the effect. Node colors indicate the sum of all outgoing arrows, i.e., the outgoing effects of an individual stock on others. The graph is arranged with the force layout, which places nodes closer to each other when they are more connected. The following stocks and exchange-traded funds are included: American International Group (AIG), American Express (AXP), Bank of America (BAC), Bank of NY Mellon (BK), Citigroup (C), Goldman Sachs (GS), JPMorgan Chase (JPM), Morgan Stanley (MS), Wells Fargo (WFC), SPDR S&P 500 Trust (SPY), Energy Select Sector SPDR (XLE).

For the generalized version, we see a higher connectedness of financial institutions

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

in the node colors and mainly see connections for the traditional banks. This side-by-side comparison highlights issues with approximating contemporaneous effects. In crises, investors are more alerted and react faster to shocks. Thus, most significant reactions occur within the same day. Since the generalized FEVD approximates contemporaneous connections with undirected correlations, it cannot uncover the full directed FEVD. We can also see that it connects all receiving institutions as they co-move contemporaneously. In the left graph of Figure 3, firms with high incoming and low outgoing connections are overestimated. AIG has mainly outgoing connections in this period, and thus its effect gets highly underestimated. For the structural FEVD (right graph of Figure 3), in comparison, the graph is nearly star-shaped, i.e., one central node in the middle affects all others. That is, AIG has outgoing connections to all other institutions and the market components. However, this finding does not imply that connections increased necessarily, but investors valued relationships more due to the uncertainty emanating from AIG.

Figure 4 shows the average of the FEVD over time and the end of the sample estimate.[12]

From the left-hand graph, we observe mostly sensible links. First, it is evident that Goldman Sachs and Morgan Stanley, the two investment banks before the GFC, share a mutual link. Their business models are relatively similar, such that they appear to affect each other. These links could stem from an unaccounted for component. Moreover, we see that American Express appears relatively satellite. Its business model is, compared to others, different such that this result is unsurprising. The energy ETF XLE appears rather central and shows strong outgoing effects on the big banks and the market (SPY). In particular, XLE links to JPMorgan Chase, Wells Fargo, and Citigroup, which rank top three as fossil fuel investors from 2016 to 2019.[13] Perhaps surprisingly, AIG shows central for the network as it has a lot of outgoing arrows. This result not only stems from the 2008 bailout but also potentially highlights how important the insurance company is for the US banks.

The right-hand graph shows the end of the sample estimate amidst the COVID-19 pandemic crisis. Here, the effect of the market component SPY on the network is striking: most arrows emerge from SPY (color-coded in red). Recall that shocks to SPY most likely have no economic meaning since the exchange-traded fund bundles many potential

[12]Note here that, for identification reasons, the smoothness assumption forces us to use a two-sided kernel. This makes the accuracy of the estimation better in the middle of the sample than at the end.

[13]`http://priceofoil.org/content/uploads/2020/03/Banking_on_Climate_Change_2020.pdf`

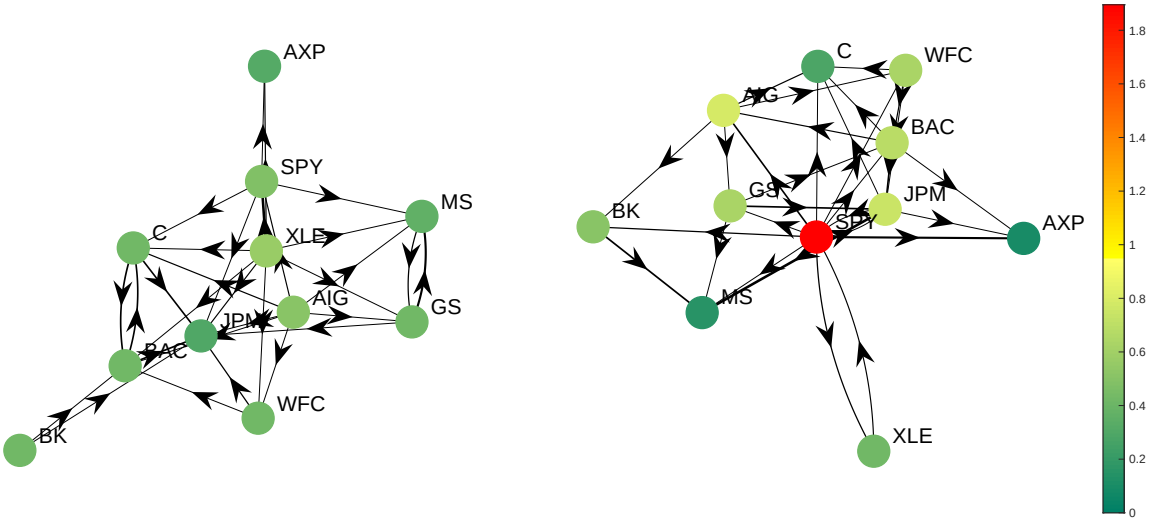

Figure 4: Structural forecast error variance decompositions for the average over time (on the left) and for June 12, 2020 (on the right), with the 25% strongest connections. The thickness of the arrows shows the strength of the effect. Node colors indicate the sum of all outgoing arrows, i.e., the outgoing effects of a respective stock on others. The graph is arranged with the force layout, which places nodes closer to each other when they are more connected. The following stocks and exchange-traded funds are included: American International Group (AIG), American Express (AXP), Bank of America (BAC), Bank of NY Mellon (BK), Citigroup (C), Goldman Sachs (GS), JPMorgan Chase (JPM), Morgan Stanley (MS), Wells Fargo (WFC), SPDR S&P 500 Trust (SPY), Energy Select Sector SPDR (XLE).

aggregate effects. However, the fact that a diversified index has a substantial impact on single stocks is interpretable. Namely, if shocks are not diversified away by the index, shocks are likely to be aggregate, i.e., they simultaneously affect more than one stock. Thus, the strong outgoing effects of SPY tell us that significant parts of increased volatility in the COVID-19 pandemic stem from aggregate components. Later in this section, we will disentangle these effects further.

Table 2 shows the entries of the right-hand network graph. The effect of the market SPY on others is pronounced. Additionally, the incoming effects on the market are relatively low, highlighting further that this crisis is mostly an aggregate shock. JPMorgan is the only bank showing a substantial effect on the market. We will analyze this peculiarity later in this section.

	AIG	AXP	BAC	BK	C	GS	MS	SPY	WFC	JPM	XLE	In
AIG	59.15	0.93	8.90	2.28	1.30	1.60	2.71	9.62	5.09	2.84	5.59	40.85
AXP	7.72	26.42	9.68	4.31	2.05	5.55	0.93	19.92	6.03	13.69	3.68	73.58
BAC	7.71	1.24	26.52	5.75	4.07	10.31	1.55	19.12	13.62	5.93	4.16	73.48
BK	9.79	1.25	6.89	39.41	1.55	5.99	2.44	20.42	2.65	7.05	2.58	60.59
C	8.23	1.10	8.89	5.75	26.38	7.20	2.05	17.49	9.14	8.93	4.84	73.62
GS	9.26	1.69	6.03	5.72	4.22	34.98	1.82	19.78	6.15	7.35	3.01	65.02
MS	7.34	0.74	6.84	10.03	2.89	12.51	18.11	22.68	6.92	8.23	3.71	81.89
SPY	5.01	0.34	3.74	4.83	1.89	2.07	0.97	59.87	2.07	11.28	7.94	40.13
WFC	11.64	0.89	6.67	3.20	2.22	5.05	0.90	16.15	44.27	5.36	3.64	55.73
JPM	7.15	1.15	8.72	6.94	3.73	10.57	0.82	16.46	9.64	31.21	3.61	68.79
XLE	4.87	0.40	2.95	2.11	5.41	1.64	0.89	27.86	1.90	4.21	47.77	52.23
Out	78.73	9.71	69.30	50.91	29.34	62.49	15.08	189.50	63.21	74.87	42.75	62.36

Table 2: Structural forecast error variance decompositions as of June 12, 2020, multiplied by 100. "Out" sums up all column entries without the respective diagonal value. "In" sums up all row entries without the respective diagonal value. The value in the bottom right corner is the average connectedness for that date. The table reads like a matrix, e.g., the effect from American Express (AXP) to American International Group (AIG) is in the first row and second column. The following stocks and exchange-traded funds are included: American International Group (AIG), American Express (AXP), Bank of America (BAC), Bank of NY Mellon (BK), Citigroup (C), Goldman Sachs (GS), JPMorgan Chase (JPM), Morgan Stanley (MS), Wells Fargo (WFC), SPDR S&P 500 Trust (SPY), Energy Select Sector SPDR (XLE).

4.4 The evolution of average connectedness and market spillovers

Analyzing the evolution of dependencies, we are interested in peaks of the general trend. To frame the dynamics, we include a list of events that are potentially important for the whole system. Table 3 gives an overview of these events.

#	Date	Event Description
1	July 31, 2007	Bear Stearns liquidates hedge funds: peak of the subprime crisis
2	September 14-16, 2008	Lehman Brothers default, AIG bailout
3	May 6, 2010	Flash Crash / Crash of 2:45
4	August 6, 2011	Downgrading of the US to AA+ amidst the European sovereign debt crisis
5	November 2014	Oil glut
6	June 2015-February 2016	Chinese stock market turbulence
7	April 20, 2020	Oil futures drop below 0 dollars amidst the COVID-19 crisis

Table 3: Important events for the dynamics of average connectedness

Figure 5 contains four lines: the average connectedness and its decompositions in market components and financial firms. Here, an average connectedness of 60 means 60% variation in stocks' volatilities are explained by shocks to others. Moreover, the value 20 in the dashed green line means 20% of financial institutions' volatilities are explained by

market components, i.e., the SPY or XLE. These spillovers are the summed entries of the cross-section of the SPY and XLE columns and the financial firms' rows (divided by N). The dotted red line shows the spillovers only happening between financial institutions, i.e., the sum of entries from financial firms to financial firms (divided by N). Analogously, the spillovers from financial firms to the market components are plotted in dash-dotted purple. Note here that there are nine financial institutions and only two market variables. Therefore, a quantitative comparison of the spillover lines is meaningless without scaling them.

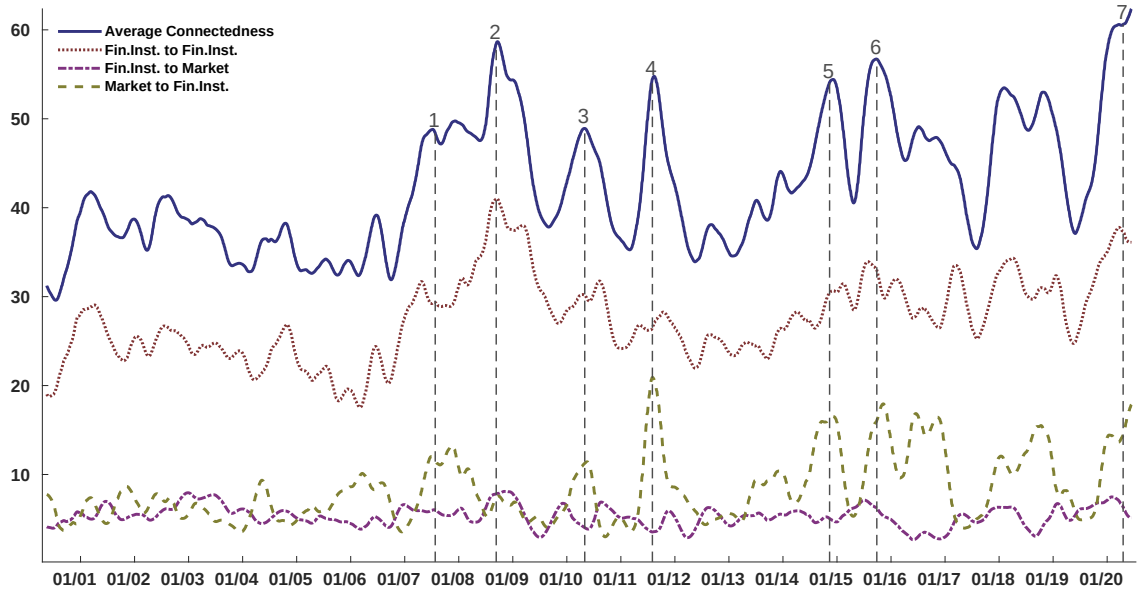

Figure 5: Average spillovers (solid blue), spillovers between the financial institutions (dotted red), spillovers from the market to financial institutions (dashed green), spillovers from financial institutions to the market (dash-dotted purple). The nine financial institutions are as in Table 1, and the two market components are the SPY and the XLE, representing the overall stock market and the energy sector, respectively. Events are described in Table 3.

First, note that the first two events during the GFC show a clear build-up to a peak of average connectedness. That is, Event 1 was mainly a market spillover during the time when the subprime crisis emerged. The spillovers from the market to financial institutions (green dashed line) increased and stayed on this level until mid-2008. At that time, the spillovers between the financial institutions (red dotted line) were already exceptionally high. In mid-2008 things changed, and the crisis rightly arrived in the nine financial

institutions. While the market effects diminished, the financial institutions got into more trouble and spilled more risk. The purple dash-dotted line also increased, showing that the financial sector had a larger impact on the market than vice versa. The intra-financial spillovers showed a clear peak during the Lehman default and AIG bailout. This is the highest measured peak for this measure during the whole sample.

Second, we see other peaks coinciding with major events (Events 3-7). These events appear to come mainly from the market components. That is, we see all these events peaking in the dashed green line. The flip side shows in the dotted red line, where intra-financial spillovers appear relatively unaffected by the major events. Perhaps surprisingly, the highest peak in market-to-financial-institution spillovers is the downgrading of the US to AA+ (Event 4) in the middle of the European sovereign debt crisis. This event caused notable troubles on the stock market and measured exceptionally high volatilities. However, it shows that the intra-financial spillovers were relatively unaffected by this event. In contrast, the COVID-19 and the oil price drop (Event 7) caused high market-to-financial-institution spillovers and led to high intra-financial spillovers. These two combined resulted in the highest average connectedness (blue line) measured in the whole sample. This highlights that the COVID-19 crisis appears to be mainly a commonality effect.

Excursion The May 6, 2010, Flash Crash, also known as the Crash of 2:45 (Event 3), was a stock market crash that lasted for approximately 36 minutes. On this day, the stock market indices lost nearly 10% within minutes, only to rebound half an hour later. Volatility measured exceptionally high and, perhaps surprisingly, connectedness did as well. Since we cannot identify any other events around this peak, the question arises whether the volatility level causes the level of average connectedness. In particular, more uncertainty in the stock market goes hand in hand with a higher average correlation between the stock returns. However, it is unclear whether a higher volatility level causes more connectedness or vice versa. Through the story, we expect that connections in the network form exogenously. That is, a higher degree of connectedness enhances the spillover of idiosyncratic shocks. By investigating the lead-lag correlation (see Appendix D) of average connectedness and average volatility, we find support for exogenous network formation. This finding suggests that average connectedness is a predictor of risk since its variations lead variations in average volatility. This finding opens up a new research question: Is average connectedness a good predictor of volatility?

4.5 On the relevance of single institutions

As a firm-level analysis, we show the Systemic Relevance of the nine financial institutions. This measure, which we introduced in Section 3.3, demonstrates the centrality of institutions. Recall that the measure is equivalent to a placebo experiment and is meant to reveal the importance of the respective financial institution. That is, it measures the difference in the amplification of the impact matrix when institution i's outgoing causal connections are mitigated. A higher value of this measure relates to a higher relevance for the system. The result is a combination of outgoing connections, ingoing connections and the general architecture of the system's network. Table 4 describes events related to this analysis.

	Date	**Event Description**
A	September 16, 2008	AIG bailout
B	November 27, 2011	Writedown of 50% of Greek bonds amidst the Greek sovereign debt crisis
C	December 14, 2012	Government commitments to AIG fully recovered
D	March 1, 2018	Trump announces steel tariffs amidst stock market corrections

Table 4: Important events at the firm level

Figure 6 depicts the Systemic Relevances. A value of 25 for institution i relates to a 25% increase in shock amplification due to i. While most of the pattern appears to be somewhat random, we can still observe some interesting dynamics. First, from eyeballing, the lines appear to be less co-moving and varying before the GFC. Most likely, this stems from increasing regulations and, consequently, more interconnections. Second, we see a sharp drop in AIG's importance after the bailout by the Fed (Event A). AIG's Systemic Relevance stayed low, except for the downgrading of the US to AA+ (small peak in 2011). After Event C, when AIG fully recovered government commitments, its Systemic Relevance went to higher levels. This story is in line with the fact that AIG should not have spread more risk after the Fed bailed them out.

Moreover, we see in Event B how a writedown of 50% on Greek bonds affected the relevance of Goldman Sachs. In particular, GS stood out in the European sovereign debt crisis since they were involved in hiding the beleaguered state's debt. At that point, it was well known that GS even bet on a Greek default. Thus, we see for GS a higher peak than for anybody else at this event. In Event 6, during the Chinese stock market

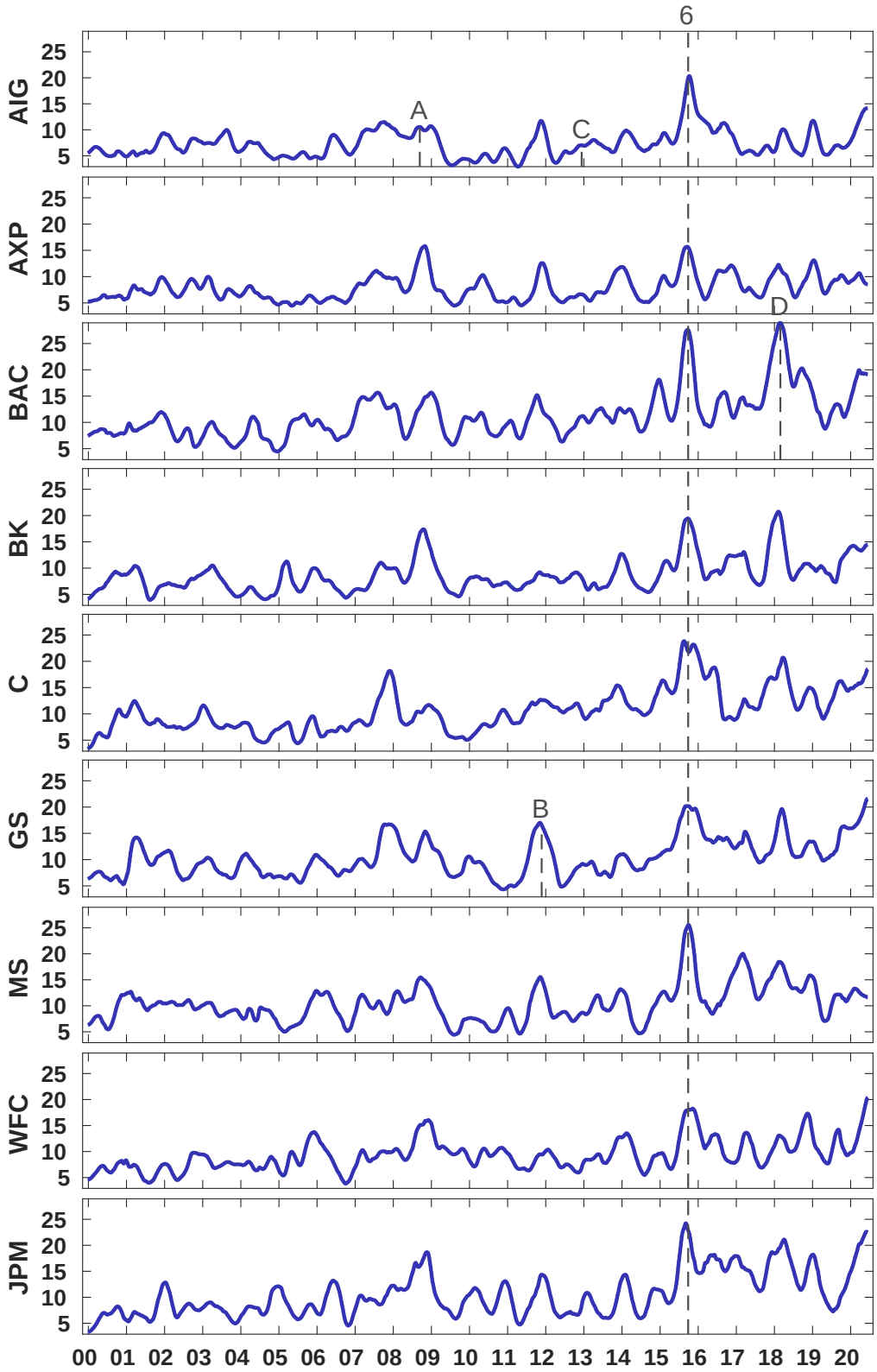

Figure 6: Systemic Relevance measured as described in Section 3.3. Firm-specific events are denoted by letters and are described in Table 4. General events are denoted by numbers as in Table 3.

turbulences, some institutions showed more relevance than others. Bank of America, Citigroup, Goldman Sachs, Morgan Stanley, and JPMorgan Chase show a notable increase in relevance during that time. This observation is in line with the exposure of US banks to China. Namely, these banks also showed the highest exposures to the Chinese market in 2015. To accurately assess common exposures like the Chinese market, it is advisable to include some sort of proxy. However, given the scope of this paper, we leave this question for future research.

Moving forward, Event D shows the highest relevance during the sample: Bank of America's relevance measured just above 25 when President Trump announced steel and aluminum tariffs. These tariffs are known to be disruptive for US firms and consumers (see Amiti et al. (2020)). However, it remains unclear whether this specific peak for Bank of America comes from the stock market corrections observable at that time or from some sort of increased exposure to steel-related industries. Lastly, we see a sharp increase of Bank of America, Goldman Sachs and JPMorgan Chases's relevance at the end of the sample and during the COVID-19 crisis. In particular, JPMorgan Chase is the most relevant institution as of June 2020.

5 Conclusion

In this paper, we tackle the question of how to analyze causal networks. We develop time-varying assumptions that can identify directed contemporaneous connections. More precisely, we model time variation non-parametrically, which allows us to estimate a time trend, ensuring identification. The identification works under parsimonious conditions and comes with a tractable estimation function. Furthermore, we seize the problem of static response matrices by allowing all parameters to be time-varying. By arguing about the causality of the contemporaneous response matrix, we introduce a new centrality measure that is economically meaningful. This centrality measure highlights the relevance of financial firms and serves as a critical tool for policymakers to monitor risk.

In the application, we study measures of connectedness. That is, we are interested in spillovers measured by the framework of Diebold and Yılmaz (2014). We highlight issues with approximating contemporaneous relations by correlations and, further, successfully address these with a structural decomposition. The application includes a contemporane-

ous causal dependency matrix, which provides new insights into financial firms' connectedness. Eventually, we argue how this estimation approach relates to the establishment of causality.

Empirically, we find that banks show decreasing idiosyncratic risks over the last 20 years. From a connectedness perspective, we find that investors mostly react within a day in uncertain periods. This finding stresses the importance of using directed contemporaneous dependencies. We exemplify the directed connectedness of financial firms by comparing generalized and structural forecast error variance decompositions. In so doing, we show the dependency structure for specific dates and the sample average.

Moreover, it appears that average connectedness based on market data mostly peaks with adverse financial events. However, decomposing it shows that most of these events are common effects, and only the GFC and the most recent COVID-19 crisis manifest an increase in financial firm spillovers. This finding highlights not only that investors care more about the connection of banks in crisis periods but also that financial firms' spillovers were rather flat over the sample.

Finally, we investigate the Systemic Relevance of financial firms. In particular, Bank of America measures high over the sample, and JPMorgan Chase stands out in June 2020. At that time, JPMorgan Chase was the biggest bank and also asserted its dominance by being the most relevant institution over the whole sample, suggesting that it warrants further monitoring by the supervisory authorities.

References

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

Tobias Adrian and Markus K Brunnermeier. CoVaR. Technical report, National Bureau of Economic Research, 2011.

Mary Amiti, Stephen J Redding, and David E Weinstein. Who's paying for the US tariffs? A longer-term perspective. Working Paper 26610, National Bureau of Economic Research, January 2020. URL `http://www.nber.org/papers/w26610`.

Joshua Angrist and Jorn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 1st edition, 2009. URL `https://EconPapers.repec.org/RePEc:pup:pbooks:8769`.

Matteo Barigozzi and Christian Brownlees. Nets: Network estimation for time series. *Journal of Applied Econometrics*, 34(3):347–364, 2019.

Monica Billio, Mila Getmansky, Andrew W Lo, and Loriana Pelizzon. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559, 2012.

J Blanchard Olivier and Danny Quah. The dynamic effects of aggregate demand and supply disturbances. *American Economic Review*, 19:655–73, 1989.

Christian T Brownlees and Robert Engle. Volatility, correlation and tails for systemic risk measurement. *Available at SSRN*, 1611229, 2012.

Rainer Dahlhaus, Wolfgang Polonik, et al. Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes. *The Annals of Statistics*, 34(6): 2790–2824, 2006.

Roberto A De Santis and Srecko Zimic. Spillovers among sovereign debt markets: Identification by absolute magnitude restrictions. 2017.

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

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

Jianqing Fan, Nancy E Heckman, and Matt P Wand. Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of the American Statistical Association*, 90(429):141–150, 1995.

Renee Fry and Adrian Pagan. Sign restrictions in structural vector autoregressions: A critical review. *Journal of Economic Literature*, 49(4):938–960, 2011.

Liudas Giraitis, George Kapetanios, Anne Wetherilt, and Filip Žikeš. Estimating the dynamics and persistence of financial networks, with an application to the sterling money market. *Journal of Applied Econometrics*, 31(1):58–84, 2016.

Yunmi Kim and Chang-Jin Kim. Dealing with endogeneity in a time-varying parameter model: Joint estimation and two-step estimation procedures. *The Econometrics Journal*, 14(3):487–497, 2011.

Markku Lanne, Helmut Lütkepohl, and Katarzyna Maciejowska. Structural vector autoregressions with Markov switching. *Journal of Economic Dynamics and Control*, 34 (2):121–131, 2010.

Daniel Lewis. Identifying shocks via time-varying volatility. *Manuscript, Harvard University*, 2017.

Daniel Lewis. Robust inference in models identified via heteroskedasticity. *FRB of New York Staff Report*, (876), 2018.

Martin Martens and Dick Van Dijk. Measuring volatility with the realized range. *Journal of Econometrics*, 138(1):181–207, 2007.

George Milunovich and Minxian Yang. On identifying structural VAR models via ARCH effects. *Journal of Time Series Econometrics*, 5(2):117–131, 2013.

Michael Parkinson. The extreme value method for estimating the variance of the rate of return. *The Journal of Business*, 53(1):61–65, 1980.

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

Roberto Rigobon and Brian Sack. Spillovers across US financial markets. Technical report, National Bureau of Economic Research, 2003.

Juan F Rubio-Ramirez, Daniel F Waggoner, and Tao Zha. Structural vector autoregressions: Theory of identification and algorithms for inference. *The Review of Economic Studies*, 77(2):665–696, 2010.

Christopher A Sims. Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, pages 1–48, 1980.

Harald Uhlig. What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2):381–419, 2005.

Sewall Wright. Correlation and causation. *Journal of Agricultural Research*, 20:557–580, 1921.

Appendices

A Structural Identification

A.1 Observationally equivalent matrices

To get further insights into the identification and estimation, we take the respective reduced form of (1),

$$u_t = A_{0,t}^{-1} B_t \epsilon_t, \quad E[u_t u_t'] = \Sigma_t = A_{0,t}^{-1} B_t B_t' \, A_{0,t}^{-1\prime}. \tag{10}$$

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

In order to understand the challenges of structural identification, we follow Rubio-Ramirez et al. (2010) and introduce the concept of observational equivalence. In short, two structural parameter points are observationally equivalent if and only if they yield the same reduced-form distribution of u_t. In our setting, this holds true if they yield the same Σ_t. To visualize this phenomenon, we redefine the structural parameters such that

$$\begin{aligned} \Sigma_t &= S_t S_t', \\ \text{with} \qquad S_t &= A_{0,t}^{-1} B_t. \end{aligned} \tag{11}$$

Consequently, the parameter set $\{\widetilde{S}_t = S_t Q | QQ' = I_N\}$ satisfies (11) and thus produces the same distribution. In the context of SVARs, we end up with the same observations. The equivalence follows directly from the orthogonality of Q and can be observed by plugging it into (11).

Note that whenever there exists more than one observationally equivalent structural parameter, the structural model is not identified. Thus, the set of orthogonal matrices defines all possible solutions in an estimation. However, the process in (1) suggests that only S_t is correct. Also, this S_t is the only matrix that identifies the real shocks ϵ_t. The fact

that we can only observe and directly estimate (11) limits identification and estimation of the real structural parameters in (1) to the set of observationally equivalent matrices. To obtain local identification of the structural parameters (A_0, B_t), we need to rule out all Qs but $Q = I_N$ such that only the true $\widetilde{S}_t = S_t I_N$ remains.

Recall that conditions on the parameter space, such as exclusion and long-run restrictions, are sufficient for local identification in a structural VAR context. However, economically motivating these restrictions is usually a fruitless challenge. Therefore, we desire weaker conditions to fit more applications. Sign restrictions, for example, are easy to motivate but only reduce the number of observationally equivalent parameters to more than one $\widetilde{S}_t$. In this case, we say the parameters are partially identified or set-identified.

Although sign restrictions are frequently applied, they fall short when it comes to the interpretation of the estimation results. Fry and Pagan (2011) point out that solutions to the estimation cannot be interpreted with probabilistic language anymore, due to the fact that all but one $\widetilde{S}_t$ are untrue.[14] Moreover, in the context of time variation, set-identification prohibits depicting point dynamics and only allows for interval dynamics, which are, as pointed out, non-probabilistic. To address this problem, this study focuses on point-identification by exploiting the mild assumptions (A1) and (A2) for time-varying parameters.

A.2 Propositions and proofs

Proposition 1. *Let $\Sigma^{-1} = SS'$ and $\dot{\Sigma}^{-1} = S\Lambda S'$ be the inverse of a $(N \times N)$ covariance matrix and its differential, where S is invertible and $\Lambda = diag(\lambda_1, .., \lambda_N)$ diagonal. If $\lambda_i \neq \lambda_j$ for all $i \neq j \in 1, ..., N$ the $(N \times N)$ matrix S in the decomposition from mappings $\Sigma^{-1} = SS'$ and $\dot{\Sigma}^{-1} = S\Lambda S'$ is unique apart from permutation and sign reversal of the columns, and Λ is unique up to the same permutation and sign reversal of its diagonal entries.*

Proof: Take any other solution of the decomposition $\tilde{S} = SQ$ and a diagonal matrix

[14] In a first draft of this paper, time variation was used to evaluate the goodness of fit of a single $\widetilde{S}_t$ in the set of partially identified parameters, but obtaining the complete set of observationally equivalent solutions proved to be infeasible.

$\tilde{\Lambda}$ with distinct values leading to same observations, such that

$$\begin{aligned} \Sigma^{-1} &= SS' = SQQ'S,' \quad (12) \\ \Rightarrow \dot{\Sigma}^{-1} &= S\Lambda S' = SQ\tilde{\Lambda}Q'S'. \quad (13) \end{aligned}$$

Multiplying the inverse of S from left and right in (12) gives $QQ' = I_N$. Hence, Q has to be orthogonal. To prove the proposition, we need to rule out all orthogonal matrices except the Qs which are permutation matrices with columns multiplied by 1s or -1s. From (13) it follows that $Q\tilde{\Lambda}Q' = \Lambda$ and therefore $Q\tilde{\Lambda} = \Lambda Q$, i.e. $\tilde{\lambda}_j q_{ij} = \lambda_i q_{ij}\ \forall i, j$. So, either $q_{ij} = 0$ or $\tilde{\lambda}_j = \lambda_i$. Trivially, it also holds that $\tilde{\lambda}_k q_{ik} = \lambda_i q_{ik}\ \forall k \neq j$. That is, every other value in row i also has q_{ik} is 0 or $\tilde{\lambda}_k = \lambda_i$. Now, when $\tilde{\lambda}_j = \lambda_i$, it has to hold that $\tilde{\lambda}_k \neq \lambda_i\ \forall k \neq j$, since all values in Λ and $\tilde{\Lambda}$ are distinct. Hence, there exists exactly one entry per column in Q. Combined with $QQ' = I_N$, it follows that $q_{ij}^2 \in \{0, 1\}$. Thus, Q has to be a permutation matrix with entries multiplied by 1s and -1s. Moreover, $\tilde{\Lambda} = Q'\Lambda Q$. □

Proposition 2. *Assume $S = A_0'B^{-1}$ and $\Lambda = -2\dot{B}B^{-1}$ with A_0 having unit diagonal, $B = diag(b_1, .., b_N)$, $\dot{B} = diag(\dot{b}_1, ..., \dot{b}_N)$ and S and Λ are as Proposition 1. Then the decomposition $\Sigma^{-1} = A_0'B^{-1}B^{-1}A_0$ and $\dot{\Sigma}^{-1} = A_0' - 2\dot{B}B^{-3}A_0$ leads to unique A_0 and unique up to sign B, $\dot{B}$, if the off-diagonal elements of A_0 are restricted to be between -1 and 1.*

Proof: Assume $\tilde{A}_0$, $\tilde{B}$ and $\tilde{\dot{B}}$ are solutions to the decomposition. That is,

$$\begin{aligned} \tilde{S} &= \tilde{A}_0'\tilde{B}^{-1}, \\ \tilde{\Lambda} &= -2\tilde{\dot{B}}\tilde{B}^{-1}. \end{aligned}$$

Further, since all off-diagonal values in $\tilde{A}_0$ have to be between -1 and 1, $|\tilde{s}_{ij}| < \tilde{s}_{jj}\ \forall j \neq i$. More precisely, since the diagonal entry in $\tilde{A}_0$ is the biggest absolute value in its row, it also has to hold for columns of $\tilde{S}$. From Proposition 1 we know that $\tilde{S} = SQ$, where Q is either diagonal or a permutation matrix with entries multiplied by 1s and -1s. Now, assume Q is non-diagonal such that the columns in S are permuted. Then, the permutation would contradict $|\tilde{s}_{ij}| < \tilde{s}_{jj}\ \forall j \neq i$. Thus, Q is diagonal with entries ± 1. □

B Structural Forecast Error Variance Decompositions

Without loss of generality, assume a structural VAR(1),

$$A_{0,t}y_t = A_{1,t}y_{t-1} + B_t\epsilon_t,$$

where this can be the companion form of a higher order VAR. The MA(∞) representation for forecast errors u_t reads

$$y_t = \sum_{k=0}^{\infty} \Phi_{k,t}u_{t-k}, \quad \Phi_{0,t} = I_N, \quad \forall t \tag{14}$$

where $\Phi_{k,t}$ is built from the autocorrelation parameters. Since we assume a structural vector auto regression, we already possess knowledge about a decomposition of u_t. That is, we use $u_t = A_{0,t}^{-1}B_t\epsilon_t$ to obtain

$$y_t = \sum_{k=0}^{\infty} \Theta_{k,t}\epsilon_{t-k}, \quad \Theta_{k,t} = \Phi_{k,t}A_{0,t-k}^{-1}B_{t-k}. \tag{15}$$

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

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

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

$$\Psi_t^2(H) = \sum_{k=0}^{H-1} ((\Phi_{k,t}A_{0,t-k}^{-1}B_{t-k})^{\cdot 2}), \tag{16}$$

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

$$d_{ij,t}^H = \frac{\psi_{ij,t}^2(H)}{\sum_g \psi_{ig,t}^2(H)}, \tag{17}$$

where $\psi_{ij,t}^2(H)$ denotes the ij-th entry of $\Psi_t^2(H)$. The ratio explains j-th percentage contribution on the total forecast variance of all variables on i. Note that, in contrast to

the generalized version of Pesaran and Shin (1998), the rows of D_t^H sum to one.

C Estimation Strategy

The estimation includes two steps. First, we estimate the reduced form, and, second, we take the resulting residual series for the structural estimation.[15] For the empirical application, the companion form

$$Y_t = A_t^* X_t + U_t \tag{18}$$

demonstrates usefulness, where

$$\begin{aligned} Y_t &= (y_t', \ldots, y_{t-p+1}')', \quad X_t = (1, y_{t-1}', \ldots, y_{t-p}')', \quad U_t = (u_t, 0, \ldots, 0)', \\ A_t^* &= \begin{bmatrix} A_{0,t}^{-1}\alpha & A_{0,t}^{-1}A_{1,t} & A_{0,t}^{-1}A_{2,t} & \cdots & A_{0,t}^{-1}A_{p,t} \\ 0 & I_N & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & I_N & 0 \end{bmatrix}. \end{aligned}$$

In the first step, we perform a time-variation estimation of (18). We use kernel-weighted least squares estimation, i.e., a special form of the generalized least squares,

$$\widehat{A}_t^* = Y W_t X' (X W_t X')^{-1}, \tag{19}$$

where W_t is defined as in (7) and $Y = [Y_1, ..., Y_T]$, $X = [X_1, ..., X_T]$. The corresponding reduced-form residuals are the first N values in $\widehat{U}_t = Y_t - \widehat{A}_t^* X_t$. We use this series in (7) to obtain the structural parameters. Similar to the rolling window estimation, this estimation approach has the advantage of great simplicity and allows us to stay in a prior-free environment, i.e., it unveils a more meaningful evolution of connectedness. Note that we explicitly need assumptions on the time variation of $A_{i,t}$. Because they are increments contributing to the long-run dependency, we impose the same assumptions on them as on $A_{0,t}$.

A crucial step in the estimation concerns the choice of bandwidths for both steps: h_1

[15] Note here that a joint estimation as in Kim and Kim (2011) could improve on the potential inference problematic of the two-step procedure.

and h_2. While arguably the results of the first step affect the second step severely, we explored robustness of the second step with respect to the first bandwidth. Due to this robustness and the fact that misspecification potentially increases estimation uncertainty in the second step, we apply a fully data-driven cross validation (CV) for this bandwidth. In order to minimize forecast errors in the first step, it is advisable to use the likelihood of the leave-one-out interpolation errors,

$$\widehat{u}^{\circ}_{t,h_1} = y_t - \widehat{y}^{\circ}_{t,h_1}, \tag{20}$$

where $\widehat{y}^{\circ}_{t,h_1}$ denotes the estimate of y_t with bandwidth h_1 and weight 0 on observation t. This step is equal to setting the t-th diagonal entry of W in (19) to 0. The bandwidth we consider is chosen with the following log-likelihood type leave-one-out criterion:[16]

$$h_1^{CV} = \arg\max_h -\frac{1}{2}\sum_t \left(\widehat{u}^{\circ}_{t,h}{}' \widehat{\Sigma}^{-1}_{\widehat{u}^{\circ}_{t,h}} \widehat{u}^{\circ}_{t,h} + log(|\widehat{\Sigma}_{\widehat{u}^{\circ}_{t,h}}|) \right), \tag{21}$$

where $\widehat{\Sigma}_{\widehat{u}^{\circ}_{t,h}}$ is the full-sample covariance of the interpolation errors. Then h_1^{CV} is the bandwidth that maximizes the explanatory power of the estimation given a training set. This set is simply the bandwidth weighted sample without the value it predicts. The maximization, however, resulted in a unreasonably high bandwidth, such that we are not able to observe any dynamics in this context. Thus, we chose a bandwidth $h_1 = \frac{150}{2q_{0.95}T}$, which corresponds to 150 trading days in our data. T is the sample length, and $q_{0.05}$ denotes the 95% quantile of the respective kernel distribution. We use a Gaussian kernel due to its good asymptotic behavior. The bandwidth assigns 95% of the kernel's weight to 150 trading days.

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

[16]The sum of squared interpolation errors is potentially inaccurate for a selection criterion because we expect the covariance matrix to be very different than the identity.

Thus, in this paper, we focus again on a bandwidth that is meaningful for the needs of our story. Namely, we choose a bandwidth that assigns 95% of the kernel's weight to 150 trading days. In particular, $h_2 = \frac{150}{2q_{0.95}T}$, where T is the sample length and $q_{0.05}$ denotes the 95% quantile of the respective kernel distribution. We use a Gaussian kernel due to its good asymptotic behavior.

D Complementary Figures

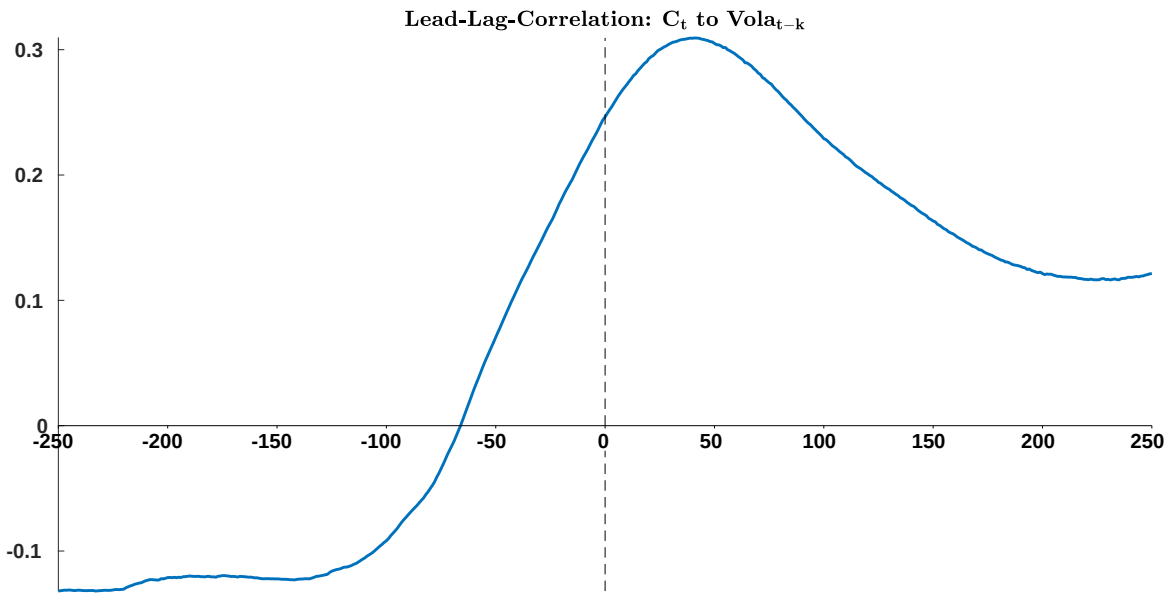

Figure 7: Lead-lag correlation: $\text{corr}(C(D_t^5), \overline{vola}_{t-k})$. The average volatility at time t is taken over all variables. The x-axis shows the difference in days.