Bayesian state-space modeling for analyzing heterogeneous network effects of US monetary policy

NIKO HAUZENBERGER and MICHAEL PFARRHOFER*

University of Salzburg

Understanding disaggregate channels in the transmission of monetary policy is of crucial importance for effectively implementing policy measures. We extend the empirical econometric literature on the role of production networks in the propagation of shocks along two dimensions. First, we set forth a Bayesian network panel state-space model that assumes time variation in the network dependence parameter. The framework is applied to a study of the network effects of Federal Reserve (Fed) monetary policy announcements on US stock market returns on the industry level using high-frequency data. Second, we account for cross-sectional heterogeneity and cluster impacts of monetary policy shocks to industries via a sparse finite Gaussian mixture model. The results suggest substantial heterogeneities in the responses of industries to surprise monetary policy shocks in line with the recent literature. Moreover, we find that the impact of policy surprises on stock returns differes strongly over time. This time-variation stems mainly from changes in the strength network effects, with magnitudes of the effects during periods coinciding with recessions being almost twice as large than in expansionary phases.

JEL: C23, C32, O47, R11
KEYWORDS: production networks, monetary policy shocks, high-frequency identification, spatio-temporal modeling

*Corresponding author: Michael Pfarrhofer. Salzburg Centre of European Union Studies, University of Salzburg. *Address*: Mönchsberg 2a, 5020 Salzburg, Austria. *Email*: michael.pfarrhofer@sbg.ac.at. *Phone*: +43 662 8044 3772. We thank Marco Del Negro, Manfred M. Fischer, Sylvia Frühwirth-Schnatter, Florian Huber and Thomas Reutterer for valuable comments and suggestions. We are particularly grateful to Michael Weber for providing the dataset. The authors gratefully acknowledge financial support from the Austrian Science Fund (FWF, grant no. ZK 35) and the Oesterreichische Nationalbank (OeNB, Anniversary Fund, project no. 18127).

1. INTRODUCTION

A growing number of papers explores how shocks on the micro- and macro-level propagate through economic networks and how such shocks relate to aggregate fluctuations. Most articles provide substantial evidence for the importance of network effects (see, for instance, Gabaix, 2011; Acemoglu *et al.*, 2012; Elliott *et al.*, 2014; Acemoglu *et al.*, 2015; Ozdagli and Weber, 2017). Recent empirical analyses, however, suffer from a set of limiting shortcomings: they mainly rely on constant parameter specifications and either focus on aggregate data or neglect heterogeneities among cross-sectional observations.

In this article, we address these issues and extend the literature on spatial panel data models (see, for instance, Elhorst, 2014; Aquaro *et al.*, 2015; LeSage and Chih, 2016), linking them to the vast literature on Bayesian state-space modeling (see Kim and Nelson, 1999). As a topical application, we focus on the transmission of monetary policy shocks through the US production network. Our approach is closely related to Ozdagli and Weber (2017), who generalize the setup proposed in Bernanke and Kuttner (2005) and Gürkaynak *et al.* (2005) for analyzing the impact of changes in monetary policy on equity prices.[1] While Bernanke and Kuttner (2005) and Gürkaynak *et al.* (2005) focus mostly on aggregate data like the S&P500 index, Ozdagli and Weber (2017) find substantial evidence for higher-order effects of monetary policy on stock market returns using disaggregate data on the industry-level, and attribute between 60 to 80 percent of the total effects to spillovers between industries.

In the empirical application, Ozdagli and Weber (2017) disregard industry-specific idiosyncrasies and time-variation in the strength of network dependencies. This is problematic for two reasons. *First*, structural breaks in macroeconomic and financial series are increasingly drawing interest in the monetary policy literature (see, for instance, Cogley and Sargent, 2005; Primiceri, 2005; Sims and Zha, 2006). More directly related to our empirical application, there is substantial empirical evidence for differences in the impact of monetary policy on stock returns over the business cycle (see Chen, 2007; Kurov, 2010; Ciccarelli *et al.*, 2013; Kontonikas *et al.*, 2013). Basistha and Kurov (2008), for instance, find that stock returns respond much stronger to surprise monetary policy shocks during recessions, under tight credit market conditions.

Second, pooling information across industries may conceal important underlying structural relationships, and potentially distorts the estimated importance of some industries in the disaggregate transmission of monetary policy shocks compared to others (see, for instance, Ehrmann and Fratzscher, 2004; Bernanke and Kuttner, 2005; Gorodnichenko and Weber, 2016). This is particularly important considering that industries and sectors differ substantially in size and use vastly different production inputs. In a theoretical framework, Pasten *et al.* (2019) show that heterogeneities in price rigidities are an important

[1] These articles are among a larger body of diverse literature focusing on measuring monetary nonneturality using high-frequency market surprises around central bank policy announcements (see Kuttner, 2001; Cochrane and Piazzesi, 2002; Gürkaynak *et al.*, 2005; Gertler and Karadi, 2015; Nakamura and Steinsson, 2018; Altavilla *et al.*, 2019; Jarociński and Karadi, 2019).

determinant how policy interventions are transmitted to the real economy, suggesting a pronounced role for industry-specific responses of stock market returns to surprise monetary policy shocks.

To address these theoretical and empirical considerations in our proposed econometric framework, we develop a Bayesian state-space model for analyzing network effects of US monetary policy, allowing for heterogeneity both over time and the cross-section. Our contributions are thus both of methodological and empirical nature. *First*, we assume the network dependence parameter to vary over time via imposing a random walk state-equation, and provide Bayesian prior and posterior distributions alongside a sampling algorithm for inference. This modeling feature is related to time-varying network structures (see, inter alia, Asgharian *et al.*, 2013; Billio *et al.*, 2016a), assuming linkage matrices to evolve over time, but keeping the overall strength of network effects constant. By contrast, our approach assumes a time-varying network dependence parameter, effectively capturing the notion that the production network is fixed, but the parameter measuring network effects changes over time. This feature is sensible for our empirical application, given that Ozdagli and Weber (2017) find the production network structure for the US is highly persistent.

Second, we address how to efficiently exploit cross-sectional information for obtaining precise inference, but allow for heterogeneous effects across units in a stochastic fashion.[2] Most of the established spatial econometric methods for panel data analysis rely on deterministic data transformations such as fixed effects. Here, we take a fully Bayesian stance by imposing a hierarchical shrinkage prior on the regression coefficients and the residual variances. In particular, we base our prior setup on sparse finite Gaussian mixtures (see Malsiner-Walli *et al.*, 2016), providing a link to the literature on random coefficient and heterogeneity models (Verbeke and Lesaffre, 1996; Allenby *et al.*, 1998; Frühwirth-Schnatter *et al.*, 2004). This allows us to account for differences in the responses of industry-level stock market returns to monetary policy surprises, while the prior structure centers the model on a parsimonious specification.

From an empirical perspective, two main findings are worth noting. *First*, we obtain substantial evidence for time-variation in the impact of monetary policy surprises on stock market returns and the strength of network dependency. In particular, US recessions tend to coincide with periods where between 40 to 60 percent of the overall effects can be attributed to network effects; expansionary economic episodes show muted network effects with magnitudes of around 20 to 30 percent. This links our findings to Kastner (2019), who identifies increases in co-movement across industries during periods of economic turmoil, and are also in line with Basistha and Kurov (2008) and Kurov (2010), who find pronounced differences in stock market responses during different economic phases. Translated to overall estimated effects of the impact of monetary policy surprises on stock

[2] For a recent paper addressing heterogeneity over the cross-sectional dimension in the spatial econometric context, see Cornwall and Parent (2017). For a textbook introduction on mixture models for panel data, see Frühwirth-Schnatter (2006).

returns, a one percentage point exogenous shock in Federal funds futures yields responses of stock returns of approximately −1.9 percentage points on average across industries, while during recessions, this effect is almost twice as large.

Second, there is substantial evidence pointing towards the necessity of addressing industry-specific reactions to monetary surprises, confirming the empirical and theoretical findings by Ehrmann and Fratzscher (2004), Chen (2007) or Pasten *et al.* (2019). Estimated effects of policy surprises on the stock market are less pronounced when disregarding the notion that some industries are more sensitive to Fed policy changes than others, up to a magnitude of roughly one percentage point on average. Moreover, averaging over time, we find that the stock market response across industries to a one percentage point surprise increase in Federal funds futures translates to a median response across all industries of approximately −2.3 percentage points, with industry-specific estimates up to negative five percentage points.

The rest of the paper is structured as follows. In Section 2, we set forth the model. Section 3 discusses the Bayesian prior setup, while Section 4 provides a sampling algorithm for inference. We apply the model in a study of the network effects of US monetary policy and discuss our findings in Section 5. Section 6 concludes.

2. A TIME-VARYING NETWORK DEPENDENCE PANEL MODEL

The baseline network model is in the spirit of spatial panel specifications and can be written for observation $i = 1, \dots, N$ as

$$y_{it} = \rho_t \sum_{j=1}^{N} w_{ij} y_{jt} + \alpha_i + \boldsymbol{x}_{it}' \boldsymbol{\beta}_i + \epsilon_{it}, \quad \epsilon_{it} \sim \mathcal{N}(0, \sigma_i^2), \tag{1}$$

where y_{it} is the response variable at time $t = 1, \dots, T$.[3] We include an intercept term α_i, K exogenous covariates in the $K \times 1$-vector $\boldsymbol{x}_{it}$ with associated observation specific parameter vector $\boldsymbol{\beta}_i$ of size $K \times 1$ and a Gaussian error term with zero mean and variance σ_i^2.

Information on the cross-sectional dependency structure is incorporated using weighted averages of the "foreign" quantities y_{jt} ($j = 1, \dots, N$) with exogenous weights w_{ij} denoting the elements of an $N \times N$ weighting matrix $\boldsymbol{W}$ subject to the typical restrictions $w_{ii} = 0$, $w_{ij} \geq 0$ and $\sum_{j=1}^{N} w_{ij} = 1$ that guarantee the stability of the model. The first novelty proposed in this paper is that the scalar parameter ρ_t features time-variation. The state equation for the network dependence parameter ρ_t is a random walk process:

$$\rho_t = \rho_{t-1} + \varsigma \xi_t, \quad \xi_t \sim \mathcal{N}(0, 1). \tag{2}$$

[3] The proposed framework for the static version of the panel model adopted for the empirical application in this paper can easily be extended to more flexible specifications, including dynamic models, and setups allowing for stochastic volatilities.

Established econometric methods typically rely on constant network dependency structures. Note that $\varsigma = 0$ implies that the model collapses to a constant parameter model, and testing whether the data suggests time-varying network dependence is thus a variance selection problem, as discussed in the context of a standard time-varying parameter model by Frühwirth-Schnatter and Wagner (2010). We exploit this fact below by imposing a suitable shrinkage prior on these variances that pushes the model towards the constant parameter specification if suggested by likelihood information.

Some of the recent literature highlights the importance of time-varying network structures (see Asgharian *et al.*, 2013; Billio *et al.*, 2016a). These studies imply the cross-sectional weights w_{ij} to evolve over time, but constrain the network dependence parameter to be a constant. Cross-sectional weights are commonly either based on observables or simple correlations, describing the network structure in a sensible way.[4] Although network multipliers (see next subsection) can also be estimated in multivariate systems by decomposing the covariance matrix of the reduced form errors, identification of specific network connections in this setting is less straight-forward (Bianchi *et al.*, 2015; Billio *et al.*, 2016b). In contrast to this approach, a time-varying ρ_t effectively captures the notion that cross-sectional weights are constant, but the overall connectedness is subject to changes. Considering the implied covariance matrix of the reduced form errors,[5] the parameter ρ_t can also be interpreted as capturing a special form of common stochastic volatility.

2.1. *Interpreting the model coefficients*

The approach to modeling network dependence pursued in this paper establishes a large system of simultaneous equations with specific parametric restrictions. Consequently, standard interpretations for linear regressions have to be adapted to account for the notion of cross-sectional dependencies. We follow LeSage and Chih (2016) and derive the impact matrix for the kth coefficient for $k = 1, \ldots, K$ with respect to a change in the kth exogenous covariate $\boldsymbol{x}_{kt} = (x_{1kt}, \ldots, x_{Nkt})'$ for all cross-sectional units as

$$\frac{\partial \boldsymbol{y}_t}{\partial \boldsymbol{x}_{kt}} = \boldsymbol{S}_{kt} = (\boldsymbol{I}_N - \rho_t \boldsymbol{W})^{-1} \boldsymbol{B}_k.$$

Here, $\boldsymbol{B}_k = \text{diag}(\beta_{1k}, \ldots, \beta_{Nk})$ with β_{ik} referring to the kth coefficient of observation i, and the term $(\boldsymbol{I}_N - \rho_t \boldsymbol{W})^{-1}$ is a network multiplier matrix governing the propagation

[4] The production network input-output matrices employed in Ozdagli and Weber (2017)–also used in this paper–are published every five years and available for 1992, 1997 and 2002. Using a time-varying production network structure does not affect their obtained results significantly. Consequently, it is feasible to focus on a fixed dependency structure a priori, and to address time-variation via the dependence parameter.

[5] For the stacked version of the model, the covariance matrix of the reduced form errors at time t is given by the expression $(\boldsymbol{I}_N - \rho_t \boldsymbol{W})^{-1} \boldsymbol{\Sigma} (\boldsymbol{I}_N - \rho_t \boldsymbol{W})^{-1\prime}$ with $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_N^2)$.

of the shocks through the network structure.[6] Following LeSage and Pace (2009), it is conventional to define $1/N \times \text{tr}(\boldsymbol{S}_{kt})$ as the *average direct* effect, $1/N \times \boldsymbol{\iota}'_N \boldsymbol{S}_{kt} \boldsymbol{\iota}_N$ as the *average total* effect, and the difference between the two as the *average indirect*, or network effect. Assuming time-varying network dependence yields an impact matrix $\boldsymbol{S}_{kt}$ for $t = 1, \ldots, T$.

It is worth mentioning that the Bayesian approach we set forth allows for adequate quantification of uncertainty surrounding all model parameters and functions thereof. Besides full posterior distributions of the total, direct and indirect effects, we obtain confidence bounds for the overall strength of the network effects over time.

3. PRIOR SPECIFICATION

We estimate the proposed model using Bayesian methods. This involves selecting suitable prior distributions for all parameters and combining them with the likelihood of the data. We first discuss the prior setup for the time-varying network dependence parameter. Conditional on a draw of the full history of this parameter $\{\rho_t\}_{t=1}^T$, inference for the other model parameters is mostly standard, and we subsequently discuss the prior setup for the regressions coefficients and the error variances.

3.1. *A spike-and-slab prior testing for time variation*

For the network dependence parameter, we propose a prior setup that stochastically determines whether time-variation is required for adequately capturing observed dynamics, nesting conventional constant parameter specifications as a special case of our framework. We adapt a variant of the well known stochastic search variable selection prior of George and McCulloch (1993) in the context of the time-varying parameter model (for a related approach, see Frühwirth-Schnatter and Wagner, 2010).

We impose a mixture of two gamma priors allowing either for substantial mass close to zero, suppressing time variation, or loose enough to allow for time-varying network dependence. In particular, we specify a mixture of two Gaussians on the signed square root of the innovation variance in Eq. (2),

$$\pm\varsigma \sim \delta \times \mathcal{N}(0, B_1) + (1 - \delta) \times \mathcal{N}(0, B_0),$$

with $B_1 \gg B_0$ and B_0 close to zero, which is equivalent to stating

$$\varsigma^2 \sim \delta \times \mathcal{G}\left(\frac{1}{2}, \frac{1}{2B_1}\right) + (1 - \delta) \times \mathcal{G}\left(\frac{1}{2}, \frac{1}{2B_0}\right).$$

The latent binary indicator δ dictates which one of the two components is active. Given $\delta = 1$, the prior on ς^2 is rather loose based on larger values of B_1 and reflects time

[6] Note that disregarding the restriction of $\text{diag}(\boldsymbol{W}) = 0$, as in Ozdagli and Weber (2017), results in upward bias of the estimated network effects due to the lack of convergence of the Neumann series expansion of the network multiplier $(\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1} = \sum_{i=1}^{\infty} \rho^i \boldsymbol{W}^i$.

variation in the network dependence parameter by allowing for non-zero variances of the error term in the state equation. For $\delta = 0$, the second component with variance B_0 close to zero is active, pushing the signed square root of the innovation variance towards zero, effectively ruling out time variation. As a byproduct, this specification yields a posterior probability measure whether time variation for these coefficients is necessary to adequately reflect the data generating process. In the empirical application, we set $B_1 = 1$ and $B_0 = 0.001$, however, our results are robust to alternative choices. The binary indicators δ are assigned a Bernoulli distribution $\delta \sim \mathcal{BER}(p)$ with prior inclusion probability $p = 0.5$. This establishes a prior that assumes contant and time-varying network dependence to be equally likely.

3.2. *Sparse finite mixtures to pool coefficients*

There are multiple possibilities to estimate observation specific parameters $\boldsymbol{\theta}_i = \{\alpha_i, \boldsymbol{\beta}_i, \sigma_i\}$ for $i = 1, \ldots, N$, with two extreme cases: either one decides to pool information over the cross-section, restricting $\boldsymbol{\theta}_i$ to be equal for all units, or one introduces truly observation specific parameters. The first restriction, especially in the empirical application of this paper considering the established literature (Ehrmann and Fratzscher, 2004; Gorodnichenko and Weber, 2016; Pasten *et al.*, 2019), is likely to be overly restrictive and may mask important structural dynamics. The second variant, however, implies estimating a large number of parameters and may thus result in imprecise estimates and overfitting issues.

In this paper, we allow for heterogeneous parameters per unit i, but introduce a hierarchical prior that exploits cross-sectional information for more precise inference and pushes similar clusters of observations towards estimated cluster-specific common means. We follow Malsiner-Walli *et al.* (2016) and introduce a sparse finite mixture of Gaussians prior for the observation specific regression coefficients $\boldsymbol{\beta}_i$, resembling a random effects specification (Verbeke and Lesaffre, 1996; Allenby *et al.*, 1998; Frühwirth-Schnatter *et al.*, 2004). The prior is given by

$$f_{\mathcal{N}}\left(\boldsymbol{\beta}_i | \{\omega\}_{m=1}^M, \{\boldsymbol{\mu}_m\}_{m=1}^M, \boldsymbol{V}\right) = \sum_{m=1}^{M} \omega_m f_{\mathcal{N}}(\boldsymbol{\beta}_i | \boldsymbol{\mu}_m, \boldsymbol{V}), \tag{3}$$

where $f_{\mathcal{N}}$ denotes the Gaussian probability density function, $\{\omega_m\}_{m=1}^M$ are mixture weights and $\{\boldsymbol{\mu}_m\}_{m=1}^M$ refer to group-specific means for a pre-determined number of M clusters. By introducing an auxiliary variable η_i, Eq. (3) can be rewritten as:

$$\boldsymbol{\beta}_i | \eta_i = m, \boldsymbol{\mu}_m, \boldsymbol{V} \sim \mathcal{N}(\boldsymbol{\mu}_m, \boldsymbol{V}),$$

with $\eta_i = m$ denoting an integer indicating that $\boldsymbol{\beta}_i$ belongs to the mth cluster. Consequently, $\Pr(\eta_i = m)$ refers to the assignment probability of industry i to cluster m. Moreover, $\boldsymbol{V} = \text{diag}(v_1, \ldots, v_K)$ denotes a common K-dimensional diagonal prior covariance matrix. We select independent inverse gamma priors for the diagonal elements

v_k $(k = 1, \ldots, K)$ of the prior covariance matrix, $v_k \sim \mathcal{G}^{-1}(d_0, d_1)$, with $d_0 = 3$ and $d_1 = 0.03$ set weakly informative.

We specify a shrinkage prior on the mixture component weights. A priori, M is chosen to be larger value (for the empirical results, we choose $M = 5$), translating into an over-fitting mixture specification. A natural choice for achieving shrinkage of the components, following Malsiner-Walli *et al.* (2016), is to assign a Dirichlet prior on $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_M)$ subject to the typical restrictions $\sum_{m=1}^{M} \omega_m = 1$ and $\omega_m \geq 0$ for $m = 1, \ldots, M$:

$$\boldsymbol{\omega}|\kappa \sim \mathcal{D}(\kappa, \ldots, \kappa).$$

Here, κ is of crucial importance, since it determines how irrelevant clusters are treated; shrinking κ empties clusters and thus ensures a parsimonious mixture representation with only a moderate number of groups. Shrinkage is achieved via a gamma distributed prior on $\kappa \sim \mathcal{G}(\vartheta, \vartheta M)$, with $\mathcal{G}$ denoting the gamma distribution with shape ϑ and scale ϑM. In the empirical application, we follow Malsiner-Walli *et al.* (2016) and define $\vartheta = 10$, introducing heavier shrinkage with increasing M.[7]

To achieve further parsimony, we combine the shrinkage prior on the weights with shrinkage on the mixture-specific means (Yau and Holmes, 2011; Malsiner-Walli *et al.*, 2016):

$$\boldsymbol{\mu}_m \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{V}_0),$$

with $\boldsymbol{\mu}_0$ referring to a common mean and $\boldsymbol{V}_0 = \boldsymbol{LRL}$ denoting the prior covariance matrix for the component means $\boldsymbol{\mu}_m$. $\boldsymbol{L} = \text{diag}(\sqrt{l_1}, \ldots, \sqrt{l_K})$ is a K-dimensional diagonal matrix, collecting the coefficient-specific shrinkage parameters l_j for $j = 1, \ldots, K$ and $\boldsymbol{R} = \text{diag}(R_1^2, \ldots, R_K^2)$ refers to a K-dimensional diagonal matrix with the jth element R_j^2 given by the range of $(\mu_{1j}, \ldots, \mu_{Mj})$. In what follows, we specify a normal gamma shrinkage prior (Griffin and Brown, 2010) on $\boldsymbol{\mu}_m$ assuming that l_j $(j = 1, \ldots, K)$ is gamma distributed,

$$l_j \sim \mathcal{G}(e_0, e_1).$$

In the empirical application, we specify $e_0 = e_1 = 0.1$.[8] To complete the setup for the regression coefficients, we specify an improper Gaussian prior on the common mean $\boldsymbol{\mu}_0 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q})$, centered on zero and with precision $\boldsymbol{Q}^{-1} = \mathbf{0}_K$.

Sofar, we remained silent on how we specify the priors for the unit-specific error variances σ_i^2. Here, we choose to cluster variances for the M groups given by $\{\tilde{\sigma}_m^2\}_{m=1}^{M}$ and

[7] Note that the expectation of κ is given by $\text{E}(\kappa) = 1/M$ and the variance is $\text{Var}(\kappa) = 1/(\vartheta M^2)$.

[8] As suggested by Malsiner-Walli *et al.* (2016), $e_0 < 1$ is important to strongly push $\boldsymbol{\mu}_m$ towards a common mean to avoid overlapping component-specific densities. This specification contrasts to Yau and Holmes (2011), who choose a Lasso prior on l_j with $e_0 = 1$ (see also Park and Casella, 2008).

use a conjugate inverse gamma hierarchical prior (Frühwirth-Schnatter, 2006),

$$\tilde{\sigma}_m^2 \sim \mathcal{G}^{-1}(\xi, \Xi), \quad \Xi \sim \mathcal{G}^{-1}(\psi, \Psi).$$

with hyperparameters specified as $\xi = 2.5 + (T-1)/2$, $\psi = 0.5 + (T-1)/2$ and $\Psi = 100\psi/(\xi R_y^2)$. R_y denotes the range of the dependent variable. The hierarchical structure again implies that the group variances $\tilde{\sigma}_m^2$ arise from a common distribution (Malsiner-Walli *et al.*, 2016). The cluster specific variance $\tilde{\sigma}_m^2$ is assigned to all observations i that are associated with the mth cluster.[9]

4. POSTERIOR COMPUTATION

Combining the likelihood of the model with the proposed prior distributions yields a set of well-known conditional posterior distributions that can be used for setting up a Markov Chain Monte Carlo (MCMC) sampling algorithm. Most of the quantities involved are standard, and we discuss details on the posteriors for the mixture model set forth in Malsiner-Walli *et al.* (2016) alongside the sampling algorithm in Appendices A and B.

Producing draws for the full history of the time-varying network dependence parameter, however, is novel to the literature. In the following, we propose a sampling algorithm for the time-varying network dependence parameter. Due to the non-Gaussian setup, Kalman-filter based methods (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994) are inapplicable. Simulation from the posterior distribution can be carried out using a Metropolis-Hastings algorithm. We denote the current state of the respective quantity by $s-1$ and s refers to a proposal from the candidate density. The procedure is similar to the algorithm proposed in the context of Bayesian stochastic volatility models in Jacquier *et al.* (2002). We rely on three proposal densities:

1. For all points in time other than the first and last observation, a draw $\rho_t^{(s)}$ is generated from the proposal distribution given by $\rho_t^{(s)} \sim \mathcal{N}(\mu_t, S_t)$, with $\mu_t = (\rho_{t-1}^{(s-1)} + \rho_{t+1}^{(s-1)})/2$ and $S_t = \varsigma^2/2$.

2. Since no initial value ρ_0 is available, we rely on Jacquier *et al.* (2002) who show that this quantity can be obtained by drawing from a Gaussian distribution $\rho_0 \sim \mathcal{N}(\mu_0, S_0)$. Under the prior $\rho_0 \sim \mathcal{N}(\mu_0, \varsigma_0^2)$, the corresponding moments are $S_0 = (\varsigma_0^2\varsigma^2)/(\varsigma_0^2 + \varsigma^2)$ and $\mu_0 = \varsigma_0^2(\mu_0/\varsigma_0^2 + \rho_1^{(s-1)}/\varsigma^2)$. The proposal at $t = 1$ is then given by $\rho_1^{(s)} \sim \mathcal{N}(\mu_1, S_1)$ where $\mu_1 = (\rho_0 + \rho_1^{(s-1)})/2$ and $S_1 = \varsigma^2/2$.

3. A similar problem arises for the final value at $t = T$, due to no ρ_{T+1} being available. Jacquier *et al.* (2002) suggest drawing from the modified candidate density $\rho_T^{(s)} \sim \mathcal{N}(\mu_T, S_T)$ with $\mu_T = \rho_{T-1}^{(s-1)}$ and $S_T = \varsigma^2$.

[9] Identification issues in mixture models arising from label switching may be resolved by implementing a random permutation sampler and ex post clustering of the posterior draws, or using economic theory to impose restrictions on the component means or variances (see Frühwirth-Schnatter, 2001).

For each point in time, we generate a proposal for $\rho_t^{(s)}$ that can be used to calculate the acceptance probability of the Metropolis-Hastings algorithm. To simplify notation, we define $\tilde{y}_{it}(\rho_t^{(s)}) = \rho_t^{(s)} \sum_{j=1}^{N} w_{ij} y_{jt} \times \sigma_i^{-1}$ and $\tilde{\boldsymbol{y}}_t(\rho_t^s) = \left(\tilde{y}_{1t}(\rho_t^{(s)}), \dots, \tilde{y}_{Nt}(\rho_t^{(s)})\right)'$ as the vector of network lags depending on the current value of $\rho_t^{(s)}$, with σ_i^2 referring to the clustered error variance assigned to industry i, and set $\tilde{\epsilon}_{it} = (y_{it} - \alpha_i - \boldsymbol{x}'_{it}\boldsymbol{\beta}_i) \times \sigma_i^{-1}$, where we again stack these quantities in $\tilde{\boldsymbol{\epsilon}}_t = (\tilde{\epsilon}_{1t}, \dots, \tilde{\epsilon}_{Nt})'$. Let

$$\mathcal{L}\left(\rho_t^{(s)}\right) = \det(\boldsymbol{I}_N - \rho_t^{(s)}\boldsymbol{W}) \times \exp\left\{-0.5\left(\tilde{\boldsymbol{\epsilon}}_t - \tilde{\boldsymbol{y}}_t(\rho_t^{(s)})\right)'\left(\tilde{\boldsymbol{\epsilon}}_t - \tilde{\boldsymbol{y}}_t(\rho_t^{(s)})\right)\right\},$$

then the acceptance probability ζ of the proposal $\rho_t^{(s)}$ implied by the likelihood is

$$\zeta = \min\left(\frac{\mathcal{L}\left(\rho_t^{(s)}\right)}{\mathcal{L}\left(\rho_t^{(s-1)}\right)}, 1\right).$$

The candidate draw $\rho_t^{(s)}$ is accepted with probability ζ, while in the opposite case, we retain the previous draw $\rho_t^{(s-1)}$. After obtaining the full history for ρ_t, it is easy to simulate the variance ς^2, and the latent binary indicator δ.

The conditional posterior of $\delta = 1|\varsigma^2$ is given by

$$\begin{aligned}
\delta = 1|\varsigma^2 &\sim \mathcal{BER}\left(u_1/(u_0 + u_1)\right), \\
u_1 &= B_1^{-1/2} \exp\{-\varsigma/2B_1\} p, \\
u_0 &= B_0^{-1/2} \exp\{-\varsigma/2B_0\}(1-p).
\end{aligned}$$

Conditional on the latent binary indicator δ and the full history of the network dependence parameter, it can be shown that the conditional posterior distribution of ς^2 is a generalized inverse Gaussian distribution. The parameter can thus be drawn using

$$\varsigma^2|\delta, \rho_1, \dots, \rho_T \sim \mathcal{GIG}\left((1-T)/2, \sum_{t=2}^{T}(\rho_t - \rho_{t-1})^2, (\delta B_1 + (1-\delta)B_0)^{-1}\right).$$

This completes the section on model estimation. We proceed by applying the proposed econometric framework to a study of time-varying effects in the transmission of US monetary policy shocks through the production network.

5. NETWORK EFFECTS OF US MONETARY POLICY

5.1. *Data and model specification*

For the sake of brevity, we only provide a brief overview of the data and refer to Ozdagli and Weber (2017) and Appendix C for more details. The event returns for industries used as dependent variables y_{it} are constructed based on returns for all common stocks trading on the NYSE, Amex or Nasdaq around press releases by the Federal Open Market Committee (FOMC). In particular, the dependent variable is defined as the difference between the last trade observation before and the first observation after the event window.

To establish the cross-sectional dependency structure via the weighting matrix $\boldsymbol{W}$, following Ozdagli and Weber (2017) we use input-output (IO) tables capturing dollar trade flows between industries published by the *Bureau of Economic Analysis* (BEA) and the *United States Department of Commerce*. Acemoglu *et al.* (2012) and Carvalho and Tahbaz-Salehi (2019) provide a theoretical justification for this choice, due to the network propagation pattern being defined by $(\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}$ which is closely related to the economy's Leontief inverse. Industries i for $i = 1, \ldots, N$ are aggregated at the four-digit IO-level, resulting in $N = 89$ cross-sectional units over time. Due to missing values, we augment the baseline Gibbs sampling scheme with an additional step for imputing these values in a Bayesian fashion (see, for instance, Gelman *et al.*, 2013).

As exogenous measure of the monetary policy shocks, we rely on high-frequency changes in Federal funds futures featured in Gorodnichenko and Weber (2016). The predetermined nature of monetary policy announcement dates (eight regular FOMC meetings per year), combined with high-frequency data on forward-looking financial instruments in a tight window of 30 minutes around the press release allows for extracting the surprise component of the monetary policy action. The tight window around the announcement reduces the risk of other events than monetary policy decisions affecting futures prices and provides support for the claim of exogeneity (see also Bernanke and Kuttner, 2005; Gürkaynak *et al.*, 2005; Altavilla *et al.*, 2019).[10] The vector $\boldsymbol{x}_{it}$ in Eq. (1) thus collapses to a scalar x_t that is common to all i, while β_i is the associated observation-specific parameter capturing the sensitivity of industry i to the monetary policy shock. Moreover, we include an industry-specific intercept term α_i. The information set includes data on FOMC announcements between early 1994 and late 2008, that is, $T = 121$.

5.2. *Empirical results*

First, we assess the importance of allowing for cross-sectional heterogeneities across industries. For this purpose, we estimate restricted versions of the general model proposed in Eq. (1) reflecting the empirical approaches of Bernanke and Kuttner (2005), Gürkaynak *et al.* (2005) and Ozdagli and Weber (2017). Second, we provide a discussion of the main findings of this paper resulting from relying on a time-varying network dependence specification.

Nested specifications and benchmarks

Table 1 displays the results for restricted versions of our model. In particular, the columns labeled BK2005/GSS2005 correspond to econometric frameworks of Bernanke and Kuttner

[10] Concerns of central bank information shocks accompanying the monetary policy announcement biasing the effects caused by pure monetary policy shocks (see Nakamura and Steinsson, 2018; Jarociński and Karadi, 2019), can be neglected for the employed dataset (for details, see Ozdagli and Weber, 2017).

Table 1: Estimated impacts of monetary policy on stock returns across industries.

	BK2005/GSS2005		Ozdagli and Weber (2017)	
	homogeneous	heterogeneous	homogeneous	heterogeneous
β	−1.183	−2.352	−0.793	−1.677
	(−1.45,−0.887)	(−2.992,−1.861)	(−1.08,−0.532)	(−2.325,−1.161)
α	−0.015	−0.031	−0.011	−0.025
	(−0.082,0.06)	(−0.185,0.125)	(−0.062,0.041)	(−0.139,0.091)
σ^2	1.134	1.009	1.126	0.976
	(0.948,1.426)	(0.592,2.614)	(0.935,1.452)	(0.563,2.673)
ρ			0.332	0.295
			(0.224,0.418)	(0.194,0.403)
Direct	−1.183	−2.352	−0.797	−1.683
	(−1.45,−0.887)	(−2.992,−1.861)	(−1.085,−0.535)	(−2.331,−1.168)
Indirect			−0.386	−0.681
			(−0.594,−0.221)	(−1.073,−0.389)
Total	−1.183	−2.352	−1.189	−2.367
	(−1.45,−0.887)	(−2.992,−1.861)	(−1.584,−0.799)	(−3.23,−1.649)
Network (%)			32.9	28.8
			(22.2,41.4)	(19.0,39.4)

Notes: The numbers refer to the estimated posterior median with the 1st and 99th percentile of the posterior distribution in parentheses. Benchmark specifications are provided by the similar setups in Bernanke and Kuttner (2005), Gürkaynak *et al.* (2005), abbreviated by BK2005 and GSS2005 respectively, and Ozdagli and Weber (2017). "Homogeneous" refers to pooling information deterministically across industries, while "heterogeneous" indicates industry-specific estimates. For the models featuring heterogeneous coefficients, we take the arithmetic mean over all industries per iteration of the algorithm and report the resulting posterior percentiles.

(2005) and Gürkaynak *et al.* (2005), disregarding cross-sectional dependency structures and network effects (that is, $\rho_1 = \ldots = \rho_T = 0$). The columns labeled Ozdagli and Weber (2017) feature network econometric models without time variation (that is, $\rho_1 = \ldots = \rho_T$). A further distinction is provided by estimating the model with homogeneous and heterogeneous coefficients. Here, "homogeneous" refers to pooling information deterministically across industries, that is $\boldsymbol{\theta}_1 = \ldots = \boldsymbol{\theta}_N$, while "heterogeneous" indicates industry-specific estimates for $i = 1, \ldots, N$ as in our baseline model.

For the models featuring heterogeneous coefficients, we take the arithmetic mean over all industries per iteration of the algorithm and report the resulting posterior percentiles (the posterior median, the 1st and 99th percentile), providing a measure of the average impact of monetary policy shocks on heterogeneous industry returns.

Negative coefficients β imply stock market responses in line with standard economic theory. Monetary tightening induces a reduction of future expected dividends, and by basic asset pricing theory, higher interest rates increase the discout rate of future dividends, resulting in stock market declines. Considering the first column of Tab. 1, a one percentage point surprise increase of the federal funds rate translates to a decline in stock market returns of about 1.2 percentage points with the 98 percent credible set ranging from approximately -0.9 to -1.5 percentage points.

This example also serves to illustrate the correspondence between interpretation of network econometric models and standard linear regressions, with obtained direct, indirect and total effects directly reflecting the regression coefficient due to the assumption of independent and identically distributed error terms. Compared to the findings of Bernanke and Kuttner (2005) and Gürkaynak *et al.* (2005), our estimates are rather small. Note, however, that the empirical findings are not directly comparable, due to their focus on the aggregate S&P 500 rather than industry-specific returns, and a different sampling period. Relaxing the assumption of parameter homogeneity, we find that the effects are roughly twice as large, where a one percent surprise in the Federal funds futures causes stock returns to decline by roughly 2.4 percentage points, on average across industries. We focus on heterogeneities over the cross-sectional dimension in the next section in more detail.

Turning to the analysis of network effects, we find that the estimated effects for the network econometric specification with pooled coefficients are smaller than those obtained by Ozdagli and Weber (2017), for two reasons. First, our proposed framework directly imputes missing values using Bayesian techniques, and thus accounts for selection bias and adequate uncertainty quantification. Second, in contrast to Ozdagli and Weber (2017) we impose the restriction $w_{ii} = 0$ to guarantee the stability of the model.[11]

The estimated total effects for the homogeneous specification are roughly in line with the effects estimated from a non-network model. Higher-order spillover dynamics explain roughly 33 percent of the total effects, with the posterior credible set ranging from 22 to 41 percent. Relaxing the assumption of homoegeneous coefficients, our estimates for the total effects are roughly in line with those of the heterogeneous non-network specification in column two, at about -2.4 percentage points on average across industries in response to a surprise monetary policy shock. The estimated contribution of network effects is slightly lower than in the homogeneous parameter case, at approximately 29 percent with 98 percent posterior intervals between 19 and 39 percent.

Allowing for time-varying network dependence and cross-sectional heterogeneity

In the following, we discuss our findings for the full model featuring time-varying network dependence and cross-sectional heterogeneities. Given the importance of industry-specific idiosyncrasies identified in previous contributions (Ehrmann and Fratzscher, 2004; Gorodnichenko and Weber, 2016; Pasten *et al.*, 2019) and suggested by Tab. 1, we begin by discussing our findings for the heterogeneous regression coefficients. The mixture model provides substantial support for one common cluster, however, roughly in 25 percent of the draws we find evidence for two clusters. Differences mainly originate from idiosyncrasies in the sensitivity of industries to the monetary policy shocks, while the intercept terms α_i are pushed more strongly towards their common mean. The error variances $\tilde{\sigma}_m^2$ are heavily shrunk towards homogeneity.

Figure 1 shows the distribution of the posterior medians across industries in form of a boxplot. The upper panel displays the intercept α_i, while the middle panel depicts the

[11] Our results are comparable to the provided robustness check in Ozdagli and Weber (2017) where the main diagonal of $\boldsymbol{W}$ is set to zero, accounting for posterior uncertainty of the estimates.

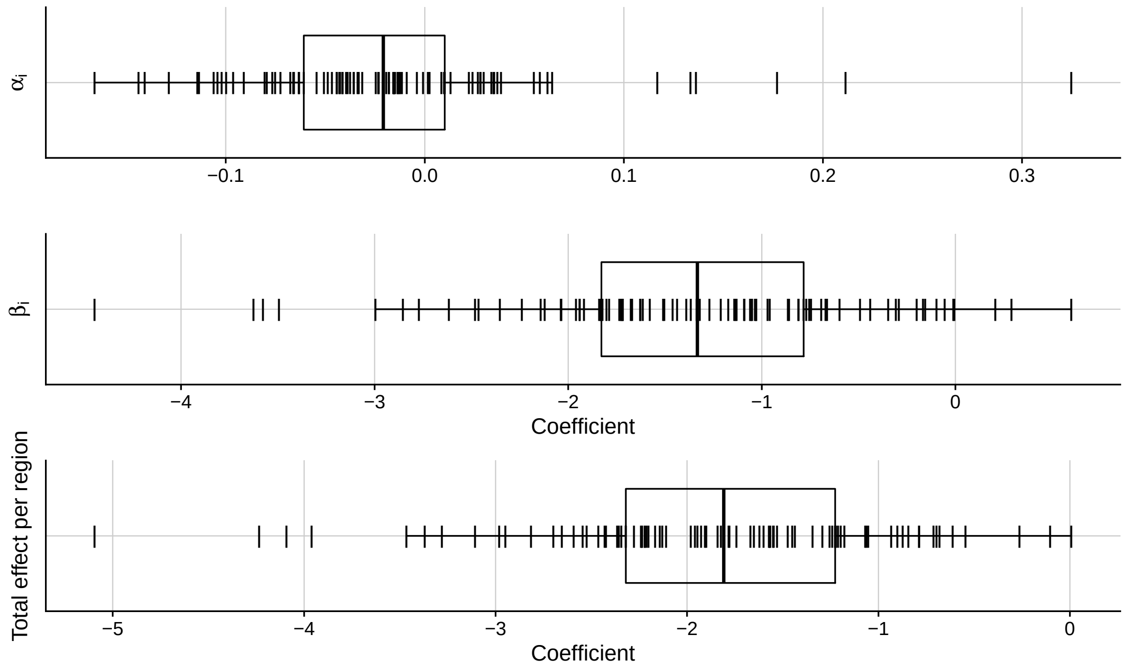

Fig. 1: Posterior median of the regression coefficients across industries.

Note: The solid black line indicates the median of the estimated posterior medians across industries, while the box covers the 25th and 75 quartile. Individual lines denote industries.

coefficients β_i associated with the monetary policy shock. Starting with the intercept coefficients α_i, half of the industries exhibit estimates between -0.05 and 0.01. This implies that the average return across industries around FOMC announcements is negative. Some few industries exhibit positive coefficients, however, these observations typically feature large posterior uncertainty surrounding the median. A substantially larger number of industries exhibit more pronounced negative estimates.

Turning to the industry-specific impacts of monetary policy surprises captured by β_i, we find that the median across industries is approximately -1.3 percentage points in response to a one percentage point increase in the instrument. This is roughly in line with our findings for the non-network specification featuring homogeneous coefficients in Tab. 1. As discussed in the context of how to interpret such network models, however, the regression coefficients cannot be interpreted directly. For this purpose, we calculate the total effect per region which corresponds to the row sums of $\boldsymbol{S}_{kt}$ depicted in the bottom panel of Fig. 1.

For simplicity, we consider the average effect over time and refer to the following paragraphs for information on time-variation of the estimates. The median impact across industries and over time to a one percent surprise increase in Federal funds futures around announcement dates is about -1.9 percentage points, with half of the industries showing declines in stock returns between -1.3 to -2.4 percentage points. We find that effects for all industries are negative, with a substantial number exhibiting effects higher than -2.5 up to more than -5 percentage points. Considering hypothetical responses to a surprise 25

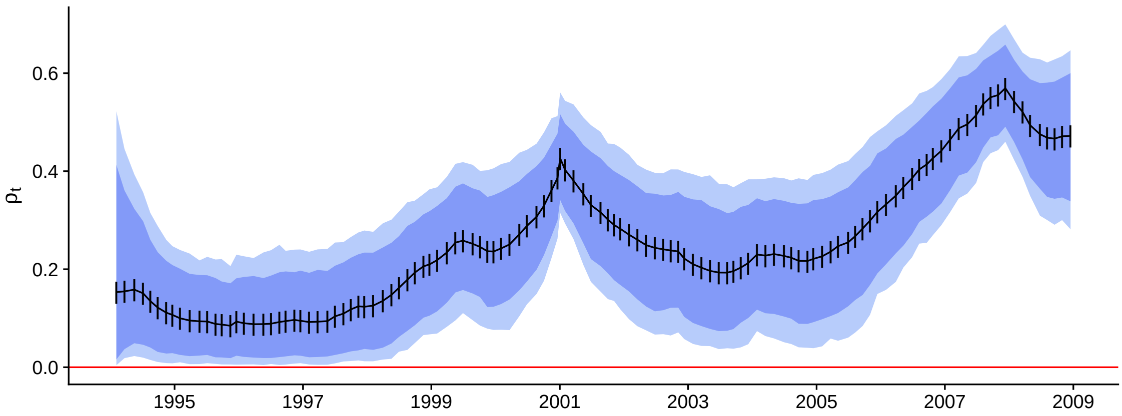

Fig. 2: Time-varying network dependence parameter.

Note: The solid black line indicates the posterior median, alongside the 98 and 90 percent credible sets in shaded blue. FOMC announcement dates are indicated by the black vertical lines.

basis points interest rate hike, this implies that a number of industries shows stock market return declines exceeding 0.75 percent. These findings provide further strong empirical evidence for the necessity of considering heterogeneous coefficients and a sophisticated econometric approach; they can be explained partly by the theoretical channels provided in Ehrmann and Fratzscher (2004) and Pasten *et al.* (2019).

We proceed with discussing our findings for time-varying network dependence. Figure 2 shows the evolution of ρ_t over time. The solid black line indicates the posterior median, alongside the 98 and 90 percent credible sets in shaded blue. FOMC announcement dates are indicated by the black vertical lines. Inference on the binary indicator dictating time variation shows that likelihood information strongly suggests time-varying network dependence, with $\delta = 1$ for all iterations of the sampling algorithm, translating into substantial differences in the parameter over time. This result corroborates findings from previous authors who suggest that the sensitivity of the stock market to monetary policy surprises changes over time and is related to the current state of the economy and the business cycle (see Chen, 2007; Basistha and Kurov, 2008; Kurov, 2010).

Interestingly, higher importance of network dependence appears to occur during recessions. Relevant recessionary episodes for the employed dataset identified by the *NBER Business Cycle Dating Committee* are the mild recession between March and November 2001, and the global financial crisis and subsequent Great Recession from December 2007 to June 2009. Largest magnitudes for ρ_t are detected during the Great Recession, with median estimates of roughly 0.55, followed by the first recession in the sample in early 2001 of approximately 0.4. The initial period covered by the sample is characterized by small contributions of network effects, with median estimates just below 0.2, while the time between the two recessions exhibits a network dependence parameter around 0.25 increasing gradually towards the outbreak of the global financial crisis.

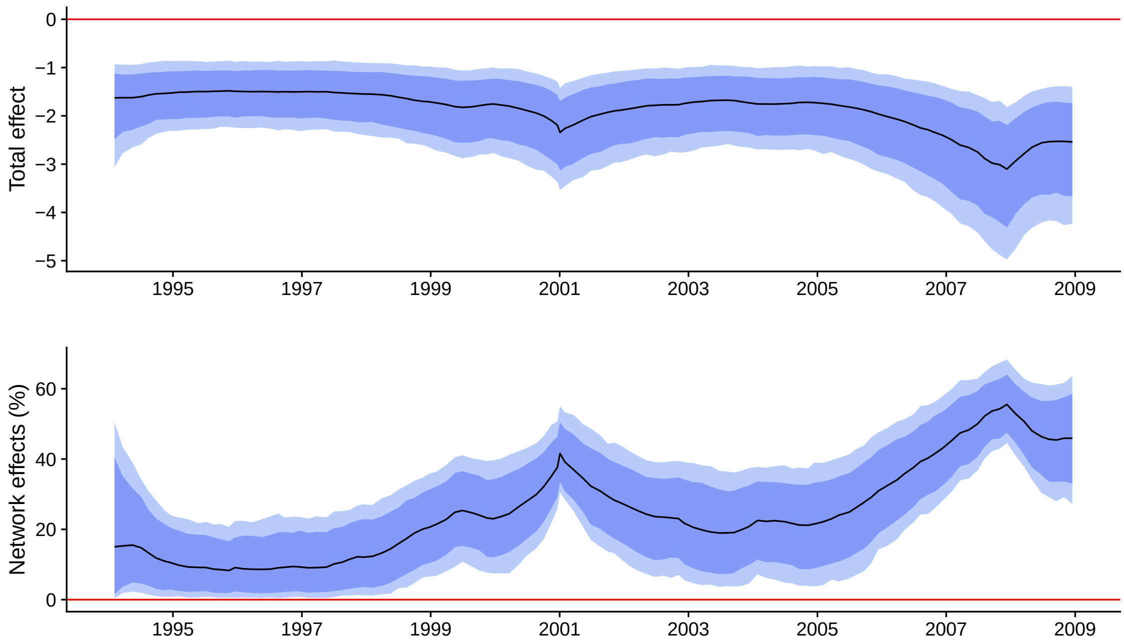

Fig. 3: Total effects of monetary policy shocks and share attributed to network effects.
Note: The solid black line indicates the posterior median, alongside the 98 and 90 percent credible sets in shaded blue.

In light of the alternative interpretation of ρ_t as a parameter related to capturing common stochastic volatility, one may interpret the network coefficient as a factor related to systemic risk (see, for instance, Billio *et al.*, 2016a). Our results imply that during recessions, idiosyncratic shocks to specific industries spill over more strongly to others and are not sterilized on aggregate. Acemoglu *et al.* (2012) and Ozdagli and Weber (2017) argue that in light of these cascade effects, monetary policy authorities must carefully consider future policy actions to stabilize financial markets. This is particularly important in the context of policies affecting systemically relevant firms (or industries) that are "too big to fail" due to their interconnectedness (Acemoglu *et al.*, 2012).

It remains to quantify the overall importance of the network effects over time. Figure 3 shows the posterior distribution of the estimated total effects in the upper panel, while the lower panel depicts the relative share of indirect effects in percent. Movements in the strength of network dependence observed in Fig. 2 clearly translate to changes in the role of network effects of monetary transmission channels. The upper panel of Fig. 3 suggests that we do not only observe substantial difference over the cross-section, but also over time. In particular, the effect size before the first, and between the two recessions covered by the sample roughly correspond to the average effect of monetary policy surprises on stock returns over time of about -1.9 percentage points. The recessionary episodes, however, exhibit larger effects up to three percentage points on average caused by a contractionary one percent policy surprise in the Federal funds futures. This result corresponds to Chen (2007) and Basistha and Kurov (2008).

Considering the lower panel of Fig. 3, we find that not only effect sizes are larger during recessions, but also that network effects tend to increase. During the recession of 2001, about 40 percent of the total effect sizes can be attributed to spillover effects, while this effect increases to approximately 60 percent during the Great Recession. This result links our findings to Kastner (2019), who finds pronounced increases in co-movement across industries during periods of economic turmoil. Expansionary economic episodes, on the other hand, show muted network effects with magnitudes of roughly 20 to 30 percent.

6. CLOSING REMARKS

This paper studies the importance of spillover effects in the transmission of monetary policy shocks through the US production network. We propose a novel Bayesian network panel state-space model to capture time-variation in the magnitude of network effects. Moreover, we address industry-specific heterogeneities via a sparse finite Gaussian mixture prior on the model coefficients. Our results suggest substantial differences in industry responses, and identify recessionary episodes as periods where network effects play a crucial role. The findings are in line with the established literature (Ehrmann and Fratzscher, 2004; Chen, 2007; Basistha and Kurov, 2008; Ozdagli and Weber, 2017), and suggest time-varying higher-order spillover effects to be of crucial importance in explaining the stock market's response to monetary policy.

REFERENCES

Acemoglu D, Carvalho VM, Ozdaglar A, and Tahbaz-Salehi A (2012), "The network origins of aggregate fluctuations," *Econometrica* **80**(5), 1977–2016.

Acemoglu D, Ozdaglar A, and Tahbaz-Salehi A (2015), "Systemic risk and stability in financial networks," *American Economic Review* **105**(2), 564–608.

Allenby GM, Arora N, and Ginter JL (1998), "On the heterogeneity of demand," *Journal of Marketing Research* **35**(3), 384–389.

Altavilla C, Brugnolini L, Gurkaynak RS, Motto R, and Ragusa G (2019), "Measuring euro area monetary policy," *Journal of Monetary Economics* **forthcoming**.

Aquaro M, Bailey N, and Pesaran MH (2015), "Quasi maximum likelihood estimation of spatial models with heterogeneous coefficients," *CESifo Working Paper* **5428**.

Asgharian H, Hess W, and Liu L (2013), "A spatial analysis of international stock market linkages," *Journal of Banking & Finance* **37**(12), 4738–4754.

Basistha A, and Kurov A (2008), "Macroeconomic cycles and the stock market's reaction to monetary policy," *Journal of Banking & Finance* **32**(12), 2606 – 2616.

Bernanke BS, and Kuttner KN (2005), "What explains the stock market's reaction to Federal Reserve policy?" *Journal of Finance* **60**(3), 1221–1257.

Bianchi D, Billio M, Casarin R, and Guidolin M (2015), "Modeling contagion and systemic risk," *SYRTO Working Paper Series* **10**.

Billio M, Caporin M, Panzica R, and Pelizzon L (2016a), "The impact of network connectivity on factor exposures, asset pricing and portfolio diversification," *SAFE Working Paper Series* **166**.

Billio M, Casarin R, Ravazzolo F, and Van Dijk HK (2016b), "Interconnections between eurozone and US booms and busts using a Bayesian panel Markov-switching VAR model," *Journal of Applied Econometrics* **31**(7), 1352–1370.

Carter CK, and Kohn R (1994), "On Gibbs sampling for state space models," *Biometrika* **81**(3), 541–553.

Carvalho VM, and Tahbaz-Salehi A (2019), "Production networks: A primer," *Annual Review of Economics* **11**.

Chen SS (2007), "Does monetary policy have asymmetric effects on stock returns?" *Journal of Money, Credit and Banking* **39**(2-3), 667–688.

Ciccarelli M, Maddaloni A, and Peydró JL (2013), "Heterogeneous transmission mechanism: monetary policy and financial fragility in the eurozone," *Economic Policy* **28**(75), 459–512.

Cochrane JH, and Piazzesi M (2002), "The FED and interest rates–A high-frequency identification," *American Economic Review* **92**(2), 90–95.

Cogley T, and Sargent TJ (2005), "Drifts and volatilities: monetary policies and outcomes in the post WWII US," *Review of Economic Dynamics* **8**(2), 262–302.

Cornwall GJ, and Parent O (2017), "Embracing heterogeneity: the spatial autoregressive mixture model," *Regional Science and Urban Economics* **64**, 148 – 161.

Ehrmann M, and Fratzscher M (2004), "Taking stock: Monetary policy transmission to equity markets," *Journal of Money, Credit and Banking* 719–737.

Elhorst JP (2014), *Spatial econometrics: from cross-sectional data to spatial panels*, Berlin, Heidelberg: Springer.

Elliott M, Golub B, and Jackson MO (2014), "Financial networks and contagion," *American Economic Review* **104**(10), 3115–53.

Frühwirth-Schnatter S (1994), "Data augmentation and dynamic linear models," *Journal of Time Series Analysis* **15**(2), 183–202.

Frühwirth-Schnatter S (2001), "Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models," *Journal of the American Statistical Association* **96**(453), 194–209.

——— (2006), *Finite mixture and Markov switching models*, Berlin/Heidelberg: Springer Science+Business Media.

Frühwirth-Schnatter S, Tüchler R, and Otter T (2004), "Bayesian analysis of the heterogeneity model," *Journal of Business & Economic Statistics* **22**(1), 2–15.

Frühwirth-Schnatter S, and Wagner H (2010), "Stochastic model specification search for Gaussian and partial non-Gaussian state space models," *Journal of Econometrics* **154**(1), 85–100.

Gabaix X (2011), "The granular origins of aggregate fluctuations," *Econometrica* **79**(3), 733–772.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, and Rubin DB (2013), *Bayesian data analysis*, Boca Raton, London and New York: Chapman and Hall/CRC.

George EI, and McCulloch RE (1993), "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association* **88**(423), 881–889.

Gertler M, and Karadi P (2015), "Monetary policy surprises, credit costs, and economic activity," *American Economic Journal: Macroeconomics* **7**(1), 44–76.

Gorodnichenko Y, and Weber M (2016), "Are sticky prices costly? Evidence from the stock market," *American Economic Review* **106**(1), 165–99.

Griffin JE, and Brown PJ (2010), "Inference with normal-gamma prior distributions in regression problems," *Bayesian Analysis* **5**(1), 171–188.

Gürkaynak RS, Sack BP, and Swanson ET (2005), "Do actions speak louder than words? The response of asset prices to monetary policy actions and statements," *International Journal of Central Banking* **1**(1).

Jacquier E, Polson NG, and Rossi PE (2002), "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business & Economic Statistics* **20**(1), 69–87.

Jarociński M, and Karadi P (2019), "Deconstructing monetary policy surprises: The role of information shocks," *American Economic Journal: Macroeconomics* **forthcoming**.

Kastner G (2019), "Sparse Bayesian time-varying covariance estimation in many dimensions," *Journal of Econometrics* **210**(1), 98 – 115.

Kim CJ, and Nelson CR (1999), *State-space models with regime switching: Classical and Gibbs-sampling approaches with applications*, Cambridge and London: MIT press.

Kontonikas A, MacDonald R, and Saggu A (2013), "Stock market reaction to fed funds rate surprises: State dependence and the financial crisis," *Journal of Banking & Finance* **37**(11), 4025–4037.

Koop G (2003), *Bayesian Econometrics*, John Wiley & Sons Inc.

Kurov A (2010), "Investor sentiment and the stock market's reaction to monetary policy," *Journal of Banking & Finance* **34**(1), 139 – 149.

Kuttner KN (2001), "Monetary policy surprises and interest rates: Evidence from the Fed funds futures market," *Journal of Monetary Economics* **47**(3), 523–544.

LeSage JP, and Chih YY (2016), "Interpreting heterogeneous coefficient spatial autoregressive panel models," *Economics Letters* **142**, 1 – 5.

LeSage JP, and Pace RK (2009), *Introduction to spatial econometrics*, Boca Raton, FL: Taylor & Francis Group, LLC.

Malsiner-Walli G, Frühwirth-Schnatter S, and Grün B (2016), "Model-based clustering based on sparse finite Gaussian mixtures," *Statistics and Computing* **26**(1), 303–324.

Nakamura E, and Steinsson J (2018), "High-frequency identification of monetary non-neutrality: the information effect," *Quarterly Journal of Economics* **133**(3), 1283–1330.

Ozdagli A, and Weber M (2017), “Monetary policy through production networks: Evidence from the stock market,” *National Bureau of Economic Research* **w23424**.
Park T, and Casella G (2008), “The Bayesian Lasso,” *Journal of the American Statistical Association* **103**(482), 681–686.
Pasten E, Schoenle R, and Weber M (2019), “The propagation of monetary policy shocks in a heterogeneous production economy,” *Journal of Monetary Economics* **forthcoming**.
Primiceri GE (2005), “Time Varying Structural Vector Autoregressions and Monetary Policy,” *Review of Economic Studies* **72**(3), 821–852.
Sims CA, and Zha T (2006), “Were there regime switches in US monetary policy?” *American Economic Review* **96**(1), 54–81.
Verbeke G, and Lesaffre E (1996), “A linear mixed-effects model with heterogeneity in the random-effects population,” *Journal of the American Statistical Association* **91**(433), 217–221.
Yau C, and Holmes C (2011), “Hierarchical Bayesian nonparametric mixture models for clustering with variable relevance determination,” *Bayesian Analysis* **6**(2), 329–351.

A. POSTERIORS FOR MIXTURE-SPECIFIC QUANTITIES

For simplicity we suppress observation-specific intercepts $\{\alpha_i\}_{i=1}^N$ in Appendix A. Conditional on the observation-specific coefficients $\{\boldsymbol{\beta}_i\}_{i=1}^N$, variances $\{\sigma_i^2\}_{i=1}^N$ and the group-allocation indicators $\{\eta\}_{i=1}^N$, the posteriors for the prior mean and covariance matrix are independent from the data and can be drawn using standard results from linear regression models (see, for instance, Koop, 2003).

Given draws for the group-allocation indicators $\boldsymbol{\eta} = (\{\eta\}_{i=1}^N)$, the posterior distribution of the mixture probabilities follows a Dirichlet distribution:

$$\boldsymbol{\omega}|\boldsymbol{\eta} \sim \mathcal{D}(\kappa_1, \dots, \kappa_M),$$

Here, we define $\kappa_m = \kappa + N_m$ with N_m referring to the number of industries assigned to cluster m. Conditional on the group means $\{\boldsymbol{\mu}_m\}_{m=1}^M$, the common prior prior covariance matrix $\boldsymbol{V}$ and the mixture weights $\boldsymbol{\omega}$, the regime indicators $\boldsymbol{\eta}$ follow a multinomial distribution with

$$\Pr(\eta_i = m|\omega_m, \boldsymbol{\mu}_m, \boldsymbol{V}) \propto \omega_m f_{\mathcal{N}}(\boldsymbol{\beta}_i|\boldsymbol{\mu}_m, \boldsymbol{V}), \quad \text{for } m = 1, \dots, M.$$

The full conditional posterior of $\boldsymbol{\mu} = \text{vec}(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_M)$ follows a multivariate Gaussian distribution with diagonal covariance matrix:

$$\boldsymbol{\mu}|\boldsymbol{V}, \boldsymbol{\eta}, \boldsymbol{\mu}_0, \boldsymbol{V}_0 \sim \mathcal{N}\left(\overline{\boldsymbol{\mu}}, \overline{\boldsymbol{V}_\mu}\right),$$

with the posterior variance and mean given by,

$$\begin{aligned} \overline{\boldsymbol{V}_\mu} &= \left(\boldsymbol{V}^{-1} \otimes \boldsymbol{H}'\boldsymbol{H} + \boldsymbol{I}_M \otimes \boldsymbol{V}_0^{-1}\right)^{-1}, \\ \overline{\boldsymbol{\mu}} &= \overline{\boldsymbol{V}_\mu}\left(\boldsymbol{V}^{-1} \otimes \boldsymbol{H}'\boldsymbol{\beta} + \boldsymbol{\iota}_M \otimes \boldsymbol{V}_0^{-1}\boldsymbol{\mu}_0\right). \end{aligned}$$

Here, $\boldsymbol{H}$ is an $N \times M$ matrix with ith row given by $\boldsymbol{H}_i = (I(\eta_i = 1), \dots, I(\eta_i = M))$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_N)'$ is a KN-dimensional vector and $\boldsymbol{\iota}_M$ is a M-dimensional vector of ones. Conditional on $\boldsymbol{V}_0$ and $\boldsymbol{\mu}$, the conditional posterior distribution of the common mean $\boldsymbol{\mu}_0$ is

$$\boldsymbol{\mu}_0|\boldsymbol{V}_0, \boldsymbol{\mu} \sim \mathcal{N}\left(\frac{\sum_{m=1}^M \boldsymbol{\mu}_m}{M}, \frac{1}{M}\boldsymbol{V}_0\right).$$

The final ingredient for the mixture are the shrinkage parameters $l_1, \ldots, l_K$. Conditional on $\boldsymbol{R}$ and $\boldsymbol{\mu}$ the conditional posterior distribution is given by:

$$l_j|\boldsymbol{R}, \boldsymbol{\mu} \sim \mathcal{GIG}\left(e_0 - M/2, 2e_1, \frac{\sum_{m=1}^{M}(\mu_{mj} - \mu_{0j})^2}{R_j}\right),$$

with μ_{mj}, μ_{0j} and R_j, for $m = 1, \ldots, M$ denoting the jth element of the component-specific means, common mean and of $\boldsymbol{R}$, respectively.

Finally, we sample $\tilde{\sigma}_m^2$, (for $m = 1, \ldots M$) from a inverse gamma conditional posterior distribution given by

$$\begin{aligned}
\tilde{\sigma}_m^2|\bullet &\sim \mathcal{G}^{-1}(\overline{\xi}_m, \overline{\Xi}_m) \\
\overline{\Xi}_m &= \Xi + \frac{1}{2}\sum_{i:\eta_i=m}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_i - \alpha_i)'(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_i - \alpha_i). \\
\overline{\xi}_m &= \xi + TN_m/2
\end{aligned}$$

The common scaling indicator Ξ_m is also drawn from an inverse gamma conditional,

$$\begin{aligned}
\Xi|\{\sigma^2\}_{m=1}^{M}, \Psi, \psi &\sim \mathcal{G}^{-1}(\overline{\psi}, \overline{\Psi}) \\
\overline{\Psi} &= \Psi + \sum_{m=1}^{M}\tilde{\sigma}_m^{-2}, \\
\overline{\psi} &= \psi + M\xi.
\end{aligned}$$

B. MCMC ALGORITHM

The set of conditional posterior distributions in Section 3 and Appendix A is used to generate draws for all parameters of the model by a standard MCMC sampling algorithm. Specifically, the sampler iterates through the following steps:

1. Sample the observation-specific regression coefficients $\boldsymbol{\beta}_i$ on an equation-by-equation basis. The posterior takes a standard form (see, for instance Koop, 2003) conditional on the full history of the network dependence parameter $\{\rho_t\}_{t=1}^{T}$.

2. Given draws for the observation-specific coefficients $\{\boldsymbol{\beta}_i\}_{i=1}^{N}$, variances $\{\sigma_i^2\}_{i=1}^{N}$ and the group-allocation indicators, the posteriors for the prior mean and covariance matrix are independent from the data. For the corresponding posterior distributions and moments, see Appendix A and Malsiner-Walli *et al.* (2016).

3. Conditional on $\{\boldsymbol{\beta}_i\}_{i=1}^{N}$ and the group-allocation indicators, the posterior for the cluster-specific variances $\{\sigma_i^2\}_{i=1}^{N}$ can be sampled based on the quantities provided again in Appendix A and Malsiner-Walli *et al.* (2016).

4. Obtaining draws for the full history of the network dependence parameter ρ_t is achieved by employing the proposed Metropolis-Hastings algorithm in Section 3.

5. Conditional on $\{\rho_t\}_{t=1}^{T}$ it is easy to sample the process variances ς^2. Given ς^2, a draw from the posterior of the binary indicator δ that governs time-variation is obtained using the corresponding posteriors Section 3.

6. Given a draw for all model parameters, it is straightforward to obtain a draw for the missing values in the dependent variable (Gelman *et al.*, 2013).

This completes the MCMC algorithm employed to simulate from the posterior distribution. After choosing starting values and a sufficient burn-in period we store draws from the conditional posterior distributions. In particular, we discard the initial $4,000$ draws, while Bayesian inference is performed based on each third of the subsequent $6,000$ draws resulting in a set of $2,000$ draws from the posterior.

C. DATA

In this section we discuss some important patterns of the cross-sectional weights matrix $\boldsymbol{W}$ derived from IO-tables that was kindly provided by Michael Weber. Figure C.1 shows the constructed industry-by-industry flows based on the IO-tables for 1992. The left panel is the employed $\boldsymbol{W}$ matrix (with zero diagonal), while the right panel shows the diagonal capturing flows within each industry. At industry level, the network is sparse, while at the same time a small number of industries supply to a wide range of other customers across different industries (see also Acemoglu *et al.*, 2012; Carvalho and Tahbaz-Salehi, 2019). Therefore, although not all industries are directly linked to each other, these general purpose industries provide a connection of almost all industries already after the first round of higher-order effects.

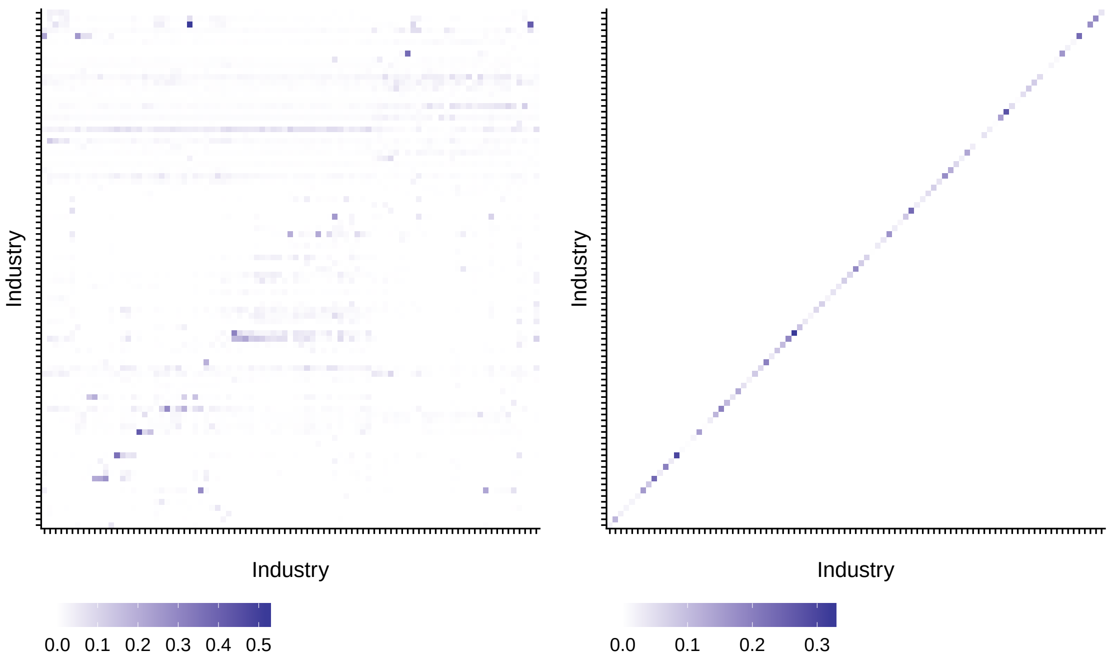

Fig. C.1: Normalized industry-by-industry flows based on IO-tables for 1992.

Note: The left panel shows the employed linkage matrix $\boldsymbol{W}$ (with zero diagonal), that was previously used in Ozdagli and Weber (2017, with non-zero diagonal elements, depicted in the right panel).