SPATIAL ECONOMIC ANALYSIS
https://doi.org/10.1080/17421772.2025.2482071

Fast two-stage variational Bayesian approach to estimating panel spatial autoregressive models with unrestricted spatial weights matrices

Deborah Gefang [a], Stephen G. Hall [a,b,c] and George S. Tavlas [b,d]

ABSTRACT

We propose a fast two-stage variational Bayesian (VB) algorithm to estimate unrestricted panel spatial autoregressive models. Using Dirichlet–Laplace shrinkage priors, we uncover the spatial relationships between cross-sectional units without imposing any a priori restrictions. Monte Carlo experiments show that our approach works well for both long and short panels. We are also the first in the literature to develop VB methods to estimate large covariance matrices with unrestricted sparsity patterns, which are useful for popular large data models such as Bayesian vector autoregressions. In empirical applications, we examine the spatial interdependence between euro area sovereign bond ratings and spreads.

KEYWORDS

Variational inference, spatial panel data models, simultaneous equations, large data

JEL C11, C33, C55

HISTORY Received 21 August 2023; in revised form 19 February 2025

1. INTRODUCTION

The spatial autoregressive (SAR) models, first proposed by Cliff and Ord (1973), have been widely used in the literature to investigate spatial dependence in cross-sectional units (e.g., Anselin, 1988; Baltagi et al., 2003, 2013; Lee & Yu, 2010). In practice, the spatial weights matrices are usually set a priori based on geographical distances or economic characteristics (e.g., Cliff & Ord, 1973; Anselin, 1988; Case, 1991). This is not surprising, as a spatial weights matrix potentially involves $N^2 - N$ interrelationships between N spatial units, which makes it difficult to estimate, especially when N is large.

In recent years, a number of variable selection and parameter shrinkage methods have been developed to estimate the spatial weights matrices of panel SAR models. Among them, methods resorting to the least absolute shrinkage and selection operator (Lasso) of Tibshirani (1996) and its variants have gained a lot of attention. For example, Basak et al. (2018) propose estimating a

CONTACT Deborah Gefang dg171@le.ac.uk
[a]Department of Economics, University of Leicester, Leicester, UK
[b]Bank of Greece, Athens, Greece
[c]Department of Economics, University of Pretoria, Hatfield, South Africa
[d]Hoover Institution, Stanford University, Stanford, CA, USA

Supplemental data for this article can be accessed online at https://doi.org/10.1080/17421772.2025.2482071

triangular weights matrix under the assumption of recursive ordering. Ahrens and Bhattacharjee (2015) develop a two-step Lasso estimator to identify the weights matrix. Lam and Souza (2020) estimate the weights matrix using adaptive Lasso with sparse adjustment in mind. Most of the studies, however, usually impose sometimes unrealistic restrictions on the model's coefficients or covariances. Krock et al. (2021) develop a graphical Lasso approach to estimating the unrestricted covariances. Their method, however, does not address the impacts of any possible exogenous variables. Moreover, to our knowledge, few of these studies focus on short panels where N is large while T is small. Only recently, Krisztin and Piribauer (2023) and Piribauer et al. (2023) used a hierarchical prior setup to identify sparsity when N exceeds T by a large margin, assuming the same spatial parameter for all the cross-sectional units.

This paper contributes to the SAR literature by developing a fast, two-stage variational Bayesian (VB) approach to estimating panel SAR models with unknown spatial weights matrices. We do not impose any restrictions on the spatial weights matrix or the covariance functions, allowing our approach to let the data speak. The prior we used for Bayesian regularisation is the Dirichlet–Laplace (D-L) prior of Bhattacharya et al. (2015). The D-L prior is appealing because practitioners only need to select a single hyperparameter, making it more tractable than many other shrinkage priors.[1]

Our second contribution is to develop VB methods to estimate large covariance matrices with a global-local shrinkage prior. We are among the first in the literature to develop a VB estimator for large covariance matrices with unknown sparsity patterns. Our VB methods using the D-L prior can be easily extended to accommodate other popular priors, such as the graphical Lasso of Wang (2012), the half-Cauchy prior of Makalic and Schmidt (2016), and the graphical horseshoe prior of Li et al. (2019). This is non-trivial because VB is a more computationally efficient alternative to Markov Chain Monte Carlo (MCMC), and our approach can be applied to estimate other popular models involving large covariance matrices, such as large Bayesian vector autoregressions (BVARs).

We have conducted a wide range of simulation studies using both dense and sparse spatial weights matrices.[2] Monte Carlo experiments demonstrate that the two-stage VB approach is accurate and computationally efficient when $T \gg N$, which is typically more relevant for macroeconomic and financial data. When $N \gg T$, which is often the case with microeconomic data, two-stage VB estimates tend to exhibit slightly larger biases, depending on the priors. However, there is ample evidence that two-stage VB can recover the true spatial weights matrix well.

In empirical applications, we use two-stage VB to estimate a two-equation simultaneous spatial model using panel data comprising sovereign bond ratings and spreads of ten eurozone countries. The research is motivated by the importance of gaining a deeper understanding of cross-country contagion dynamics between southern and northern euro area countries, particularly during events like the European debt crisis (e.g., Gibson et al., 2021; Hall et al., 2023). Utilising rolling windows of 2 years, we compare how changes in sovereign ratings and spreads in the south versus the north affect each individual member country. Our findings reveal significant differences between these two groups: shocks originating in the south or north impact specific countries differently, and shocks affecting one country group have varying impacts on southern and northern countries. Furthermore, our results underscore the substantial impacts of global financial crises, showing severe effects on southern euro area countries during the European debt crisis, while northern countries were relatively less affected.

The rest of the paper is organised as follows. Section 2 extends the traditional panel SAR models to an unrestricted panel SAR. Section 3 develops the two-stage VB methodology. Section 4 presents Monte Carlo studies. Section 5 applies the two-stage VB approach to sovereign bond ratings and spreads data from ten euro area countries. Section 6 provides concluding remarks. Detailed VB derivation formulas and more extensive Monte Carlo results can be found in the Appendices A to D in the online supplemental data. Sources for data used in the empirical example are listed in Appendix E in the online supplemental data.

2. UNRESTRICTED PANEL SAR MODEL

In this section, we start from a traditional standard panel SAR and then relax the restrictions imposed upon it in steps, with the aim of giving a flavour of the differences between the traditional model and the unrestricted panel SAR model that we set out to estimate using two-stage VB.

Let Y, X and V denote the $T \times N$ matrix of endogenous variables, $T \times (Nm)$ matrix of exogenous variables, and $T \times N$ matrix of disturbances, respectively. A traditional panel SAR model takes the following form:

$$y_t = \lambda W_n y_t + X_t \beta + u_t, \quad |\lambda| < 1 \tag{1}$$

where $y_t = (Y_{t1}, Y_{t2}, \ldots, Y_{tN})'$ is the $N \times 1$ vector of observations of the dependent variables, W_n is an $N \times N$ known spatial weights matrix with zero diagonal entries, $X_t = \begin{pmatrix} X_{t,1} \\ X_{t,2} \\ \ldots \\ X_{t,N} \end{pmatrix}$ is the $N \times m$ matrix of exogenous variables, with $X_{t,i}$ denoting the $1 \times m$ row vector of exogenous variables associated with dependent variable y_{ti}, β is an $m \times 1$ vector of parameters, λ is a scalar parameter, and $u_t = (V_{t1}, V_{t2}, \ldots, V_{tN})'$ is the $N \times 1$ i.i.d. (independent and identically distributed) error terms with mean zero and diagonal covariance matrix $\begin{pmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ .. & .. & \ldots & .. \\ 0 & 0 & \ldots & \sigma^2 \end{pmatrix}$.

Model (1) imposes the following unrealistic restrictions on the data generating process: (1) W_N is predetermined, in a fashion that is not related to the variations in the data; (2) λ and β remain the same across equations associated with different dependent variables and (3) the covariance matrix of u_t is diagonal with the same diagonal entries.

Relaxing those restrictions, model (1) can be written as:

$$y_t = \tilde{\lambda} \odot (\tilde{W}_N y_t) + \tilde{X}_t \tilde{\beta} + u_t, \quad |\tilde{\lambda}|_\infty < 1 \tag{2}$$

where $\tilde{W}_N$ is an $N \times N$ unknown spatial weight matrix with zero diagonal entries, $\tilde{\lambda}$ is a $N \times 1$ parameter vector, $\odot$ is the Hadamard product, $\tilde{X}_t = \begin{pmatrix} X_{t,1} & 0 & \ldots & 0 \\ 0 & X_{t,2} & \ldots & 0 \\ .. & .. & \ldots & .. \\ 0 & 0 & \ldots & X_{t,N} \end{pmatrix}$, $\tilde{\beta}$ is a $Nm \times 1$ parameter vector, and u_t is i.i.d. with mean zero and diagonal covariance matrix $\tilde{\Sigma}$ with diagonal entries that can be different from each other. In this model, the dimensions of $X_{t,i}$ and $X_{t,j}$ for $i \neq j$ can differ from each other. Let the dimension of $X_{t,i}$ be $1 \times m_i$. The dimension of the parameter vector $\tilde{\beta}$ is thus$(\sum_{i=1}^{n} m_i) \times 1$.

Note that model (2) is quite flexible. For example, with appropriate restrictions, it can be easily transformed back into the traditional form described in (1) or a panel SAR containing the simultaneous cross-sectional spatial relationship as described in Yang and Lee (2017) and Liu and Saraiva (2019).

If our main concerns in panel SAR models are the spillover effects, there would be little research interest in separately identifying $\tilde{\lambda}$ and $\tilde{W}_N$. What we care about is the product

$(\tilde{\lambda} \otimes l_N) \odot \tilde{W}_N$, where l_N is a $N \times 1$ column of ones and $\otimes$ is the Kronecker product, as $(\tilde{\lambda} \otimes l_N) \odot \tilde{W}_N$ is the $N \times N$ parameter matrix which captures the spillover effects between spatial units.

Let $\Lambda = (\tilde{\lambda} \otimes l_N) \odot \tilde{W}_N$. Model (2) can be written as:

$$y_t = \Lambda y_t + \tilde{X}_t \tilde{\beta} + u_t, \tag{3}$$

where $I_N - \Lambda$ is non-singular and the characteristic roots of $I_N - \Lambda$ lie within the unit circle.

The attractiveness of model (3) is that it turns an unrestricted panel SAR model into a system of simultaneous equations (SEM). As shown in Zellner and Theil (1962) and Fox (1979), the i^{th} equation in model (3) is just identified if $\sum_{i=1}^{N} m_i = N - 1 + m_i$ and over-identified if $\sum_{i=1}^{N} m_i > N - 1 + m_i$. Under these circumstances, a myriad of estimation methods, such as two-stage least squares (2SLS), three-stage least squares (3SLS), maximum likelihood methods, and simultaneous generalised method of moments (GMM), can be used to uncover the structural parameters Λ and $\tilde{\beta}$. Moreover, standard tests can be developed to test the restrictions on $\tilde{\lambda}$, $\tilde{W}_N$, and $\tilde{\beta}$, if those restrictions are of concern to the researchers.

This paper proposes to estimate Λ and $\tilde{\beta}$ in two stages as it is computationally simple. To estimate the parameters associated with the i^{th} individual dependent variable, in the first stage, we estimate

$$Y_{/i} = X\Upsilon_i + E_i, \tag{4}$$

where $Y_{/i}$ is the $T \times (N-1)$ matrix of dependent variables except for the i^{th} dependent variable, and E_i is a $T \times (N-1)$ matrix of error terms whose precision matrix might not be diagonal.

Making use of the estimated Υ_i, in the second stage, we estimate

$$y_i = \hat{Y}_{/i}(\Lambda_{i\bullet})' + X_{\bullet,} i\tilde{\beta}_i + u_i, \tag{5}$$

where y_i is the $T \times 1$ vector of the i^{th} dependent variable, $\hat{Y}_{/i} = X\Upsilon_i$, $\Lambda_{i\bullet}$ is the $1 \times (n-1)$ vector of the i^{th} row of Λ with Λ_{ii} dropped, and $\tilde{\beta}_i$ is the corresponding coefficients in $\tilde{\beta}$.

In Equation (5), the t^{th} row of $\hat{Y}/i$ contains the predicted spatial values of y_t from Equation (1), excluding its i^{th} element y_{it}. Since the exogenous variables are used as instrumental variables to derive the predicted values, Equation (5) is free from the endogeneity problem caused by reverse causality in Equation (1), where y_{it} may depend on y_{jt} for $j \neq i$ and vice versa. As proved by Zellner and Theil (1962), while the variance-covariance matrix of u_t is diagonal, the two-stage least squares method is asymptotically as efficient as the three-stage least squares method in solving simultaneous equations modelled in Equations (1) and (3).

3. TWO-STAGE VB

The number of unknown parameters involved in estimating model (3) is huge. Assuming the number of exogenous variables is m for each spatially dependent variable, we need to estimate Equations (4) and (5) N times to uncover a total of $N^2(N-1)m + (N^2(N-1)/2) + N^2 + Nm$ parameters. This can easily become computationally costly or even impossible for sampling methods such as MCMC.

As a more computationally efficient alternative to MCMC, VB has been increasingly used in sophisticated models involving large data where MCMC is too computationally expensive or untenable (e.g., Gefang et al., 2020, 2023; Loaiza-Maya et al., 2022).

Tran et al. (2021) provide a practical tutorial on VB methods. Instead of sampling from the posterior $p(\theta|y)$, VB approximates the posterior density by $q(\theta)$ belonging to some tractable distribution families. The best approximation q^* is found by minimising the Kullback–Leibler

divergence ($KL(q(\theta), p(\theta|y))$), which is equivalent to maximising the evidence lower bound (ELBO):

$$ELBO(q) = \int q(\theta)(\log p(\theta) + \log p(y|\theta) - \log q(\theta))d\theta$$

Tran et al. (2021) have categorised VB algorithms into two classes: mean field VB (MFVB) and fixed form VB (FFVB). Let $\theta = (\theta_1, \theta_2)$. MFVB assumes that q can be written as $q(\theta) = q_1(\theta_1)q_2(\theta_2)$, indicating that the optimisation problems can be solved by repeatedly updating $q_2(\theta_2)$ conditional on $q_1(\theta_1)$, and then updating $q_1(\theta_1)$ conditional on $q_2(\theta_2)$ until convergence. FFVB assumes a fixed parametric form for q, but usually involves more sophisticated stochastic optimisations with a lot of refinements. In this paper, we use MFVB to approximate the conditional posterior densities, following Ormerod and Wand (2010) and Blei et al. (2017).

In the two-stage VB, we identify the parameters in model (3) equation by equation. In each stage, we update the parameters using the approximate q densities through iterations. The convergence of the algorithm can be measured by checking if the changes in ELBO across iterations are less than a convergence criterion. However, when the number of parameters is large, calculating ELBO can be time-consuming. Therefore, it is more convenient to check for convergence by examining if the VB estimates of parameters stop changing across iterations.

3.1. First-stage VB

In the first stage, we estimate model (4) in order to construct the predicted value of $Y_{/i}$.

Let $\gamma = vec(Y_i)$. We set hierarchical D-L prior for the j^{th}, for $j = 1, \ldots, np$, element of γ as follows:

$$\gamma_j|\phi, \tau \sim DE(\phi_j\tau), \quad \phi_j \sim Dir(a, \ldots, a), \quad \tau \sim G(npa, 1/2). \tag{6}$$

where $DE(\bullet)$ denotes Double Exponential or Laplace distribution, $Dir(\bullet)$denotes Dirichlet distribution, $G(\bullet)$denotes Gamma distribution, $n = N - 1$, and $p = \sum_{i=1}^{N} m_i$.

The above hierarchical DL prior for γ_j can be expressed as:

$$\gamma_j \sim N(0, \psi_j\phi_j^2\tau^2), \quad \psi_j \sim Exp(1/2), \tag{7}$$

indicating the prior of γ is $N(\mathbf{0}, \underline{V})$ where $\underline{V} = diag(\psi_1\phi_1^2\tau^2, \ldots, \psi_{np}\phi_{np}^2\tau^2)$

Next, we set Exponential priors and D-L priors for the elements of Ω, the precision matrix of E_i, as follows:

$$\begin{aligned}
&\omega_{ii} \sim Exp(\underline{s}), \quad i = 1, \ldots, n \\
&\omega_{ij} \sim N(0, \psi_{\omega,ij}\phi_{\omega,ij}^2\tau_\omega^2), \quad \psi_{\omega,ij} \sim Exp(1/2), \quad i < j = 2, \ldots, n, \\
&\phi_{\omega,ij} \sim Dir(a_\omega, \ldots., a_\omega), \quad \tau_\omega \sim G\left(\frac{n^2 - n}{2}a_\omega, 1/2\right)
\end{aligned} \tag{8}$$

where, with a slight abuse of notations, we use ω_{ii} and ω_{ij} to denote the diagonal and off-diagonal elements of Ω.

The conditional posterior of γ is $N(\tilde{\gamma}, \tilde{V})$, where

$$\begin{aligned}
\tilde{V} &= (\underline{V}^{-1} + \Omega \otimes (X'X))^{-1}, \\
\tilde{\gamma} &= \tilde{V}(\Omega \otimes X')vec(Y).
\end{aligned} \tag{9}$$

The conditional posterior of τ is $giG(npa - np, 1, \sum_{j=1}^{np}(2|\gamma_j|/\phi_j))$.[3]

The conditional posterior of ϕ_j is derived as follows: First, we have $\xi_j \sim giG(a - 1, 1, 2|\gamma_j|)$. Next let $\Xi = \sum_{j=1}^{np}\xi_j$. The conditional posterior of ϕ_j can then be found to be ξ_j/Ξ.

The conditional posterior of $1/\psi_j$ is Inverse Gaussian with mean $\sqrt{\phi_j^2\tau^2/\gamma_j^2}$ and scale parameter 1.

Following Wang's (2012) Block Gibbs sampler to update the relevant parameters and hyperparameters, we use the last column and row of Ω as an example on how to update Ω.

Let $S = E_i'E_i$ and H be the $n \times n$ matrix with 0 diagonal elements and the off diagonal element at i^{th} row and j^{th} column be $\psi_{\omega,ij}\phi_{\omega,ij}^2\tau_\omega^2$. Partition Ω, S and H as follows:

$$\Omega = \begin{pmatrix} \Omega_{-n,-n} & \omega_{-n,n} \\ \omega'_{-n,n} & \omega_{nn} \end{pmatrix} \; S = \begin{pmatrix} S_{-n,-n} & s_{-n,n} \\ s'_{-n,n} & s_{nn} \end{pmatrix} \; H = \begin{pmatrix} H_{-n,-n} & h_{-n,n} \\ h'_{-n,n} & h_{nn} \end{pmatrix} \tag{10}$$

where $-n$ denotes the set of all indices except for n.

Let $b_1 = \omega_{n,n} - \omega'_{-n,n}\Omega_{-n,-n}^{-1}\omega_{-n,n}$ and $b_2 = \omega_{-n,n}$. The conditional posteriors are as follows:

$$\begin{aligned}
&b_1 \sim G\left(\frac{T}{2}+1,\ \frac{s_{nn}+\underline{s}}{2}\right) \\
&b_2 \sim N(-\ Cs_{-n,n},\ C),\quad C = ((\underline{s}+s_{nn})\Omega_{-n,-n}^{-1} + H*^{-1}) - 1,\quad H* = diag(h_{-n,n}) \\
&1/\psi_{w,ij} \sim iG\left(\sqrt{\frac{\tau_\omega^2\phi_{\omega,ij}^2}{\omega_{ij}^2}},\ 1\right) \\
&\tau_\omega \sim giG\left(\frac{n^2-n}{2}(a_3-1),\ 1,\ \sum_{i<j}\frac{2|\omega_{ij}|}{\phi_{w,ij}}\right) \\
&\xi_{w,ij} \sim giG(a_\omega - 1,\ 1,\ 2|\omega_{ij}|),\quad \phi_{\omega,ij} = \xi_{\omega,ij}/\sum_{i<j}\xi_{w,ij}.
\end{aligned} \tag{11}$$

Since the conditional posteriors are all of standard forms, we can then use them to derive the MFVB approximation densities and ELBO. To save space, we have relegated q densities and ELBO to Appendix A in the online supplemental data.

Algorithm 1 (First-stage VB estimation). Initialise the parameters, then update the following until convergence.

- Update $\bar{\gamma}$ and $\bar{V}$

$$\bar{V} = (V^{-1} + \bar{\Omega} \otimes (X'X))^{-1}$$

$$\bar{\gamma} = \bar{V}(\bar{\Omega} \otimes X')vec(Y_{/i})$$

- Update $\bar{\tau}$

$$\bar{\tau} = \frac{\sqrt{\chi}K_{npa-np+1}(\sqrt{\chi})}{K_{npa-np}(\sqrt{\chi})}$$

- Update ψ_j^{-1}

$$\overline{\psi_j^{-1}} = \sqrt{\frac{\overline{\phi_j^2}\,\overline{\tau^2}}{\bar{\gamma}_j^2 + \bar{V}_{jj}}}$$

- Update $\overline{\phi_j}$

$$\overline{\phi_j} = \frac{\overline{\xi_j}}{\sum_{j=1}^{np} \overline{\xi_j}}$$

- Update $\overline{b_1}$

$$\overline{b_1} = \frac{T/2}{\bar{s}_{n,n}}$$

- Update $\overline{b_2}$ and $\bar{C}$

$$\bar{C} = (2\bar{s}_{n,n}\Omega_{-n.-n}^{-1} + \bar{H}^{*-1})^{-1}$$

$$\overline{b_2} = (-\bar{C}\bar{s}_{-n,n})$$

- Update $\overline{\tau_\omega}$

$$\overline{\tau_\omega} = \frac{\sqrt{\chi_\omega} K_{\frac{n^2-n}{2}(a_\omega-1)+1}\left(\sqrt{\chi_\omega}\right)}{K_{\frac{n^2-n}{2}(a_\omega-1)}\left(\sqrt{\chi_\omega}\right)}$$

- Update $\overline{\psi_{\omega,jn}^{-1}} = \rho_\omega$

$$\overline{\psi_{\omega,jn}^{-1}} = \rho_\omega$$

- Update $\overline{\phi_{\omega,jn}}$

$$\overline{\phi_{\omega,jn}} = \frac{\overline{\xi_{\omega,jn}}}{\sum_{j<k} \overline{\xi_{\omega,jk}}},$$

- Update ELBO

3.2. Second-stage VB

We provide detailed technical explanations of the second-stage VB in Appendix B in the online supplemental data. Below, we briefly outline the priors and posteriors of both the parameters and hyperparameters.

$$\text{Let } Z = (\hat{Y}_{/i} \;\; X_{\bullet,i})' \text{ and } \theta = [(\Lambda_{i\bullet})' \;\; \tilde{\beta}_i].$$

We elicit hierarchical D-L prior for θ as follows:

$$\theta_j|\tilde{\phi}, \tilde{\tau} \sim DE(\tilde{\phi}_j\tilde{\tau}), \quad \tilde{\phi}_j \sim Dir(\tilde{a}, \ldots, \tilde{a}), \quad \tau \sim G(k\tilde{a}, 1/2) \tag{12}$$

where $k = N-1+, mi$.

This prior for θ_j can be expressed as:

$$\theta_j \sim N(0, \tilde{\psi}_j\tilde{\phi}_j^2\tilde{\tau}^2), \quad \tilde{\psi}_j \sim Exp(1/2) \tag{13}$$

Hence, the prior of θ is $N(\mathbf{0}, \tilde{V})$ where $\tilde{V} = diag(\tilde{\psi}_1\tilde{\phi}_1^2\tilde{\tau}^2, \ldots, \tilde{\psi}_k\tilde{\phi}_k^2\tilde{\tau}^2)$.

Next we set a Gamma prior for σ^{-2}:

$$\sigma^{-2} \sim G(\nu, \tilde{S}). \tag{14}$$

It can be found that the conditional posteriors are of θ is $N(\bar{\tilde{b}}, \bar{\tilde{V}})$, with $\bar{\tilde{V}} = [\sigma^{-2} Z'Z + \tilde{V}^{-1}]^{-1}$ and $\bar{\tilde{b}} = \sigma^{-2}\bar{\tilde{V}} Z' y_i$ and the conditional posterior of τ is $giG(k\tilde{a} - k, 1, \sum_{j=1}^{k}(2|\theta_j|/\tilde{\phi}_j))$.

To find the conditional posteriors of $\phi \sim j$, we integrate $\tilde{\tau}$ out following Bhattacharya et al. (2015). First, we have $\tilde{\xi}_j \sim giG(\tilde{a} - 1, 1, 2|\theta_j|)$. Next let $\tilde{\Xi} = \sum_{j=1}^{k} \tilde{\xi}_j$. The conditional posterior of $\tilde{\phi}_j$ can then be found to be $\tilde{\xi}_j / \tilde{\Xi}$.

The conditional posterior of $1/\tilde{\psi}_j$ is Inverse Gaussian with mean $\sqrt{\tilde{\phi}_j^2 \tilde{\tau}^2 / \theta_j^2}$ and scale parameter 1, while the conditional posterior of σ^{-2} is $G(T/2 + \nu, 1/2(y_i - Z\theta)'(y_i - Z\theta) + \tilde{S})$.

Making use of the q densities described in Appendix B in the online supplemental data, the second-stage VB estimation can be conducted as following:

Algorithm 2 (Second-stage VB estimation). Initialise the parameters, then update the following until convergence.

- Update $\bar{\tilde{V}}$ and $\bar{\theta}$

$$\bar{\tilde{V}} = \left(\frac{(T/2) + \nu}{\bar{\tilde{S}}} Z'Z + \tilde{V}^{-1} \right)^{-1}$$

$$\bar{\theta} = \left(\frac{(T/2) + \nu}{\bar{\tilde{S}}} \right) \bar{\tilde{V}} Z' y_i$$

- Update $\bar{S}$

$$\bar{S} = \frac{1}{2}[||y_i - Z\bar{\theta}||^2 + tr(Z'Z\bar{\tilde{V}})] + \tilde{S}$$

- Update $\bar{\tilde{\tau}}$

$$\bar{\tilde{\tau}} = \frac{\sqrt{\tilde{\chi}} K_{k\tilde{a}-k+1}(\sqrt{\tilde{\chi}})}{K_{k\tilde{a}-k}(\sqrt{\tilde{\chi}})}$$

- Update $\overline{\frac{1}{\tilde{\psi}_j}}$

$$\overline{\frac{1}{\tilde{\psi}_j}} = \sqrt{\frac{\bar{\tilde{\phi}}_j^2 \bar{\tilde{\tau}}^2}{\bar{\theta}_j^2 + \bar{\tilde{V}}^{jj}}}$$

- Update $\bar{\tilde{\phi}}_j$

$$\bar{\tilde{\phi}}_j = \frac{\bar{\tilde{\xi}}_j}{\sum^k \bar{\tilde{\xi}}_j}$$

- Update ELBO

4. MONTE CARLO STUDIES

In the Monte Carlo studies, we investigate a range of panel SAR models with different spatial autocorrelation coefficients, spatial weights matrices, and various sample sizes. Our focus is data

generation process 1 (DGP 1). In this scenario, for estimation simplicity, we assume that the dimension of the exogenous variables associated with each spatial unit is of dimension $T \times 1$. To examine the impact of the changes in the dimension of the exogenous variables, we conducted further simulation exercises using the same data generation process of DGP 1, but allowing the dimension of exogenous variables associated with each spatial unit to be of $T \times 5$ and $T \times 10$ (DGP 2).

4.1. DGP 1

The true model takes the following general form:

$$y_t = \lambda W_N y_t + 0.9 X_t + u_t, \tag{15}$$

where we generate X_{ti} and V_{ti} independently from $N(0, 1)$ and $0.1N(0, 1)$, respectively.

We make use of two types of true spatial weights matrices: one sparse and the other dense.

For the sparse W_{N_0}, each cross-sectional unit is connected to the two units ahead and behind it. The matrix W_{N_0} is then normalised by rows.

For the dense W_{N_0}, $2N/5$ elements ahead and $2N/5$ elements behind each diagonal element are randomly generated from $U(-0.5, 1.5)$. The matrix W_{N_0} is similarly normalised by rows.

We allow for negative values in the spatial weights matrix to capture phenomena such as more developed regions attracting labour from less developed regions, which can exert negative spatial effects on the latter's economic development.

The true value of λ is set to either $\lambda_0 = 0.4$ or $\lambda_0 = 0.6$.

The sample sizes we considered are $N = 30$ and $T = 20$, $N = 30$ and $T = 80$, $N = 50$ and $T = 30$, $N = 50$ and $T = 100$, $N = 100$ and $T = 50$, and $N = 100$ and $T = 200$.

For each case, we conduct 1000 Monte Carlo replications. To expedite the process, we monitor changes in parameters rather than in ELBO to assess convergence of the two-stage VB. To better understand the performance of VB with D-L priors, we compare the results with VB using horseshoe priors. Following Gefang et al. (2020), we set the standard $a = 1/np$ in the first stage and $\tilde{a} = 1/N - 1 + m_i$ for the D-L priors described in (6) and (12). For horseshoe priors, we used the standard Gamma priors as described in Makalic and Schmidt (2016).

Algorithm 3 (Monte Carlos). Given N, select the sparse or dense W_{N_0} according to model specifications, then repeat the following steps 1000 times.

- Independently generate X_{ti} and V_{ti} from $N(0, 1)$ and $0.1N(0, 1)$, respectively, for $i = 1, \ldots, N$ and $t = 1, \ldots, T + 1$.
- Generate y_t by computing $y_t = (I_n - \lambda_0 W_{N_0})^{-1} \left\{ 0.9 \begin{pmatrix} X_{t1} & 0 & \ldots & 0 \\ 0 & X_{t2} & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & X_{tN} \end{pmatrix} + \begin{pmatrix} V_{t1} \\ V_{t2} \\ \ldots \\ V_{tN} \end{pmatrix} \right\}$, for $t = 1, \ldots, T + 1$
- Using the first T observations of y_t and X_{ti} for estimation
  - First-stage VB: Using the $T \times N$ matrix X, which contains all the exogenous variables, to calculate the predicted values of $Y_{/i}$ for $i = 1, \ldots, N$.
  - Second-stage VB: Substitute the endogenous variables by their associated predicted values, then estimate β, λ and W_N.
- Compute the predictive log likelihood $\log P(y_{T+1} | X_{T+1}, \beta, \lambda, W_N)$.

Results of Monte Carlo simulations provide strong evidence that the two-stage VB is able to recover the true parameters in the data generating process, especially when $T \gg N$. When $T \ll N$, two-stage VB estimates tend to have larger biases depending on the priors. More

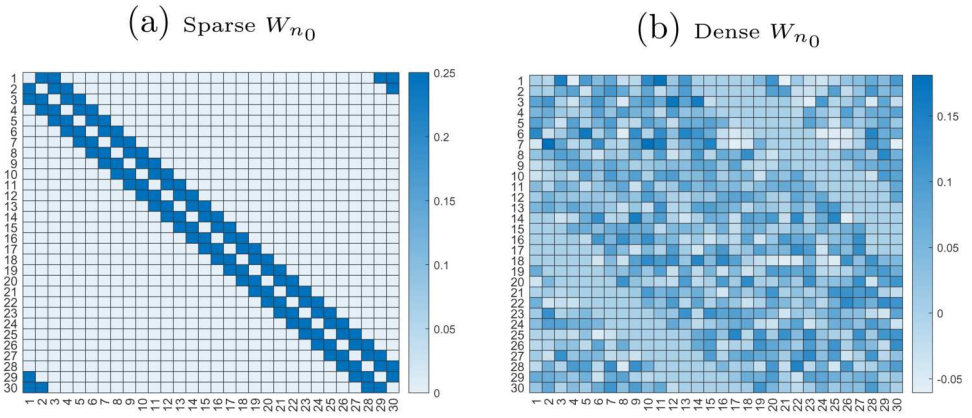

Figure 1. True spatial weights matrices.

important, there is clear evidence that the true spatial connections can be identified regardless of the length of the panel, long or short.

To give a flavour, we present the heatmaps of the spatial weights matrices for $N = 30$.[4] The true spatial weights matrices are displayed in Figure 1. We then plot the spatial weights matrices estimated using D-L priors in Figure 2, followed by those estimated using horseshoe priors in Figure 3. In Figures 2 and 3, the true values for λ_0 and T, and the characteristics of W_{n_0} used in the Monte Carlo setups are listed at the top of the corresponding subfigures.

An inspection of Figures 1–3 shows that both D-L and horseshoe priors are able to uncover the true sparse spatial matrix quite well under all circumstances. In comparison with horseshoe priors, D-L priors perform better when the true spatial weights matrix is dense. Holding λ_0 unchanged, the true dense spatial weights matrix can be better recovered when T is set to 80 instead of 20. Holding T unchanged, the dense spatial weights matrix is better estimated when λ_0 is larger. Same pattern can be also observed in Appendix C in the online supplemental data in heatmaps involving $N = 50$ and $N = 100$.

Table 1 summarises the Monte Carlo results. The structured similarity index measure (SSIM) index measures the structural similarity between the heatmap of the estimated spatial weights matrix and its true counterpart, with values closer to 1 indicating higher similarity.[5] Peak signal-to-noise ratio (PSNR) calculates the similarity between the heatmaps of the estimated and true spatial weights matrices, with higher values indicating greater similarity.[6] The Frobenius norm of the difference between the estimated and true spatial weight matrices is reported asF_{norm}. A smaller F_{norm} indicates greater similarity between the matrices.[7] To facilitate comparison between different estimated spatial matrices, we also compute the ratio of the Frobenius norm of the difference between the estimated matrix and its true counterpart to the Frobenius norm of the true spatial matrix, denoted as RF_{norm}. A smaller RF_{norm} thus implies a better estimation. Finally, we report the log predictive likelihoods ($\log \hat{p}$) and then the energy scores (ES) proposed in Gneiting et al. (2008).

The SSIM, PSNR and Frobenius norm related values reported in Table 1 are consistent with our visual assessment from Figures 1–3 and those in Appendix C in the online supplemental data. However, the log predictive likelihoods and energy scores reveal that when $T \gg N$, there is little difference between the D-L and horseshoe priors. In contrast, when $N \gg T$, horseshoe priors consistently outperform D-L priors, likely due to the fact that horseshoe priors tend to produce less biased estimates of β.

When $N \gg T$, the estimated λ is very similar to its true value when D-L priors are used for VB. However, the estimated βs are notably downwards biased. Conversely, when horseshoe priors are used, the situation is reversed. This phenomenon is expected because horseshoe priors

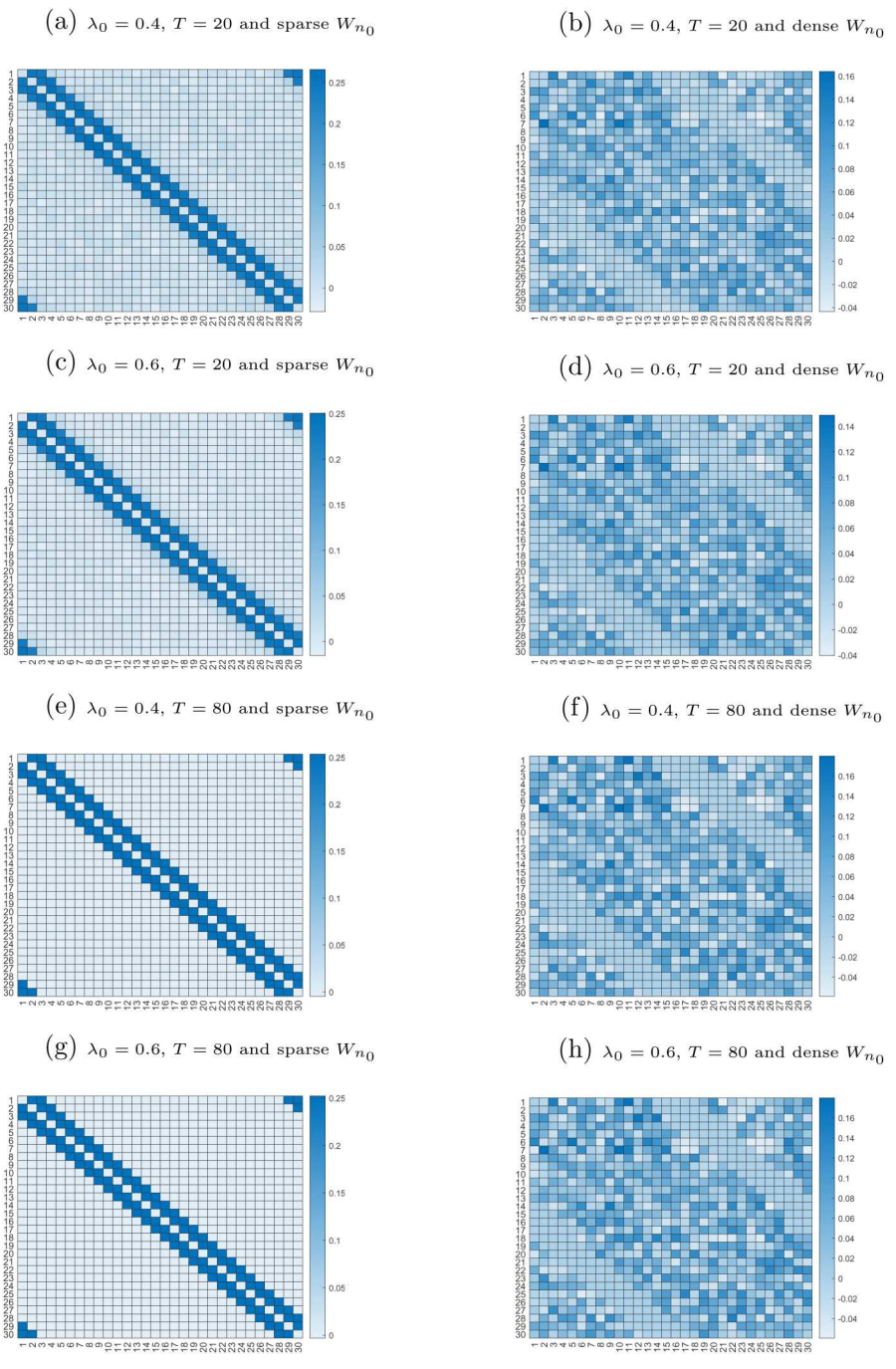

Figure 2. Estimated spatial weights matrices with D-L priors.

tend to impose smaller or no shrinkage on larger coefficients, while shrinking smaller coefficients more aggressively compared to D-L priors.

When $N \ll T$, D-L priors exhibit excellent performance in uncovering the true λ and β in all cases. In contrast, horseshoe priors, despite performing well in uncovering parameters when W_N is sparse, tend to produce biased λ estimates.

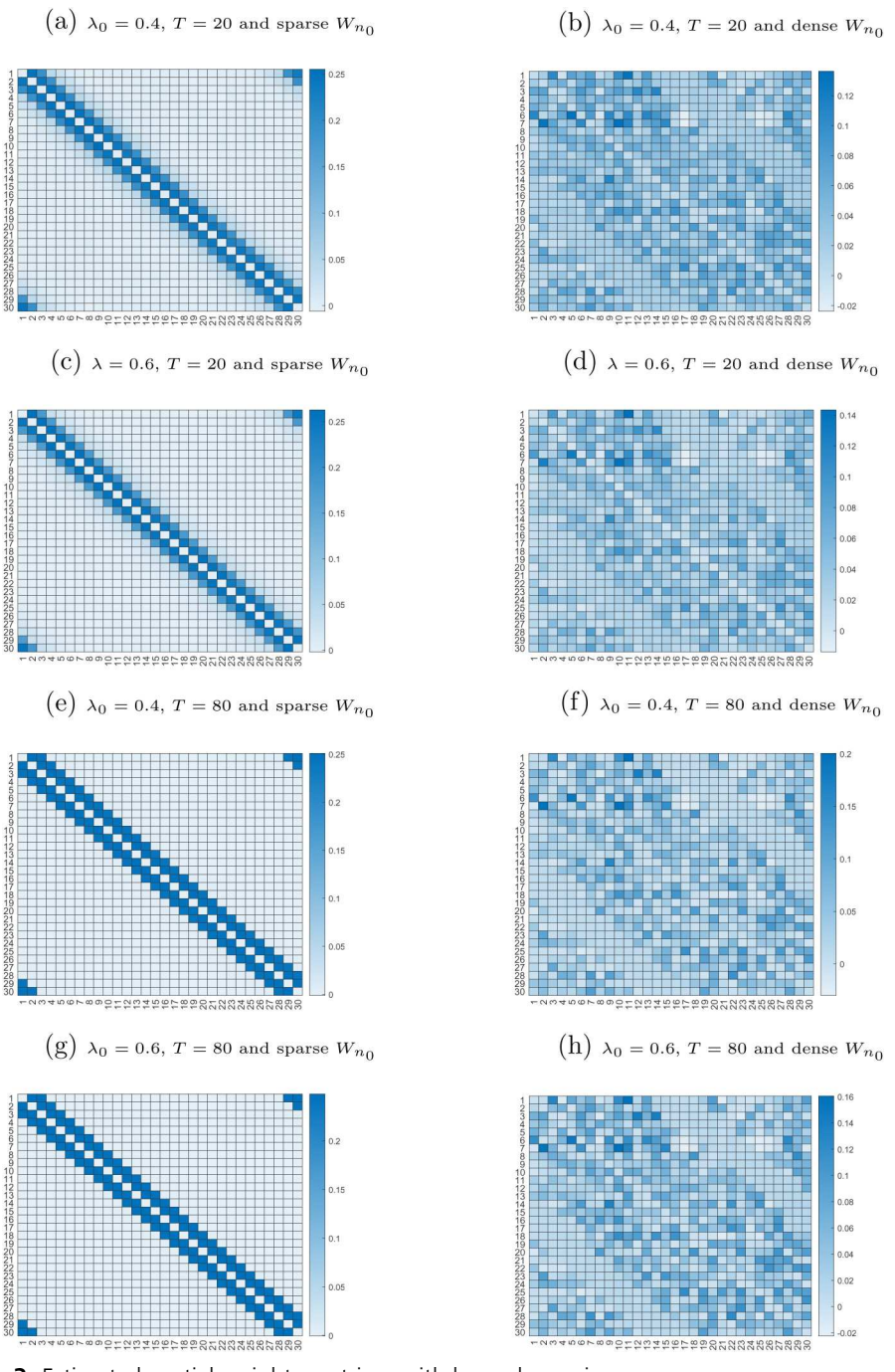

Figure 3. Estimated spatial weights matrices with horseshoe priors.

We use high-performance computing (HPC) services for estimation. All computations are performed using one compute node with four processor cores. On average, each Monte Carlo replication for a model with $N = 30$ and $T = 80$ takes approximately 20 s, while for the same model with $N = 30$ and $T = 20$, it requires about 60 s.

Similar patterns are observed for models with $N = 50$ and $N = 100$: two-stage VB requires significantly more time to converge when $N \gg T$. To provide a sense of the convergence speed of two-stage VB, each Monte Carlo replication for a model with $N = 100$ and $T = 50$ takes approximately 20 min, whereas this reduces to 5 min for the same model with $N = 100$ and $T = 200$.

The convergence speeds for models with the same N and T are relatively similar, regardless of the priors used and whether the spatial weights matrix is sparse or dense.

4.2. DGP 2

The only difference between DGP 2 and DGP 1 is that, in DGP 2, we allow the dimension of X_{ti}, the exogenous variables associated with each spatial unit to be 1×5 and 1×10 respectively, with each element of the exogenous variables again drawn from a standard normal distribution. Since the simulation results for DGP 1 indicate that D-L priors are effective in uncovering the true spatial weight matrices, particularly when they are dense, for brevity, we focus on evaluating the performance of the D-L prior on simulated data where $N = 30$ and $\lambda_0 = 0.6$.

The heatmaps of the estimated spatial weight matrices are presented in Figure 4. Upon visual inspection, it is hard to distinguish the differences between subfigures (a), (c), (e) and (g) plotted in Figure 4, and indeed, any differences between them and subfigures (c) and (d) of Figure 2. Likewise, subfigures (b), (d), (f) and (h) of Figure 4 look similar, and they resemble the subfigures (d) and (h) of Figure 2. A closer study of Table 2, however, reveals that when $N \gg T$, increasing the dimension of the exogenous variables tends to result in less accurate estimates of the spatial weight matrices, and more importantly, cause the coefficients of the exogenous variables to be noticeably downward biased. In contrast, when $T \gg N$, changes in the dimension of the exogenous variables have little impact.

4.3. VB estimates versus MCMC posterior draws

To evaluate how VB point estimates compare with MCMC posterior draws, we consider an illustrative example based on data generated from DGP 1, with $N = 15$, $T = 40$, $\lambda_0 = 0.6$, $\beta_0 = 0.9$ and a sparse spatial weight matrix.

Using the same priors for VB estimation as outlined in Section 4.1, the MCMC algorithm was executed for 7000 iterations, with the initial 2000 iterations discarded as burn-in.

Figures 5 and 6 show histograms of the MCMC posterior draws (in light blue) alongside the corresponding point VB estimates (in red) and true parameters (in dark blue), under the D-L and horseshoe priors, respectively. The estimate for β and λ are presented first, followed by four non-zero elements from the first two rows of the spatial weight matrix.

As anticipated, in most cases, the VB estimates closely align with the MCMC posterior means or modes. Additional plots for the non-zero elements of the spatial weight matrix, provided in Appendix D in the online supplemental data, further corroborate these findings.

5. EMPIRICAL APPLICATIONS

The relationship between sovereign bond ratings and spreads in eurozone countries is of great interest to researchers and policymakers alike, as discussed in studies such as Gibson et al. (2021) and Hall et al. (2023). A key inquiry is whether changes in ratings and spreads, respectively, among southern euro area countries and their northern counterparts have different impacts on individual member countries, particularly during the European debt crisis, which persisted from late 2009 to 2015. To address this question, we employ the two-stage VB approach on a dataset comprising five southern euro area countries – Spain, Greece, Ireland, Italy and Portugal – and five northern euro area countries – Austria, Belgium, Finland, France and the Netherlands.

We use monthly data of Gibson et al. (2021). Sovereign rating is denoted by Rate, which is the combined sovereign ratings given by Standard and Poor's, Fitch and Moody's then

Table 1. Summaries of simulation results (GDP 1).

Priors	**D-L**		**Horseshoe**		**D-L**		**Horseshoe**		**D-L**		**Horseshoe**	
W_{n_0}	**Sparse**	**Dense**	**Sparse**	**Dense**	**Sparse**	**Dense**	**Sparse**	**Dense**	**Sparse**	**Dense**	**Sparse**	**Dense**
	N = 30, T = 20				**N = 50, T = 30**				**N = 100, T = 50**			
$\lambda_0 = 0.4$												
SSIM	0.9378	0.9113	0.9556	0.7942	0.9547	0.9049	0.9856	0.8317	0.9690	0.9343	0.9989	0.9011
PSNR	40.7064	34.9154	33.9569	32.2450	41.9435	37.7057	36.8405	36.4465	42.9151	41.7627	46.7688	41.5224
F_{norm}	0.2766	0.5387	0.6016	0.7326	0.3998	0.6512	0.7194	0.7527	0.7149	0.8163	0.4587	0.8392
RF_{norm}	0.1010	0.3043	0.2197	0.4138	0.1131	0.3801	0.2035	0.4394	0.1430	0.4711	0.0917	0.4843
log$p$	−0.1535	−0.1630	0.2146	0.4239	−0.3504	−0.3537	0.3325	0.5413	−0.5959	−0.5843	0.6874	0.6629
ES	0.1738	0.1661	0.1223	0.0973	0.2112	0.1971	0.1080	0.0795	0.2719	0.2481	0.0772	0.0727
λ	0.3976	0.4072	0.1338	0.0942	0.3835	0.3954	0.1739	0.0616	0.3607	0.3760	0.2908	0.0354
	(0.0886)	(0.0854)	(0.0282)	(0.0285)	(0.0829)	(0.0783)	(0.0198)	(0.0176)	(0.0686)	(0.0649)	(0.0085)	(0.0087)
β	0.6393	0.6359	0.9101	0.8893	0.5760	0.5729	0.9171	0.8966	0.4812	0.4797	0.9092	0.8998
	(0.0681)	(0.0675)	(0.0509)	(0.0392)	(0.0528)	(0.0521)	(0.0369)	(0.0271)	(0.0386)	(0.0376)	(0.0212)	(0.0181)
$\lambda_0 = 0.6$												
SSIM	0.9790	0.8916	0.9537	0.7400	0.9825	0.8919	0.9922	0.8036	0.9854	0.9326	0.9999	0.8623
PSNR	39.7683	34.3537	33.4767	31.1463	40.4761	37.4611	39.6654	35.7122	40.9419	41.9692	59.6920	39.9803
F_{norm}	0.3081	0.5747	0.6358	0.8314	0.4733	0.6698	0.5196	0.8191	0.8972	0.7972	0.1036	1.0023
RF_{norm}	0.1125	0.3246	0.2321	0.4696	0.1339	0.3910	0.1470	0.4782	0.1794	0.4600	0.0207	0.5784
log$p$	−0.1507	−0.1638	0.2868	0.1793	−0.3535	−0.3547	0.5926	0.2575	−0.6112	−0.5847	0.8449	0.4693
ES	0.1987	0.1742	0.1440	0.1460	0.2447	0.2025	0.1038	0.1164	0.3209	0.2529	0.0762	0.0992
λ	0.6187	0.6312	0.4440	0.2883	0.6100	0.6251	0.5088	0.2016	0.5919	0.6146	0.5713	0.2048
	(0.0582)	(0.0534)	(0.0660)	(0.0762)	(0.0566)	(0.0497)	(0.0472)	(0.0470)	(0.0516)	(0.0442)	(0.0038)	(0.0439)
β	0.6417	0.6355	0.9346	0.9016	0.5781	0.5724	0.9245	0.9050	0.4811	0.4796	0.9033	0.9025
	(0.0694)	(0.0686)	(0.0733)	(0.0649)	(0.0541)	(0.0525)	(0.0453)	(0.0420)	(0.0399)	(0.0378)	(0.0170)	(0.0226)

	N = 30, T = 80				N = 50, T = 100				N = 100, T = 200			
$\lambda_0 = 0.4$												
SSIM	0.9989	0.9996	0.9990	0.9243	0.9986	0.9992	0.9998	0.9269	0.9991	0.9992	0.9999	0.9634
PSNR	57.8630	57.8767	51.7717	36.6654	58.0320	58.1413	59.2160	40.1701	60.8426	60.9244	74.2362	45.8173
F_{norm}	0.0384	0.0383	0.0774	0.4404	0.0627	0.0619	0.0547	0.4903	0.0908	0.0899	0.0194	0.5118
RF_{norm}	0.0140	0.0216	0.0282	0.2488	0.0177	0.0362	0.0155	0.2862	0.0182	0.0519	0.0039	0.2954
log$p$	0.8840	0.8840	0.8690	0.7580	0.8812	0.8809	0.8739	0.7364	0.8801	0.8799	0.8797	0.7992
ES	0.0626	0.0584	0.0637	0.0694	0.0619	0.0574	0.0623	0.0669	0.0617	0.0570	0.0616	0.0635
λ	0.4003	0.4003	0.3783	0.2946	0.4030	0.3998	0.3863	0.2262	0.4031	0.4029	0.3976	0.2237
	(0.0097)	(0.0096)	(0.0069)	(0.0191)	(0.0094)	(0.0097)	(0.0044)	(0.0202)	(0.0067)	(0.0067)	(0.0019)	(0.0143)
β	0.9000	0.9001	0.9030	0.9028	0.8996	0.9000	0.9019	0.9013	0.8996	0.8999	0.9001	0.9010
	(0.0147)	(0.0145)	(0.0124)	(0.0141)	(0.0145)	(0.0144)	(0.0108)	(0.0123)	(0.0103)	(0.0101)	(0.0074)	(0.0081)
$\lambda_0 = 0.6$												
SSIM	0.9995	0.9998	0.9988	0.8987	0.9994	0.9996	0.9996	0.9180	0.9996	0.9996	0.9999	0.9654
PSNR	61.2033	61.2621	53.5347	35.5434	61.3757	61.6279	60.1760	39.6870	63.7999	64.2264	79.7412	45.9792
F_{norm}	0.0261	0.0259	0.0632	0.5011	0.0427	0.0415	0.0490	0.5183	0.0646	0.0615	0.0103	0.5024
RF_{norm}	0.0095	0.0147	0.0231	0.2831	0.0121	0.0242	0.0139	0.3026	0.0129	0.0355	0.0021	0.2899
log$p$	0.8840	0.8840	0.8805	0.7158	0.8809	0.8809	0.8800	0.8045	0.8804	0.8799	0.8806	0.8064
ES	0.0737	0.0623	0.0739	0.1330	0.0725	0.0597	0.0725	0.0635	0.0718	0.0581	0.0717	0.0664
λ	0.6004	0.6005	0.6012	0.7220	0.6000	0.6003	0.6029	0.5913	0.6033	0.6029	0.6017	0.5105
	(0.0065)	(0.0065)	(0.0053)	(0.1840)	(0.0065)	(0.0065)	(0.0040)	(0.1368)	(0.0045)	(0.0045)	(0.0016)	(0.0101)
β	0.9000	0.9001	0.9016	0.9098	0.9000	0.9000	0.9007	0.9047	0.8990	0.8999	0.8994	0.9011
	(0.0151)	(0.0147)	(0.0126)	(0.0176)	(0.0150)	(0.0145)	(0.0110)	(0.0149)	(0.0105)	(0.0101)	(0.0075)	(0.0085)

Note: Standard deviations are presented in parentheses.

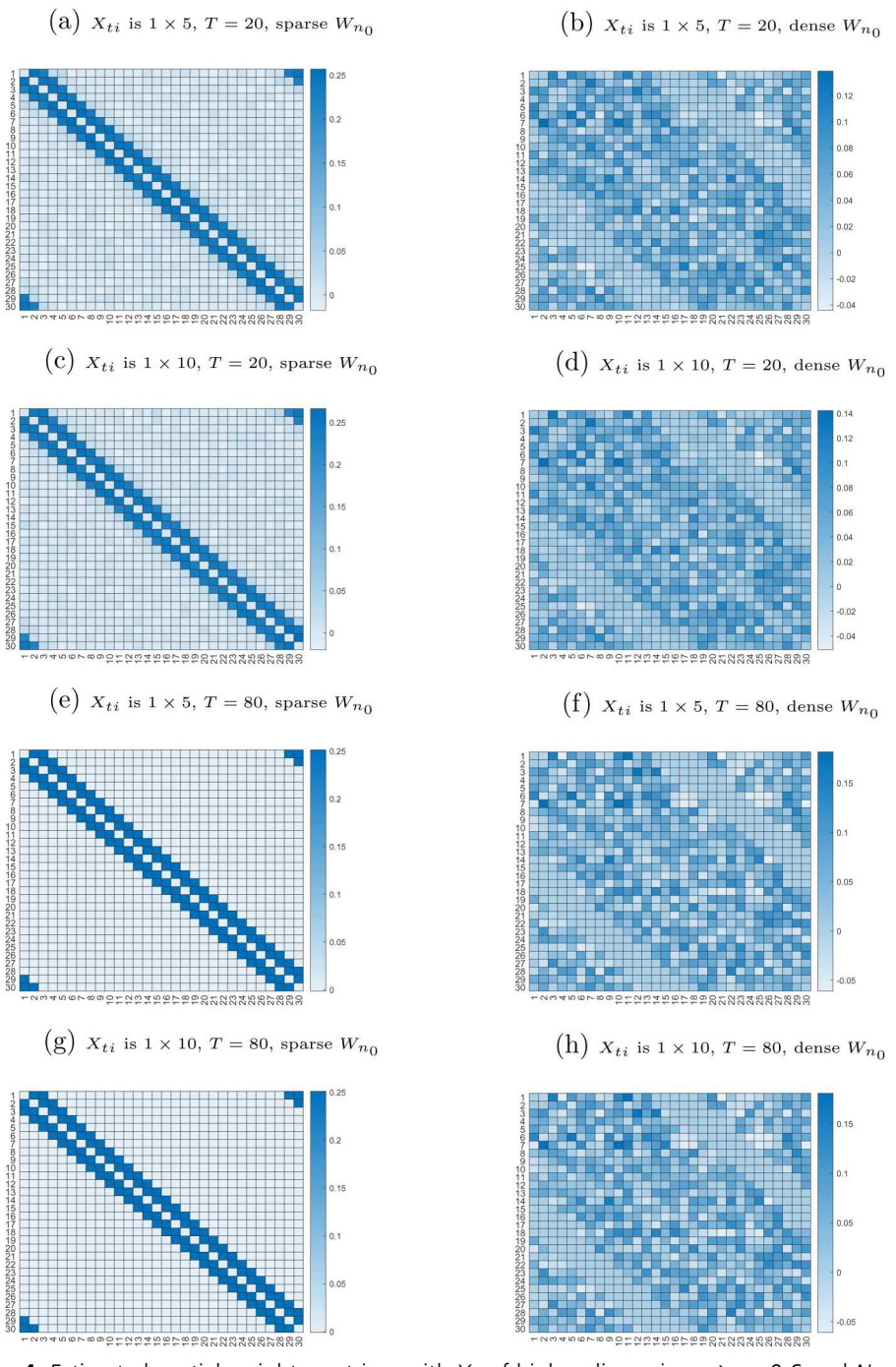

Figure 4. Estimated spatial weights matrices with X_{ti} of higher dimensions, $\lambda_0 = 0.6$ and $N = 30$.

transformed into numerical scale with 'triple A' having value 1 and 'selected default' having the value 22. Spread is defined as the difference between the yield on 10-year government bond of a country and that of Germany. Hence a rise in the rate or spread implies a worsening situation. Following Gibson et al. (2017), we use the following variables that capture the economic and political fundamentals as control variables: GDPgrowth, which is the real GDP growth rate;

Table 2. Summaries of simulation results (GDP 2).

Priors	$X_{ti}$is of 1×5		$X_{ti}$is of 1×10	
W_{n_0}	Sparse	Dense	Sparse	Dense
		N = 30, T = 20		
ssim	0.9632	0.8244	0.9658	0.7888
psnr	39.7788	31.9357	40.3890	31.1356
F_{norm}	0.3077	0.7592	0.2869	0.8324
RF_{norm}	0.1124	0.4288	0.1047	0.4702
log p	−1.5562	−1.5565	−2.1017	−2.1041
ES	0.8014	0.7042	1.3512	1.1878
λ	0.6615	0.6725	0.6796	0.6925
	(0.0822)	(0.0753)	(0.0935)	(0.0842)
β_1	0.3839	0.3892	0.2638	0.2722
	(0.1736)	(0.1714)	(0.2112)	(0.2100)
β_2	0.3842	0.3885	0.2635	0.2691
	(0.1783)	(0.1774)	(0.2042)	(0.2025)
β_3	0.3853	0.3921	0.2643	0.2713
	(0.1811)	(0.1777)	(0.2064)	(0.2074)
β_4	0.3825	0.3884	0.2823	0.2901
	(0.1810)	(0.1774)	(0.2015)	(0.2024)
β_5	0.3851	0.3882	0.2532	0.2595
	(0.1752)	(0.1737)	(0.2053)	(0.2043)
β_6			0.2658	0.2719
			(0.2116)	(0.2073)
β_7			0.2662	0.2730
			(0.2084)	(0.2085)
β_8			0.2595	0.2658
			(0.2067)	(0.2025)
β_9			0.2632	0.2751
			(0.2107)	(0.2082)
β_{10}			0.2659	0.2704
			(0.1981)	(0.1967)
		N = 30, T = 80		
ssim	0.9999	1.0000	0.9999	1.0000
psnr	67.6057	67.4311	70.5797	70.5724
F_{norm}	0.0125	0.0128	0.0089	0.0089
RF_{norm}	0.0046	0.0072	0.0032	0.0050
log p	0.8829	0.8834	0.8774	0.8775
ES	0.0701	0.0619	0.0725	0.0629

(*Continued*)

Table 2. Continued.

Priors	$X_{ti}$is of 1×5		$X_{ti}$is of 1×10	
W_{n_0}	Sparse	Dense	Sparse	Dense
λ	0.6027	0.6022	0.6013	0.6010
	(0.0030)	(0.0030)	(0.0022)	(0.0022)
β_1	0.9003	0.9004	0.9001	0.8993
	(0.0154)	(0.0153)	(0.0165)	(0.0153)
β_2	0.9000	0.9002	0.8998	0.9002
	(0.0153)	(0.0150)	(0.0158)	(0.0158)
β_3	0.9000	0.9002	0.9006	0.9004
	(0.0145)	(0.0144)	(0.0156)	(0.0160)
β_4	0.8999	0.8999	0.9002	0.8995
	(0.0152)	(0.0147)	(0.0149)	(0.0161)
β_5	0.8998	0.8990	0.9007	0.8996
	(0.0144)	(0.0147)	(0.0168)	(0.0155)
β_6			0.9001	0.8999
			(0.0153)	(0.0163)
β_7			0.8996	0.8993
			(0.0161)	(0.0157)
β_8			0.9005	0.8995
			(0.0155)	(0.0157)
β_9			0.8996	0.8997
			(0.0159)	(0.0165)
β_{10}			0.8992	0.8997
			(0.0156)	(0.0160)

Note: Standard deviations are presented in parentheses.

News, which is fiscal news constructed using European Commission forecasts;$Debt/GDP$, which is the ratio of government debt to GDP;CA/GDP, which is the ratio of current account balance to GDP; P/P^*, which is the ratio of a country's harmonised consumer price to that of Germany; and Pol, which is the index of political uncertainty reflecting the climate for foreign investors and political uncertainty. The monthly data runs from January 2000 to April 2019. For brevity, we report the data sources in Appendix E in the online supplemental data.

Taking account of the feedback loop between sovereign bond ratings and sovereign spreads, our dynamic model takes the following form:

$$\begin{aligned} Rate_{it} &= c_{11,i} + c_{12,i} * Spread_{it} + c_{13,i} * GDPgrowth_{it} + c_{14,i} * \frac{Debt_{it}}{GDP_{it}} + c_{15,i} * News + \\ &\quad c_{16,i} * \frac{P}{P^*it} + c_{17,i} * Rate_{i(t-1)} + W_i^{rate} * Rate_t + \varepsilon_{it}^{rate} \\ Spread_{it} &= c_{21,i} + c_{22,i} * Rate_{it} + c_{23,i} * GDPgrowth_{it} + c_{24,i} * \frac{Debt_{it}}{GDP_{it}} + c_{25,i} * Pol_{it} + \\ &\quad c_{26,i} * \frac{CA_{it}}{GDP_{it}} + c_{27,i} * Spread_{i(t-1)} + W_i^{sp} * Spread_t + \varepsilon_{it}^{sp} \end{aligned} \tag{16}$$

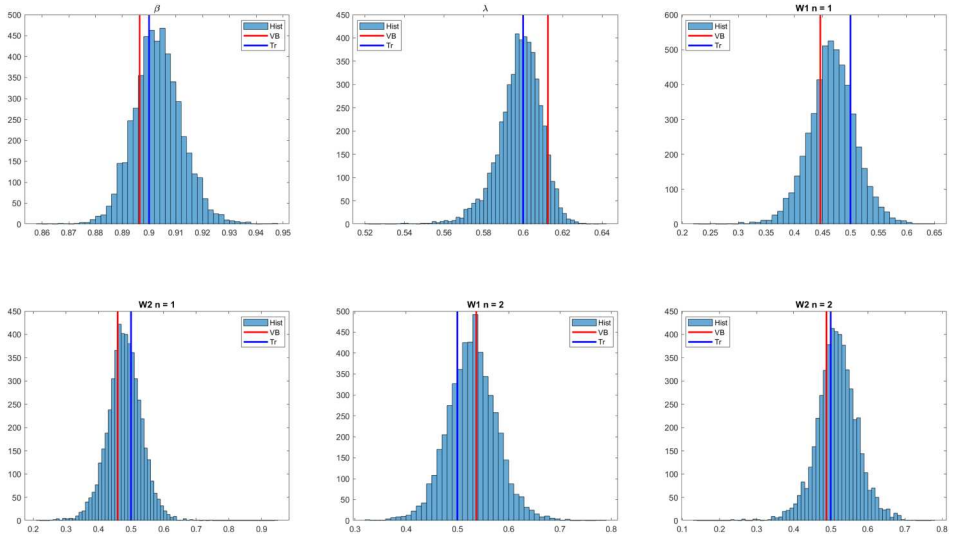

Figure 5. Plots of MCMC posterior draws against VB estimates and true parameter values, with D-L priors.

where W_i^{rate} and W_i^{spread} are the i^{th} row of the 10×10 spatial weights matrices W^{rate} and W^{spread}, respectively. Note that W^{rate} and W^{spread} are both with zero diagonals and the rest of the elements to be estimated.

Model (16) can be rewritten as

$$\begin{bmatrix} Rate_t \\ Spread_t \end{bmatrix} = \begin{bmatrix} I_{10} - W^{rate} & -c_{12} \\ -c_{22} & I_{10} - W^{spread} \end{bmatrix}^{-1} \left(\begin{bmatrix} c_{17} & 0 \\ 0 & c_{27} \end{bmatrix} \begin{bmatrix} Rate_{t-1} \\ Spread_{t-1} \end{bmatrix} + Ax + \begin{bmatrix} \varepsilon_t^{rate} \\ \varepsilon_t^{spread} \end{bmatrix} \right) \tag{17}$$

where $Rate_t = (Rate_{1t}, \ldots, Rate_{10t})'$, $Spread_t = (Spread_{1t}, \ldots, Spread_{10t})'$, $\varepsilon_t^{rate} = (\varepsilon_{1t}^{rate}, \ldots, \varepsilon_{10t}^{rate})'$, $\varepsilon_t^{spread} = (\varepsilon_{1t}^{spread}, \ldots, \varepsilon_{10t}^{spread})'$, $\mathbf{c}_{12} = diag(\mathbf{c}_{12,1}, \ldots, \mathbf{c}_{12,10})$, $\mathbf{c}_{22} = diag(\mathbf{c}_{22,1}, \ldots, \mathbf{c}_{22,10})$, $\mathbf{c}_{17} = diag(\mathbf{c}_{17,1}, \ldots, \mathbf{c}_{17,10})$, $\mathbf{c}_{27} = diag(\mathbf{c}_{27,1}, \ldots, \mathbf{c}_{22,10})$, $\mathbf{x}$ is the vector containing all the exogenous variables, and A is the matrix containing their respective parameters. In the spirit of Debarsy et al. (2012), we use Equation (17) to calculate the impulse response functions and evaluate how rate and spread changes in one country spill over to the other countries.

Due to the interconnected nature of economic variables, W_i^{rate} and W_i^{spread} tends to be dense rather than sparse. Hence we exclusively employ D-L priors in this empirical experiment, in contrast to Gibson et al. (2021) who used fixed, a priori, weights. Also, in contrast to Gibson et al. (2021), we show the time-varying behaviour of spillovers and spatial structures in Figures 7 and 8. The priors are set following the same approach as in the earlier Monte Carlo simulations.

We use a rolling window of 24 months to trace how the spillovers fluctuate over time. For the i^{th} country, the average rating spillovers from the south is computed by taking the mean of the cumulative impacts of a one notch value increase in ratings of southern euro area countries; and the average rating spillovers from the north is computed by the average of the cumulative impacts of a one notch value increase in northern euro area countries. In both cases, the impact responses to a shock of own country are excluded. In the same fashion, we calculate the average spread spillovers from the south and the north. We set the amount of shock to the spread to be one basis point.

Figure 7 presents the average spillovers from the southern euro area countries and those from the north, where spillovers are measured by the cumulative impacts over 5 years.[8] Two salient

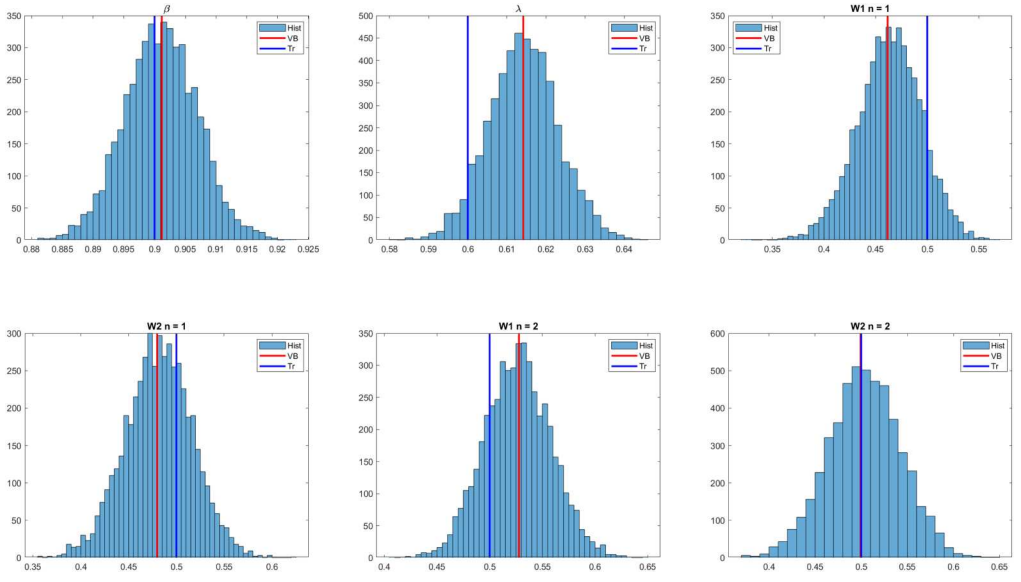

Figure 6. Plots of MCMC posterior draws against VB estimates and true parameter values, with horseshoe priors.

patterns emerge. First, spatial spillovers between countries only become apparent after the outbreak of the global financial crisis in 2007. Before that, the spillovers between countries are minimal. Second, shocks to a country's spread consistently increase other countries' spreads, but the impacts of shocks to sovereign ratings are mixed.

Let's examine the impacts of rating changes in a northern euro area country. We observe that shocks to rates of the southern countries tend to improve the sovereign ratings of France, Finland and the Netherlands (causing their rates to decrease), but worsen the ratings of Austria and Belgium (causing their rates to increase), especially during the period of the euro area crisis. Conversely, shocks to rates of other northern countries tend to slightly worsen the sovereign ratings of a northern country.

For a southern country, aside from Greece and Ireland, shocks to rates of northern countries tend to have minimal cumulative impacts on its ratings. However, for Greece and Ireland, especially during the euro area crisis, positive shocks to northern countries' ratings cause their own ratings to deteriorate. For all southern countries, shocks that worsen the ratings of other southern countries also significantly worsen their own ratings, far more than similar shocks to northern countries' ratings.

In all countries, an increase in the spreads of northern countries is consistently followed by an increase in their own spreads, increasing borrowing costs. Similarly, increases in the spreads of southern countries also lead to higher borrowing costs, but the impacts are much more pronounced than those from northern countries in all cases. During the euro area crisis, the spreads of France, Finland, and Spain increased significantly following shocks to southern countries' spreads.

Since model (16) is dynamic, the spatial spillovers measured by impulse responses are complicated by time dependence. To gain a deeper understanding of the spatial structure, we proceed by plotting the average spatial weights associated with each country in Figure 8. For the i^{th} country, the average rating spatial weights from the south is computed by taking the mean of the non-zero elements in W_i^{rate} that are associated with the southern euro area countries; and the average rating spatial weights from the north is computed by taking the mean of the

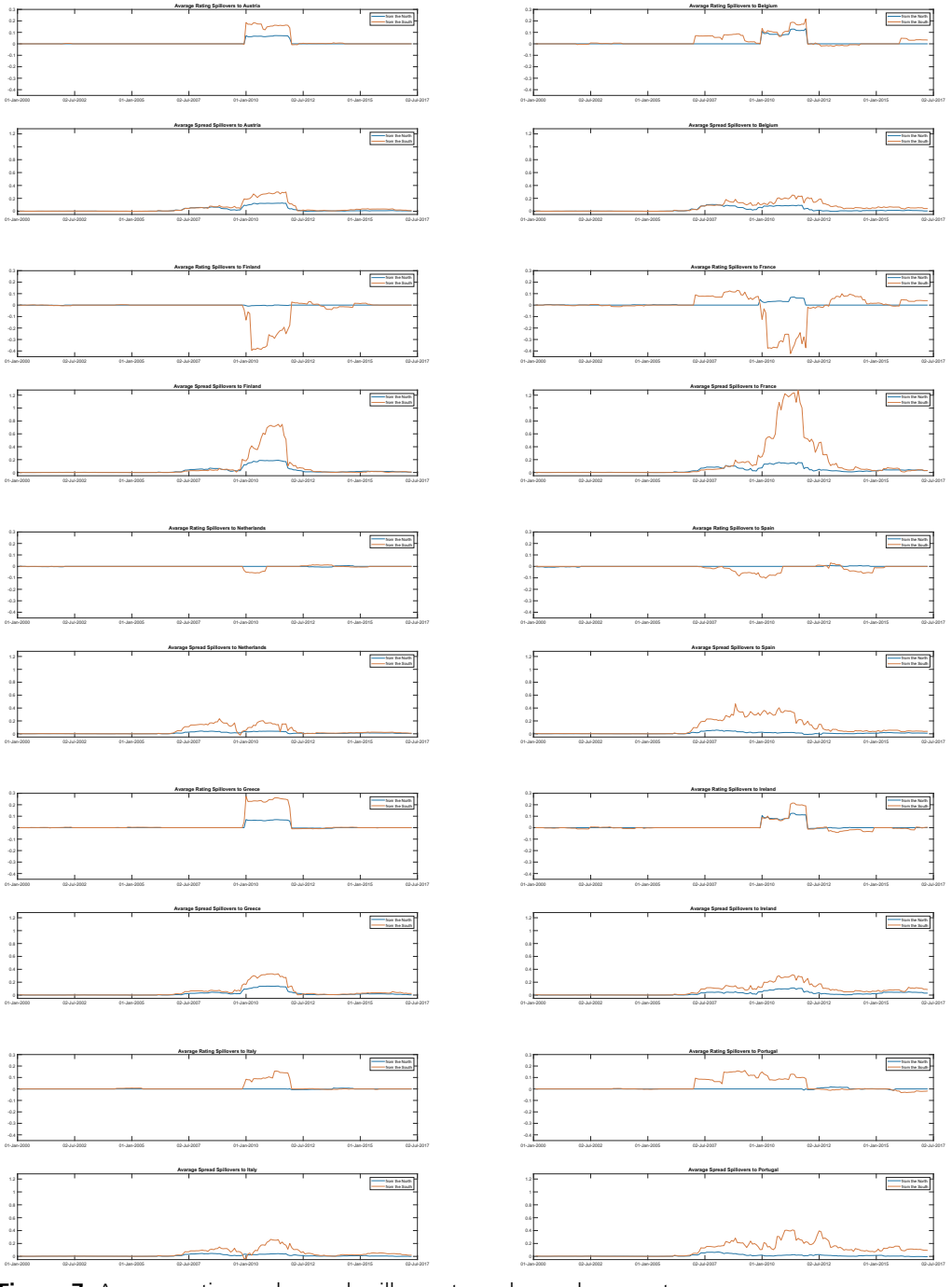

Figure 7. Average rating and spread spillovers to each member country.

non-zero elements in W_i^{rate} that are associated with the northern euro area countries. In the same way, we calculate the average spread spatial weights from the south and the north. The plots in Figure 8 provide further evidence for the marked differences in how a country is spatially linked with the north and the south. In general, northern countries are less spatially influenced by other

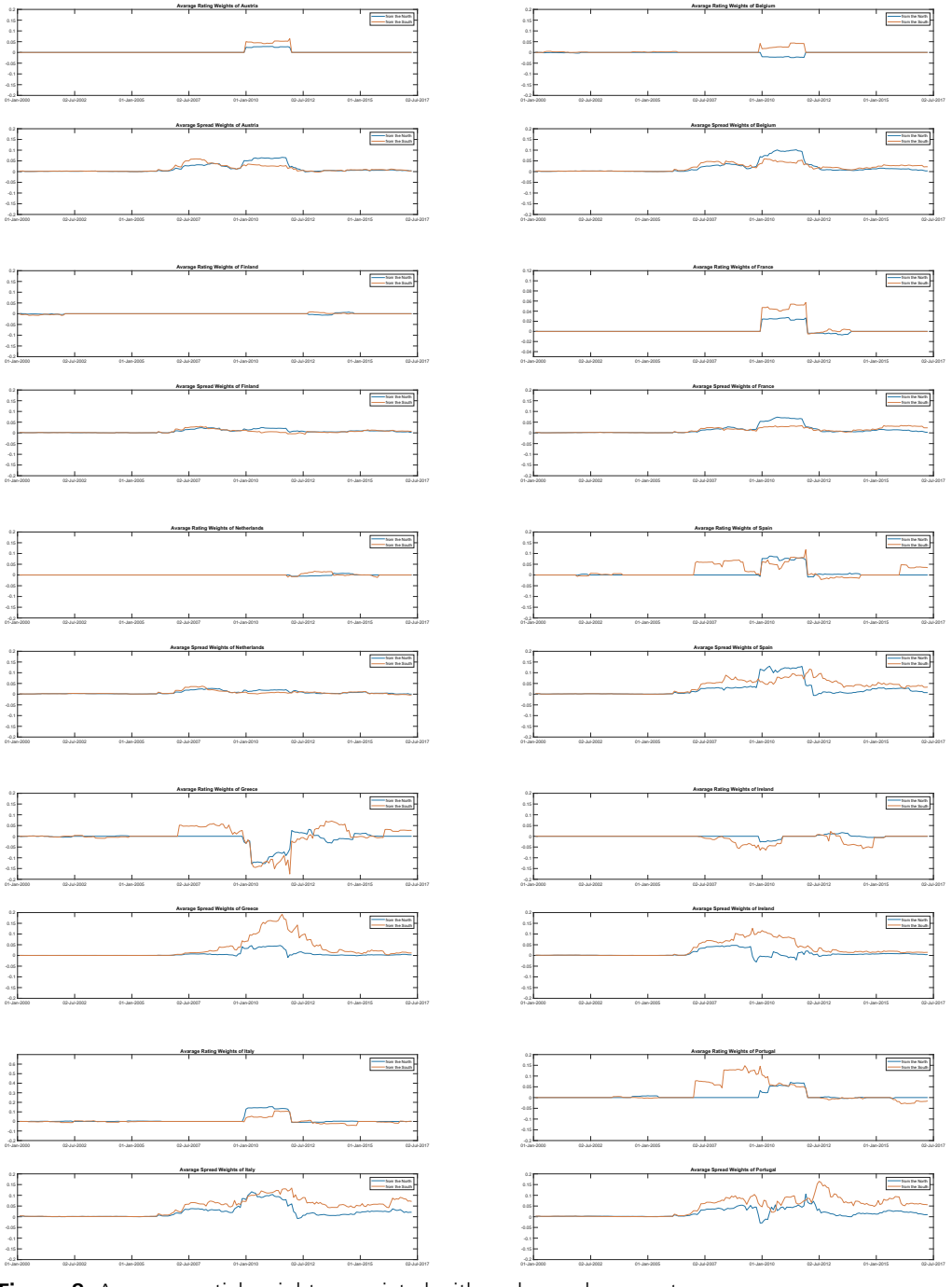

Figure 8. Average spatial weights associated with each member country.

countries than the southern countries, warranting us to take a closer look at how a southern country's rate and spread are spatially related to those of other countries.

In terms of the spatial relationships between one country's ratings and the ratings of other countries, Greece and Ireland stand out during the crises period, with other countries' ratings

spatially influencing these two countries' ratings negatively, highlighting the limitations of traditional spatial weight matrix where all the elements are non-negative. Apart from Greece and Ireland, a country's sovereign ratings tend to be positively associated with other countries' ratings. Interestingly, ratings of Spain and Portugal are more closely associated with those of other southern countries than those of the northern countries from 2007 to 2010, but become more influenced by northern countries' ratings in the euro area crisis after 2010. Italy's sovereign ratings, by contrast, are more spatially influenced by ratings of the northern countries than those of the south throughout time.

For spreads, negative weights are rare, indicating an increase in one country's spread tends to be associated with spread increases in other countries too. Apart from Spain, whose spreads are more spatially influenced by the northern countries for a brief period during the euro area crisis, spreads of all southern countries are more closely spatially linked with other southern countries than with the northern ones.

These results stand in sharp contrast to Gibson et al. (2021). In that paper, Europe's north and the south spatial weighting matrix had fixed equal weights which were symmetric. For example, the effect of Germany on the Netherlands was the same as the effect of the Netherlands on Germany. This is true of many fixed weight schemes, such as those using distance. These results not only suggest that there is time variation in the spillover effects but also that they are far from symmetric. Such a finding would not have been possible using the fixed weight scheme of Gibson et al. (2021).

6. CONCLUSIONS

In applied work, if the spatial weight matrix set a priori were to be far from its true value, the empirical analysis using SAR models would be misleading. In this paper, we have developed a two-stage VB approach to estimating panel SAR models with unknown spatial weights matrices so as to let the data speak.

Our two-stage VB with D-L priors method can be easily extended by using other popular priors such as the Lasso, horseshoe and spike and a slab priors. Furthermore, the success of two-stage VB shows the potential of combining VB with more sophisticated methods, such as three-stage least squares, full information likelihood and GMM to estimate panel SAR models, especially those involving $N \gg T$. Monte Carlo experiments show that our two-stage VB is rather fast, and it can recover the spatial impacts well for both the long and short panels.

As an empirical example, we apply the two-stage VB to the sovereign bond ratings and sovereign spreads data of 10 eurozone countries to uncover the impacts of spatial spillovers of the southern countries, which were more severely hit by the eurozone debt crisis, and those of the northern countries. Without pre-imposing any spatial weights matrices that might be unrealistic, we are able to shed new lights on the spillover behaviours of the south and the north. To our best knowledge, our findings are among the first in the literature to delineate how an individual country is affected by the spillovers from other countries in the eurozone.

ACKNOWLEDGEMENT

We would like to thank the co-editor and the anonymous referee for their constructive comments and suggestions. We also thank Maria Monopoli for her excellent research assistance. All remaining errors are the authors' own. This paper is a revised version of the working paper 'Fast Two-Stage Variational Bayesian Approach to Estimating Panel Spatial Autoregressive Models with Unrestricted Spatial Weights Matrices', which is available here: https://ideas.repec.org/p/arx/papers/2205.15420.html.

DATA AVAILABILITY STATEMENT

Data sources are provided in Appendix E in the online supplemental data. The Matlab codes can be found at: https://github.com/DBayesian/GHT2025_HS and https://github.com/DBayesian/GHT2025_DL

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

NOTES

[1] Global-local shrinkage priors such as the horseshoe (Carvalho et al., 2009) and the R2-D2 prior (Zhang et al., 2022) are known for improving on the D-L prior in terms of the mass at the origin and tail thickness. For comparison, we extended our VB algorithms to include the horseshoe prior.

[2] This research used the ALICE High Performance Computing Facility at the University of Leicester.

[3] $y \sim giG(p, a, b)$ if $f(y) \propto y^{p-1}exp[-1/2(ay + b/y)]$.

[4] Heatmaps related to $N = 50$ and $N = 100$ are plotted in Appendix C in the online supplemental data.

[5] SSIM between two images x and y of size $N \times N$ is calculated as follows:

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

where μ_x and μ_y are the means of x and y, respectively; σ_x^2 and σ_y^2 are the respective variances of x and y; σ_{xy} is the covariance; and c_1 and c_2 are constants to avoid instability when any of the following four terms $2\mu_x\mu_y$, $2\sigma_{xy}$, $\mu_x^2 + \mu_y^2$ and $\sigma_x^2 + \sigma_y^2$ are close to zeros. We refer readers to Zhou et al. (2004) for more technical details.

[6] PSNR between two images A and B of dimensions$N \times M$ is calculated as follows:

$$PSNR = 10log_{10}\frac{R^2}{(\sum_{i=1}^{N}\sum_{j=1}^{M}(A_{ij} - B_{ij})^2)/MN}$$

where the denominator is the mean square error and the numerator R is the maximum fluctuation in the input image data type. For example, if the input image has a double-precision floating-point data type, then R is 1. If it has an 8-bit unsigned integer data type, R is 255. More details about image types can be found in Matlab's User's Guide for Image Processing Toolbox™ (R2024b) (Matlab Help Center, 2024).

[7] The Frobenius norm of an $N \times M$ matrix D is derived as follows:

$$F_{norm} = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{M}|D_{ij}|^2}$$

[8] The impulse response functions typically converge to zero after 2 to 3 months.

ORCID

Deborah Gefang http://orcid.org/0000-0002-5117-4903
Stephen G. Hall http://orcid.org/0000-0001-6068-8749
George S. Tavlas http://orcid.org/0000-0002-9046-1956

REFERENCES

Ahrens, A., & Bhattacharjee, A. (2015). Two-step lasso estimation of the spatial weights matrix. *Econometrics*, *3* (1), 128–155. https://doi.org/10.3390/econometrics3010128

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Kluwer Academic.

Baltagi, B. H., Egger, P., & Pfaffermayr, M. (2013). A generalized spatial panel data model with random effects. *Econometric Reviews*, *32*(5–6), 650–685. https://doi.org/10.1080/07474938.2012.742342

Baltagi, B. H., Song, S. H., & Koh, W. (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics*, *117*(1), 123–150. https://doi.org/10.1016/S0304-4076(03)00120-9

Basak, G. A., Bhattacharjee, A., & Das, S. (2018). Causal ordering and inference on acyclic networks. *Empirical Economics*, *55*(1), 213–232. https://doi.org/10.1007/s00181-018-1454-3

Bhattacharya, A., Pati, D., Pillai, N. S., & Dunson, D. B. (2015). Dirichlet–Laplace priors for optimal shrinkage. *Journal of the American Statistical Association*, *110*(512), 1479–1490. https://doi.org/10.1080/01621459.2014.960967

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877. https://doi.org/10.1080/01621459.2017.1285773

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2009). Handling sparsity via the horseshoe. *Proceedings of the 12th International Conference Artifical Intelligence and Statistics, PMLR*, *5*, 73–80. https://proceedings.mlr.press/v5/carvalho09a/carvalho09a.pdf.

Case, A. (1991). Spatial patterns in household demand. *Econometrica 59*(4), 953–965. https://doi.org/10.2307/2938168

Cliff, A. D., & Ord, J. K. (1973). *Spatial autocorrelation*. Pion.

Debarsy, N., Ertur, C., & LeSage, J. P. (2012). Interpreting dynamic space–time panel data models. *Statistical Methodology*, *9*(1–2), 158–171. https://doi.org/10.1016/j.stamet.2011.02.002

Fox, J. (1979). Simultaneous equation models and two-stage least squares. *Sociological methodology*, *10*, 130–150. https://doi.org/10.2307/270769

Gefang, D., Koop, G., & Poon, A. (2020). Computationally efficient inference in large Bayesian mixed frequency VARs. *Economics Letters*, *191*, 1–6. https://doi.org/10.1016/j.econlet.2020.109120

Gefang, D., Koop, G., & Poon, A. (2023). Forecasting using variational Bayesian inference in large vector autoregressions with hierarchical shrinkage. *International Journal of Forecasting*, *39*(1), 346–363. https://doi.org/10.1016/j.ijforecast.2021.11.012

Gibson, H. D., Hall, S. G., Gefang, D., Petroulas, P., & Tavlas, G. S. (2021). Cross-country spillovers of national financial markets and the effectiveness of ECB policies during the euro-area crisis. *Oxford Economic Papers*, *73* (4), 1454–1470. https://doi.org/10.1093/oep/gpab046

Gibson, H. D., Hall, S. G., & Tavlas, G. S. (2017). Self-fulfilling dynamics: The interactions of sovereign spreads, sovereign ratings and bank ratings during the euro financial crisis. *Journal of International Money and Finance*, *73*(B), 371–385. https://doi.org/10.1016/j.jimonfin.2017.03.006

Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L., & Johnson, N. A. (2008). Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test*, *17*(2), 211–235. https://doi.org/10.1007/s11749-008-0114-x

Hall, S. G., Gefang, D., & Tavlas, G. S. (2023). A test to select between spatial weighting matrices. *Journal of Spatial Econometrics*, *4*(1), 1. https://doi.org/10.1007/s43071-022-00032-9

Krisztin, T., & Piribauer, P. (2023). A Bayesian approach for the estimation of weight matrices in spatial autoregressive models. *Spatial Economic Analysis*, *18*(1), 44–63. https://doi.org/10.1080/17421772.2022.2095426

Krock, M., Kleiber, W., & Becker, S. (2021). Nonstationary modeling with sparsity for spatial data via the basis graphical lasso. *Journal of Computational and Graphical Statistics*, *30*(2), 375–389. https://doi.org/10.1080/10618600.2020.1811103

Lam, C., & Souza, P. C. (2020). Estimation and selection of spatial weight matrix in a spatial lag model. *Journal of Business & Economic Statistics*, *38*(3), 693–710. https://doi.org/10.1080/07350015.2019.1569526

Lee, L.-F., & Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, *154*(2), 165–185. https://doi.org/10.1016/j.jeconom.2009.08.001

Li, Y., Craig, B. A., & Bhadra, A. (2019). The graphical horseshoe estimator for inverse covariance matrices. *Journal of Computational and Graphical Statistics*, *28*(3), 747–757. https://doi.org/10.1080/10618600.2019.1575744

Liu, X., & Saraiva, P. (2019). GMM estimation of spatial autoregressive models in a system of simultaneous equations with heteroskedasticity. *Econometric Reviews*, *38*(4), 359–385. https://doi.org/10.1080/07474938.2017.1308087

Loaiza-Maya, R., Smith, M. S., Nott, D. J., & Danaher, P. J. (2022). Fast and accurate variational inference for models with many latent variables. *Journal of Econometrics*, *230*(2), 339–362. https://doi.org/10.1016/j.jeconom.2021.05.002

Makalic, E., & Schmidt, D. F. (2016). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, *23*, 179–182.

Matlab Help Center. (2024). MATLAB User's Guide for Image Processing Toolbox™ (R2024b). The MathWorks, Inc. https://uk.mathworks.com/help/images/.

Ormerod, J. T., & Wand, M. P. (2010). Explaining variational approximations. *The American Statistician*, *64*(2), 140–153. https://doi.org/10.1198/tast.2010.09058

Piribauer, P., Glocker, C., & Krisztin, T. (2023). Beyond distance: The spatial relationships of european regional economic growth. *Journal of Economic Dynamics and Control*, *155*, 1–19. https://doi.org/10.1016/j.jedc.2023.104735

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tran, M. N., Nguyen, T. N., & Dao, V. H. (2021). A practical tutorial on variational Bayes. arXiv preprint arXiv:2103.01327.

Wang, H. (2012). Bayesian graphical lasso models and efficient posterior computation. *Bayesian Analysis*, *7*(4), 867–886. https://doi.org/10.1214/12-BA729

Yang, K., & Lee, L.-F. (2017). Identification and QML estimation of multivariate and simultaneous equations spatial autoregressive models. *Journal of Econometrics*, *196*(1), 196–214. https://doi.org/10.1016/j.jeconom.2016.04.019

Zellner, A., & Theil, H. (1962). Three-stage least squares: Simultaneous estimation of simultaneous equations. *Econometrica*, *30*(1), 54–78. https://doi.org/10.2307/1911287

Zhang, Y. D., Naughton, B. P., Bondell, H. D., & Reich, B. J. (2022). Bayesian regression using a prior on the model fit: The R2-D2 shrinkage prior. *Journal of the American Statistical Association*, *117*(538), 862–874. https://doi.org/10.1080/01621459.2020.1825449

Zhou, W., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612.