Fast Two-Stage Variational Bayesian Approach to Estimating Panel Spatial Autoregressive Models with Unrestricted Spatial Weights Matrices

Deborah Gefang[1], Stephen G. Hall[2], and George S. Tavlas[3]

[1]University of Leicester
[2]University of Leicester, Bank of Greece and University of Pretoria
[3]Bank of Greece and Hoover Institution, Stanford University

Abstract

This paper proposes a fast two-stage variational Bayesian (VB) algorithm to estimate unrestricted panel spatial autoregressive models. Using Dirichlet-Laplace priors, we are able to 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. We find marked differences between the spillover behaviours of the northern euro area countries and those of the south.

1 Introduction

The spatial autoregressive (SAR) models, first proposed by Cliff and Ord (1973), have been widely used in the literature to investigate the spatial dependence in cross-sectional units (e.g., Anselin, 1988; Baltagi et al., 2003, 2013, Lee and Yu, 2010). In practice, the spatial weights matrices are usually set a priori based on geographical distances or economic characters (e.g. Cliff and 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 least absolute shrinkage and selection operator (Lasso) of Tibshirani's (1996) and its variants have gained a lot of attention. For example, Basak et al. (2018) propose to estimate a 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 (2019) estimate the weights matrix using adaptive Lasso with sparse adjustment in mind. Most of the studies, however, usually impose sometime 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 deal with the impacts of any possible exogenous variables. Moreover, to our knowledge, few of those studies focus on the short panels where N is large while T is small. Only recently, Krisztin and Piribauer (2023) and Piribauer et al. (2023) use a hierarchical prior set-up 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 spatial weights matrix or the covariance functions, hence our approach lets the data speak. The prior we used for Bayesian regularization is the Dirichlet–Laplace (D-L) prior of Bhattacharya et al. (2015). With D-L prior, the entire posterior distribution concentrates at the optimal rate. This nice feature remains unchanged when the number of parameters to be estimated is much larger than the number of observations, providing strong theoretical justifications for the two-stage VB's effectiveness in uncovering the spatial dependencies in a short panel.

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 VB estimator for large covariance matrices with unknown sparsity patterns. Our VB methods using D-L prior can be easily extended to allow for other popular priors such as the graphic Lasso of Wang (2012), the half-Cauchy prior of Makalic and Schimidt (2016) and the graphical horseshoe prior of Li et al. (2019). This is not trivial as VB is a more computationally efficient alternative of Markov Chain Monte Carlo (MCMC), and our approach can be used in estimating other popular models involving large covariance matrices such as large Bayesian vector autoregressions (BVARs).[1]

We have conducted a wide range of simulation studies using a traditional panel SAR model and a panel SAR model that takes account of the simultaneous relationships between cross-sectional groups.[2] Monte Carlo experiments show that two-stage VB is accurate and computationally efficient when $T \gg N$, which usually is more pertinent to macroeconomic and financial data. When $N \gg T$, which tends to be more relevant to microeconomic data, two-stage VB estimates tend to have slightly larger biases and empirical standard

---

[1]Matlab code is provided.

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

deviations. Tighter priors can help lessen that problem.

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 euro-zone countries. The research is motivated by the importance of a more in-depth understanding in how cross-country contagion works in the southern euro area countries and their northern counterparts, especially in the European debt crises (e.g. Gibson et al., 2021, Hall et al., 2022). Using rolling windows of 2 years, we compare and contrast how a rate or spread change in the south and the north affect each individual member country. Our results provide ample evidence for the marked differences between the two country groups, both in how a shock to the south and a shock to the north impact a particular country in a very different way and in how the shocks to a particular country group affect a southern and a northern country very differently. Our results also highlight the big impacts of the global financial crises and how the south was severely affected by the European debt crises while the north was much less so.

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. Section 4 conducts Monte Carlo studies. Section 5 applies two-stage VB to the sovereign bond ratings and spreads data of ten euro area countries. Section 6 concludes. Sources for data used in the empirical example are listed in the Appendix. Detailed VB derivation formulas and more extensive Monte Carlo results are relegated to Online Supplements.[3]

---

[3]Online Supplements can be found at https://github.com/DBayesian/GKT2022

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 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:

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

where $y_t = (Y_{t1}, Y_{t2}, ..., Y_{tN})'$ 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, $X_t = \begin{pmatrix} X_{t,1} \\ X_{t,2} \\ ... \\ X_{t,N} \end{pmatrix}$ 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 $u_t = (V_{t1}, V_{t2}, ..., V_{tN})'$ is the $N \times 1$ i.i.d error terms with mean zero and diagonal covariance matrix $\begin{pmatrix} \sigma^2 & 0 & ... & 0 \\ 0 & \sigma^2 & ... & 0 \\ .. & .. & ... & .. \\ 0 & 0 & ... & \sigma^2 \end{pmatrix}$.

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:

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

where $\widetilde{W}_N$ 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, $\tilde{X}_t = \begin{pmatrix} X_{t,1} & 0 & ... & 0 \\ 0 & X_{t,2} & ... & 0 \\ .. & .. & ... & .. \\ 0 & 0 & ... & X_{t,N} \end{pmatrix}$,

$\tilde{\beta}$ 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 $X_{t,i}$ to 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).

Since our main concerns in panel SAR models are the spillover effects, there is therefore little research interest in separately identifying $\tilde{\lambda}$ and $\widetilde{W}_N$. What we care about is the product $(\tilde{\lambda} \otimes l_N) \odot \widetilde{W}_N$, where l_N is a $N \times 1$ column of ones and $\otimes$ is the Kronecker product, as $(\tilde{\lambda} \otimes l_N) \odot \widetilde{W}_N$ is the $N \times N$ parameter matrix which captures the spillover

effects between spatial units.

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

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

where $I_N - \Lambda$ is nonsingular 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 Thell (1962) and Fox (1979), the ith 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 likelihoods methods and simultaneous generalized method (GMM), can be used to uncover the structural parameters Λ and $\tilde{\beta}$. Moreover, standard tests can be developed to test the restrictions on $\tilde{\lambda}$, $\widetilde{W}_N$ and $\tilde{\beta}$, if those restrictions are of the researchers' concerns.

This paper proposes to estimate Λ and $\tilde{\beta}$ in two stages as it is computationally simple. To estimate the parameters associated with the ith 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 ith 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

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

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

Note that in a panel SAR model, we can have $N \gg T$ and $(N - 1 + m_i) \gg T$, which makes it difficult or even impossible to uncover B_i and $\Lambda_{i\bullet}$ using traditional estimation techniques.

3   Two-stage VB

As detailed in Ormerod and Wand (2010) and Blei et al. (2017), the essence of VB is to use appropriate densities from a mean field variational family to approximate the posterior densities through minimizing the Kullback-Leibler divergence, which is equivalent to maximising the evidence lower bound (ELBO). As a more efficient alternative to MCMC, VB has been increasingly used in sophisticated models involving large data where MCMC is too computationally expensive or even untenable (e.g. Gefang et al., 2020, 2022, Loaiza-Maya et al., 2022).

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 by iterations. The convergence of the algorithm can be measured by the changes in ELBO across iterations is less than a convergence criteria. When the number of parameters is large, however, calculating ELBO can be time consuming. It is therefore more convenient to check if convergence has occurred by examining if the VB estimates of parameters stop changing

8

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 $\gamma = vec(\Upsilon_i)$. We set hierarchical D-L prior for the jth, for $j = 1, ..., np$, element of γ as follows:

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

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

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

$$\omega_{ii} \sim Exp(\underline{s}), \quad i = 1, ..., n$$
$$\omega_{ij} \sim N(0, \psi_{\omega,ij} \phi_{\omega,ij}^2 \tau_\omega^2), \quad \psi_{\omega,ij} \sim Exp(1/2), \quad i < j = 2, ..., n, \tag{7}$$
$$\phi_{\omega,ij} \sim Dir(a_\omega, ..., a_\omega), \quad \tau_\omega \sim G(\frac{n^2 - n}{2} a_\omega, 1/2)$$

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

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

9

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:

$$\Omega = \begin{pmatrix} \Omega_{-n,-n} & \omega_{-n,n} \\ \omega'_{-n,n} & \omega_{nn} \end{pmatrix} \quad S = \begin{pmatrix} S_{-n,-n} & s_{-n,n} \\ s'_{-n,n} & s_{nn} \end{pmatrix} \quad H = \begin{pmatrix} H_{-n,-n} & h_{-n,n} \\ h'_{-n,n} & h_{nn} \end{pmatrix} \quad (8)$$

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

Relegating technical details to Online Appendix A, we outline the VB approximation densities as follows:

3.1.1 $q(\gamma)$

$$q(\gamma) \sim N(\bar{\gamma}, \overline{V}), \quad (9)$$

where

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

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

and

$$V^{-1} = diag(\frac{1}{\overline{\psi_1}\,\overline{\phi_1^2}\,\overline{\tau^2}}, ..., \frac{1}{\overline{\psi_{np}}\,\overline{\phi_{np}^2}\,\overline{\tau^2}})$$

3.1.2 $q(\tau)$

$$q(\tau) \sim giG[npa - np, 1, \sum_{j=1}^{np} 2(\bar{\gamma}_j^2 + \overline{V}_{jj})^{1/2}\frac{1}{\overline{\phi_j}}], \quad (10)$$

10

Let $\chi = \sum_{j=1}^{np} 2(\bar{b}_j^2 + \overline{V}_{jj})^{1/2} \frac{1}{\phi_j}$, we have

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

$$\overline{\tau^2} = \bar{\tau}^2 + \chi \left[\frac{K_{npa-np+2}(\sqrt{\chi})}{K_{npa-np}(\sqrt{\chi})} - \left(\frac{K_{npa-np+1}(\sqrt{\chi})}{K_{npa-np}(\sqrt{\chi})} \right)^2 \right]$$

where $K_*[\bullet]$ is the modified Bessel functions of the second kind.

3.1.3 $q(\psi_j)$

$$q(\psi_j^{-1}) \sim iG\left(\sqrt{\frac{\overline{\phi_j^2} \, \overline{\tau^2}}{\overline{\gamma_j^2} + \overline{V}_{jj}}}, 1 \right), \tag{11}$$

where $iG(\bullet)$ denotes s Inverse Gaussian distribution.

Let $\rho = \sqrt{\frac{\overline{\phi_j^2} \, \overline{\tau^2}}{\overline{\gamma_j^2} + \overline{V}_{jj}}}$,

$$\overline{\psi_j^{-1}} = \rho$$

and

$$\overline{\psi}_j = 1 + 1/\rho$$

3.1.4 $q(\phi_j)$

$$q(\xi_j) \sim giG(a - 1, 1, 2\sqrt{\overline{\gamma_j^2} + \overline{V}_{jj}}), \tag{12}$$

where $giG(\bullet)$ denotes the generalized inverse Gaussian distribution.

Let $\varpi = 2\sqrt{\overline{\gamma_j^2} + \overline{V}_{jj}}$, we have

$$\overline{\xi_j} = \frac{\sqrt{\varpi}K_a(\sqrt{\varpi})}{K_{a-1}(\sqrt{\varpi})},$$

and

$$var(\xi_j) = \varpi\{\frac{K_{a+1}(\sqrt{\varpi})}{K_{a-1}(\sqrt{\varpi})} - [\frac{K_a(\sqrt{\varpi})}{K_{a-1}(\sqrt{\varpi})}]^2\},$$

where $var(\bullet)$ denotes the variance.

Scaling ξ_i, we have

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

and

$$\overline{\phi_j^2} = \overline{\phi_j}^2 + \frac{var(\xi_j)}{(\sum_{j=1}^{np}\overline{\xi_j})^2}$$

Thus, the optimal q density of $\phi_{i,j}$ takes the following form:

$$q(\phi_j) \sim giG(a-1, \sum_{j=1}^{np}\overline{\xi_j}, \frac{2\sqrt{\overline{\gamma_j^2} + (\overline{V}_{jj})^2}}{\sum_{j=1}^{np}\overline{\xi_j}}) \tag{13}$$

3.1.5 $q(b_1)$

Let $b_1 = \omega_{n,n} - \omega'_{-n,n}\Omega^{-1}_{-n,-n}\omega_{-n,n}$.

$$q(b_1) \sim G(\frac{T}{2}, \overline{s}_{n,n}), \tag{14}$$

where

$$\overline{s}_{n,n} = \frac{1}{2}(s_{n,n} + tr(X'X\overline{V}_n) + \underline{s}),$$

12

and

$$\overline{V_n} = V_{(n-1)\times p+1:n\times p,(n-1)\times p+1:n\times p}.$$

Hence

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

3.1.6 $q(b_2)$

Here we use b_2 to denote $\omega_{-n,n}$. Let $\overline{s}_{-n,n} = s_{-n,n} + \tilde{s}_{-n,n}$, where $\tilde{s}_{-n,n}$ is a $(n-1) \times 1$ vector with the jth element being $tr(X'XA_j)$ and $A_j = V_{(j-1)\times p+1:j\times p,(j-1)\times p+1:j\times p}$.

$$q(b_2) \sim N(-\overline{C}\overline{s}_{-n,n}, \overline{C}), \tag{15}$$

where

$$\overline{C} = (2\overline{s}_{n,n}\Omega^{-1}_{-n,-n} + \overline{H}^{*-1})^{-1},$$

and

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

Note that $\overline{H}^* = diag(\overline{h}_{-n,n})$, and j^{th} element of $\overline{h}_{-n,n}$ is $\overline{\psi_{\omega,jn}\phi^2_{\omega,jn}\tau^2_\omega}$.

3.1.7 $q(\tau_\omega)$

$$q(\tau_\omega) \sim giG[\frac{n^2-n}{2}(a_\omega - 1), 1, \sum_{j<k} 2(\overline{\omega_{jk}}^2 + \overline{C}_{jj})^{1/2}\frac{1}{\overline{\phi_{\omega,jk}}}], \,^4 \tag{16}$$

---

[4]Note that $\overline{C}_{jj}$ changes when k changes.

Let $\chi_\omega = \sum\limits_{j<k} 2(\overline{\omega_{jk}}^2 + \overline{V}_{jj})^{1/2}(\overline{\phi_{\omega,jk}})^{-1}$, we have

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

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

3.1.8 $q(\psi_{\omega,jn})$

$$q(\psi_{\omega,jn}^{-1}) \sim iG\left(\sqrt{\frac{\overline{\phi_{\omega,jn}^2}\,\overline{\tau_\omega^2}}{\omega_{jn}^2 + \overline{C}_{jj}}}, 1\right), \tag{17}$$

Let $\rho_\omega = \sqrt{\frac{\overline{\phi_{\omega,jn}^2}\,\overline{\tau_\omega^2}}{\overline{\omega_{jn}}^2+\overline{C}_{jj}}}$,

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

and

$$\overline{\psi_{w,jn}} = 1 + 1/\rho_\omega$$

3.1.9 $q(\phi_{\omega,jn})$

$$q(\xi_{\omega,jn}) \sim giG(a_\omega - 1, 1, 2\sqrt{\overline{\omega_{jn}}^2 + \overline{C}_{jj}}) \tag{18}$$

14

Let $\varpi_\omega = 2\sqrt{\overline{\omega_{jn}}^2 + \overline{C}_{jj}}$, we have

$$\overline{\xi_{\omega,jn}} = \frac{\sqrt{\varpi_\omega}K_{a_\omega}(\sqrt{\varpi_\omega})}{K_{a_\omega-1}(\sqrt{\varpi_\omega})},$$

and

$$var(\xi_{\omega,jn}) = \varpi_\omega\{\frac{K_{a_\omega+1}(\sqrt{\varpi_\omega})}{K_{a_\omega-1}(\sqrt{\varpi_\omega})} - [\frac{K_{a_\omega}(\sqrt{\varpi_\omega})}{K_{a_\omega-1}(\sqrt{\varpi_\omega})}]^2\},$$

where $var(\bullet)$ denotes the variance.

Scaling $\xi_{\omega,jn}$, we have

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

and

$$\overline{\phi_{\omega,jn}^2} = \overline{\phi_{\omega,jn}}^2 + \frac{var(\xi_{\omega,jn})}{(\sum_{j<k}\overline{\xi_{\omega,jk}})^2}$$

Thus, the optimal q density of $\phi_{\omega,ij}$ takes the following form:

$$q(\phi_{\omega,jn}) \sim giG(a - 1, \sum_{j<k}\overline{\xi_{\omega,jk}}, \frac{2\sqrt{\overline{\omega_{jn}}^2 + (\overline{C}_{jj})^2}}{\sum_{j<k}\overline{\xi_{\omega,jk}}}) \tag{19}$$

3.2 Second-stage VB

We explain the technical details of second-stage VB in Online Appendix B. Below we briefly describe the priors of the parameters and hyperparameters and then provide their optimal q densities.

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

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

$$\theta_j|\tilde{\phi}, \tilde{\tau} \sim DE(\tilde{\phi}_j\tilde{\tau}), \quad \tilde{\phi}_j \sim Dir(\tilde{a}, ..., \tilde{a}), \quad \tau \sim G(k\tilde{a}, 1/2) \tag{20}$$

where $k = N - 1 + m_i$.

Next we set a Gamma prior for σ^{-2}:

$$\sigma^{-2} \sim G(\nu, \tilde{S}). \tag{21}$$

The VB optimal densities can be found as follows:

3.2.1 $q(\theta)$

$$q(\theta) \sim N(\bar{\theta}, \overline{\tilde{V}}), \tag{22}$$

where

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

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

$$\tilde{V}^{-1} = diag(\overline{\tilde{\psi}_1^{-1}} \ \overline{\tilde{\phi}_1^{-2}\tilde{\tau}^{-2}}, ..., (\overline{\tilde{\psi}_k^{-1}} \ \overline{\tilde{\phi}_k^{-2}\tilde{\tau}^{-2}})$$

3.2.2 $q(\sigma^{-2})$

$$q(\sigma^{-2}) \sim G(\frac{T}{2} + \nu, \overline{\tilde{S}}), \tag{23}$$

16

where

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

3.2.3 $q(\tilde{\tau})$

$$q(\tilde{\tau}) \sim giG[k\tilde{a} - k, 1, \sum_{j=1}^{k} 2(\overline{\theta}_j^2 + \overline{\widetilde{V}}_{jj})^{1/2} \frac{1}{\overline{\tilde{\phi}}_j}], \tag{24}$$

Let $\tilde{\chi} = \sum_{j=1}^{k} 2(\overline{\theta}_j^2 + \overline{\widetilde{V}}_{jj})^{1/2} \frac{1}{\overline{\tilde{\phi}}_j}$, we have

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

and

$$\overline{\tilde{\tau}^2} = \overline{\tilde{\tau}}^2 + \chi[\frac{K_{k\tilde{a}-k+2}(\sqrt{\tilde{\chi}})}{K_{k\tilde{a}-k}(\sqrt{\tilde{\chi}})} - (\frac{K_{k\tilde{a}-k+1}(\sqrt{\tilde{\chi}})}{K_{k\tilde{a}-k}(\sqrt{\tilde{\chi}})})^2]$$

3.2.4 $q(\tilde{\psi}_j)$

$$q(\frac{1}{\tilde{\psi}_j}) \sim iG(\sqrt{\frac{\overline{\tilde{\phi}_j^2}\,\overline{\tilde{\tau}^2}}{\overline{\theta}_j^2 + \overline{\widetilde{V}}^{jj}}}, 1), \tag{25}$$

Let $\tilde{\rho} = \sqrt{\frac{\overline{\tilde{\phi}_j^2}\,\overline{\tilde{\tau}^2}}{\overline{\theta}_j^2 + \overline{\widetilde{V}}^{jj}}}$,

$$\overline{\frac{1}{\tilde{\psi}_j}} = \tilde{\rho}$$

and

$$\overline{\tilde{\psi}}_j = 1 + 1/\tilde{\rho}$$

17

3.2.5 $q(\tilde{\phi}_j)$

$$q(\tilde{\xi}_j) \sim giG(\tilde{a} - 1, 1, 2\sqrt{\overline{\theta}_j^2 + (\overline{\tilde{V}}_{jj})^2}) \tag{26}$$

Let $\tilde{\varpi} = 2\sqrt{\overline{\theta}_j^2 + (\overline{\tilde{V}}_{jj})^2}$, we have

$$\overline{\tilde{\xi}}_j = \frac{\sqrt{\tilde{\varpi}} K_{\tilde{a}}(\sqrt{\tilde{\varpi}})}{K_{\tilde{a}-1}(\sqrt{\tilde{\varpi}})},$$

and

$$var(\tilde{\xi}_j) = \tilde{\varpi}\{\frac{K_{\tilde{a}+1}(\sqrt{\tilde{\varpi}})}{K_{\tilde{a}-1}(\sqrt{\tilde{\varpi}})} - [\frac{K_{\tilde{a}}(\sqrt{\tilde{\varpi}})}{K_{\tilde{a}-1}(\sqrt{\tilde{\varpi}})}]^2\}.$$

Scaling $\tilde{\xi}$, we have

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

and

$$\overline{\tilde{\phi}_j^2} = \overline{\tilde{\phi}}_j^2 + \frac{var(\tilde{\xi}_j)}{(\sum^k \overline{\tilde{\xi}}_j)^2}$$

Thus, the optimal q density of $\tilde{\phi}_j$ takes the following form:

$$q(\tilde{\phi}_j) \sim giG[\tilde{a} - 1, \sum^k \overline{\tilde{\xi}}_j, (2\sqrt{\overline{\theta}_j^2 + (\overline{\tilde{V}}_{jj})^2})/(\sum^k \overline{\tilde{\xi}}_j)] \tag{27}$$

4   Monte Carlo Studies

In the Monte Carlo studies, we look into two traditional panel SAR models of various sample sizes. The first model is:

$$y_t = 0.6W_N y_t + 0.9x_t + u_t, \tag{28}$$

where to specify W_N, we let each cross-sectional unit be connected with the unit ahead of it and the unit behind, and then normalize W by rows. When conducting Monte Carlo, we generate x_t and u_t independently from $N(0,1)$ and $0.1N(0,1)$, respectively.

The second model is:

$$y_{t,1} = 0.5y_{t,2} + 0.6W_{N_1}y_{t,1} + 0.4W_{N_2}y_{t,2} + 0.9x_{t,1} + u_{t,1}$$
$$y_{t,2} = 0.5y_{t,1} + 0.4W_{N_1}y_{t,1} + 0.6W_{N_2}y_{t,2} + 0.9x_{t,2} + u_{t,2} \tag{29}$$

where $y_{t,1}$ is a vector of half of the N dependent variables observed at time t, and $y_{t,2}$ is the other half. When setting W_{N_1} and W_{N_2}, we assume a variable in $y_{t,1}$ is only spatially related with the unit ahead of it and the unit behind, likewise a variable in $y_{t,2}$. Both W_{N_1} and W_{N_2} are normalized by rows. In Monte Carlo, we generate each element of $x_{t,1}$ and $x_{t,2}$ independently from $N(0,1)$, then each element of $u_{t,1}$ and $u_{t,2}$ independently from $0.1N(0,1)$.[5]

For both (28) and (29), the sample sizes 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

---

[5]We have also experimented on W_N, W_{N_1} and W_{N_2} of other forms. In addition, we have looked into models with different coefficients, $u_{t,1} \sim 0.1N(0,1)$ and $u_{t,2} \sim 0.3N(0,1)$. The Monte Carlo results provide further evidence that two-stage VB method works well.

$N = 100$ and $T = 200$. For each case, we conduct 1000 Monte Carlo replications and use the changes in parameters instead of that of ELBO to check whether two-stage VB has converged.

Results of Monte Carlo simulations, which are relegated to Online Supplement to save space, 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 have larger biases and larger empirical standard deviations, which can be reduced by setting tighter priors. More 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, tables 1-2 report the Monte Carlo results of parameter Λ and $\tilde{\beta}$ for model (28) with $N = 30$ and $T = 80$ as well as $N = 30$ and $T = 20$. Since a tablet with 30 columns is too big to fit into a page, we only report the mean and standard deviations, the latter in parenthesis, of the empirical distributions of the 5×5 sub-matrices in the four corners of Λ, which are associated with $y_{t,1}, ..., y_{t,5}$ and $y_{t,26}, ..., y_{t,30}$, and the first and last five elements in $\tilde{\beta}$.

We use high-performance computing (HPC) services for estimation. All computations are done using 1 compute node and 4 processor core. On average, each Monte Carlo replication for model (28) where $N = 30$ and $T = 80$ takes about 20 seconds, while for the same model, when $N = 30$ and $T = 20$ it takes about 60 seconds. For model (29), it again takes 20 seconds when $N = 30$ and $T = 80$. However, estimating model (29) takes 3 minutes when $N = 30$ and $T = 20$. The same pattern can be observed when estimating models where $N = 50$ and $N = 100$, that is: two-stage VB takes much longer to converge when $N \gg T$, and the larger the number of nonzero true parameters, the longer the estimation takes. To

Table 1: Estimates of Λ for model (28), where $N = 30$

T=80					...					
	0.30	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.30
	(0.02)	(0.02)	(0.02)	(0.02)	...	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
0.30		0.30	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00
(0.02)		(0.02)	(0.02)	(0.02)	...	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
0.00	0.30		0.30	0.00	...	0.00	0.00	0.00	0.00	0.00
(0.02)	(0.02)		(0.02)	(0.02)	...	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
0.00	0.00	0.30		0.30	...	0.00	0.00	0.00	0.00	0.00
(0.02)	(0.02)	(0.02)		(0.02)	...	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
0.00	0.00	0.00	0.30		...	0.00	0.00	0.00	0.00	0.00
(0.02)	(0.02)	(0.02)	(0.01)		...	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
...	...	...	...	...	...	...	...	...	...	...
0.00	0.00	0.00	0.00	0.00	...		0.30	0.00	0.00	0.00
(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	...		(0.02)	(0.02)	(0.02)	(0.02)
0.00	0.00	0.00	0.00	0.00	...	0.30		0.30	0.00	0.00
(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	...	(0.02)		(0.02)	(0.02)	(0.02)
0.00	0.00	0.00	0.00	0.00	...	0.00	0.30		0.30	0.00
(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	...	(0.02)	(0.02)	( )	(0.01)	(0.02)
0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.30		0.30
(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	...	(0.02)	(0.02)	(0.01)		(0.02)
0.30	0.00	0.00	0.00	0.00	...	(0.00	0.00	0.00	0.30	
(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	...	0.02)	(0.02)	(0.02)	(0.02)	
T=20					...					
	0.27	0.04	0.00	0.00	...	0.00	0.00	0.00	0.03	0.27
	(0.08)	(0.08)	(0.08)	(0.08)	...	(0.08)	(0.08)	(0.08)	(0.09)	(0.08)
0.27		0.27	0.04	0.00	...	0.00	0.00	0.00	0.00	0.03
(0.08)		(0.08)	(0.09)	(0.08)	...	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)
0.04	0.27		0.27	0.03	...	0.00	0.00	0.00	0.00	0.00
(0.09)	(0.08)		(0.08)	(0.09)	...	(0.08)	(0.08)	(0.08)	(0.09)	(0.08)
0.00	0.03	0.27		0.27	...	0.00	0.00	0.00	0.00	0.00
(0.08)	(0.08)	(0.08)		(0.08)	...	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)
0.00	0.00	0.03	0.27		...	0.00	0.00	0.00	0.00	0.00
(0.08)	(0.08)	(0.08)	(0.08)		...	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
...	...	...	...	...	...	...	...	...	...	...
0.00	0.00	0.00	0.00	0.00	...		0.27	0.03	0.00	0.00
(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	...		(0.08)	(0.08)	(0.08)	(0.08)
0.00	0.00	0.00	0.00	0.00	...	0.27		0.27	0.04	0.00
(0.08)	(0.08)	(0.08)	(0.09)	(0.08)	...	(0.07)		(0.08)	(0.09)	(0.08)
0.00	0.00	0.00	0.00	0.00	...	0.03	0.27		0.27	0.04
(0.09)	(0.08)	(0.08)	(0.09)	(0.09)	...	(0.08)	(0.08)		(0.08)	(0.08)
0.04	0.00	0.00	0.00	0.00	...	0.00	0.04	0.27		0.27
(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	...	(0.08)	(0.08)	(0.08)		(0.08)
0.27	0.03	0.00	0.00	0.00	...	-0.01	0.00	0.04	0.27	
(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	...	(0.08)	(0.09)	(0.08)	(0.08)	

Table 2: Estimates of $\tilde{\beta}$ for model (28), where $N = 30$

T=80										
0.90	0.90	0.90	0.90	0.90	...	0.90	0.90	0.90	0.90	0.90
(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	...	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
T=20										
0.65	0.66	0.65	0.66	0.65	...	0.65	0.65	0.65	0.65	0.65
(0.08)	(0.07)	(0.07)	(0.07)	(0.08)	...	(0.07)	(0.08)	(0.07)	(0.08)	(0.08)

give a flavour of how fast two-stage VB converges, we would like to mention that, when allowing for parallel computing, each Monte Carlo replication of model (28) where $N = 100$ and $T = 50$ takes about 20 minutes, and that number reduces to 5 minutes for the same model where $N = 100$ and $T = 200$.

5 Empirical Applications

The relationship between sovereign bond ratings and spreads in eurozone countries is of great interest to researchers and policy makers alike, as explained in papers such as Gibson et al. (2021) and Hall et al. (2022). One of the important questions posed is whether rate changes of the southern euro area countries and their northern counterparts have different impacts on a member country, especially during the European debt crisis, also known as euro area crises, that started in late 2009 and lingered on well into 2015. Equally, it is important to know whether rate changes of a country group affect all the other countries differently. To answer those questions, we apply the two-stage VB approach on a sample containing five southern euro area countries – Spain, Greece, Ireland, Italy and Portugal – and five northern euro area countries – Austria, Belgium, France and 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 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; News, which is fiscal news constructed using European Commision forecasts; $\frac{Debt}{GDP}$, which is the ratio of government debt to GDP; $\frac{CA}{GDP}$, which is the ratio of current account balance to GDP; $\frac{P}{P^*}$, 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 the Appendix.

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

$$Rate_{it} = c_{11,i} + c_{12,i} * Spread_{it} + c_{13,i} * GDPgrowth_{it} + c_{14,i} * \frac{Debt_{it}}{GDP_{it}} + c_{15,i} * News +$$

$$c_{16,i} * \frac{P}{P*}_{it} + c_{17,i} * Rate_{i(t-1)} + W_i^{rate} * Rate_t + \varepsilon_{it}^{rate}$$

$$Spread_{it} = c_{21,i} + c_{22,i} * Rate_{it} + c_{23,i} * GDPgrowth_{it} + c_{24,i} * \frac{Debt_{it}}{GDP_{it}} + c_{25,i} * Pol_{it} +$$

$$c_{26,i} * \frac{CA_{it}}{GDP_{it}} + c_{27,i} * Spread_{i(t-1)} + W_i^{sp} * Spread_t + \varepsilon_{it}^{sp}$$

$$(30)$$

where W_i^{rate} and W_i^{spread} are the ith 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 (30) can be rewritten as

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

$$(31)$$

where $Rate_t = (Rate_{1t}, ..., Rate_{10t})'$, $Spread_t = (Spread_{1t}, ..., Spread_{10t})'$, $\varepsilon_t^{rate} = (\varepsilon_{1t}^{rate}, ..., \varepsilon_{10t}^{rate})'$, $\varepsilon_t^{spread} = (\varepsilon_{1t}^{spread}, ..., \varepsilon_{10t}^{spread})'$, $\mathbf{c}_{12} = diag(\mathbf{c}_{12,1}, ..., \mathbf{c}_{12,10})$, $\mathbf{c}_{22} = diag(\mathbf{c}_{22,1}, ..., \mathbf{c}_{22,10})$, $\mathbf{c}_{17} = diag(\mathbf{c}_{17,1}, ..., \mathbf{c}_{17,10})$, $\mathbf{c}_{27} = diag(\mathbf{c}_{27,1}, ..., \mathbf{c}_{27,10})$, $\mathbf{x}$ 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 (31) to calculate the impulse response functions and evaluate how rate and spread changes in one country spill over to the other countries.

We use a rolling window of 24 months to trace how the spillovers fluctuate over time. For the ith country, the average rating spillovers from the south is computed by taking the

23

mean of the cumulative impacts of a 1 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 1 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 1 basis point.

Figure 1 presents the average spillovers from the southern euro area countries and those from the north, where spillovers are measured by the cumulative impacts of 5 years.[6] Two salient patterns leap out. First, spatial spillovers between countries only become apparent after the financial crises in 2007. Before that, the spillovers between countries are barely minimum. Second, shocks to a country's spread tend to always increase other countries' spreads, but the impacts of shocks to sovereign ratings are mixed.

Let us start with the impacts of rating changes in a northern euro area country. We can observe that shocks to rates of the southern countries tends to improve the sovereign ratings of France, Finland and Netherland (causing the values of their rates to decrease), but worsen the ratings of Austria and Belgium (causing their rates to increase), especially during the period of euro area crises. For a northern country, shocks to rates of other northern countries, by contrast, tend to slightly worsen its sovereign ratings.

For a southern country, apart from Greece and Ireland, shocks to rates of the northern countries tend to have little cumulative impacts on its ratings. For Greece and Ireland, especially in the euro area crises, positive shocks to northern countries' ratings will cause their ratings to deteriorate as well. For all the southern countries, shocks that worsen other southern countries' ratings will worsen their own ratings too, in a magnitude that is much

---

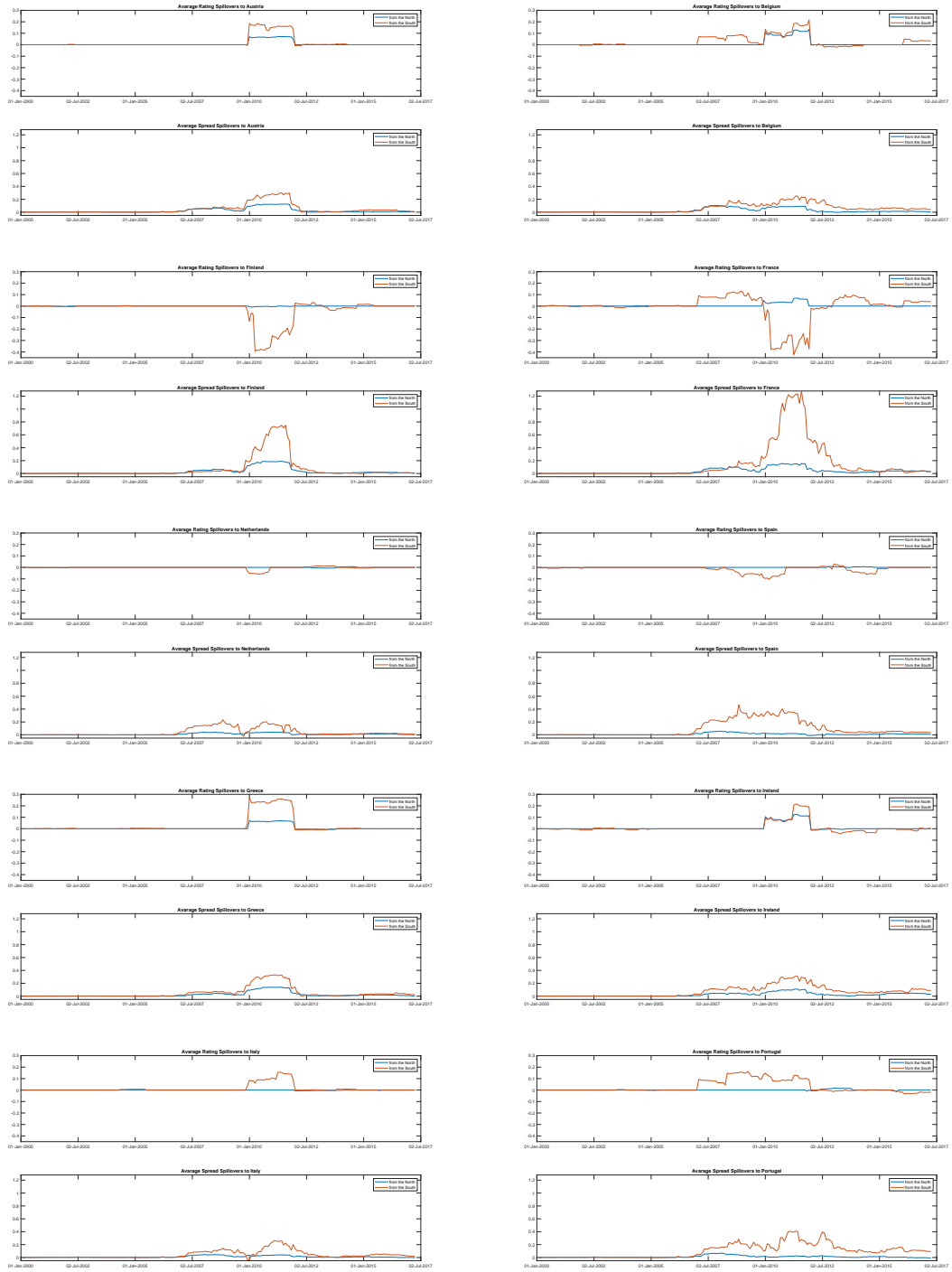[6]The impulse response functions usually converge to zero after 2 to 3 months

Figure 1: Average spillovers

higher than the impacts of the same amount of shocks to northern countries' ratings.

For all countries, an increase in the spreads of northern countries is always followed by an increase in their own spreads, making it more costly for the country to borrow, same as the impacts of an increase in the spreads of southern countries. But the latter's impacts are much larger than the former's in all cases. In euro area crises, the spreads of France, Finland and Spain increased a lot after shocks to southern countries' spreads.

Since model (30) is dynamic, the spatial spillovers measured by impulse responses are complicated by time dependence. To better understand the spatial structure, we next plot the average spatial weights associated with each country in Figure 2. For the ith 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 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. Plots in Figure 2 provides 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 countries than the southern countries, warranting we taking 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 weights matrix where all the elements are non-negative. Apart
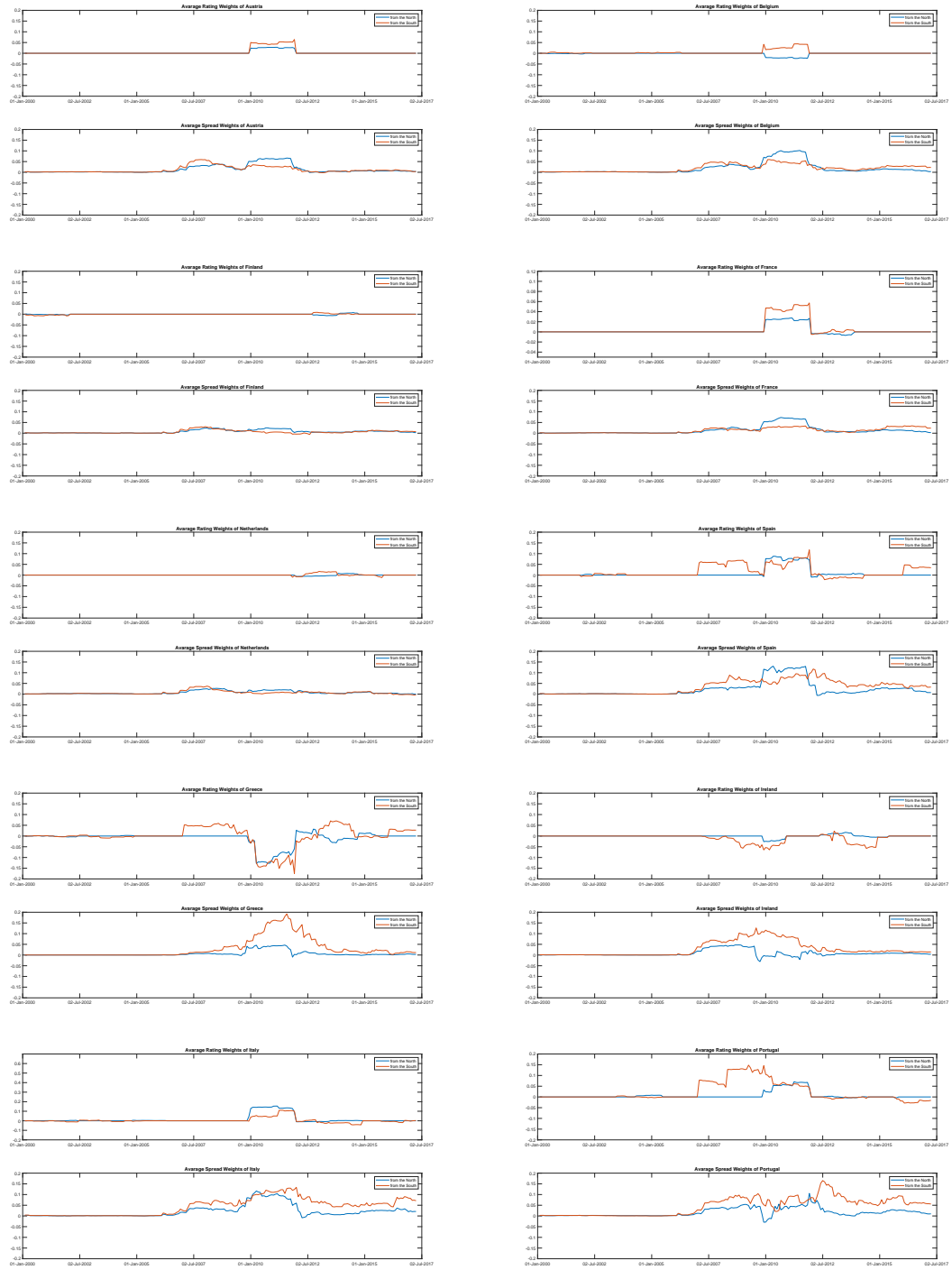
Figure 2: Spatial structures

from Greece and Ireland, a country's sovereign ratings tend to be positively associated with other countries's 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 crises 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 tend 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 euro area crises, spreads of all southern countries are more closely spatially linked with other southern countries than with the northern ones.

6   Conclusions

In applied work, if the spatial weights 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 crises, 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.

References

[1] Ahrens, A. and Bhattacharjee, A.(2015). Two-step lasso estimation of the spatial weights matrix. *Econometrics 3*, 128–55.

[2] Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic, Dordrecht.

[3] Baltagi, B. H., S. H. Song and Koh, W. (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics 117*, 123–50.

[4] Baltagi, B. H., Egger, P. and Pfaffermayr, M. (2013). A generalized spatial panel data model with random effects. *Econometric Reviews 32*, 650–85.

[5] Basak, G. A., Bhattacharjee, A. and Das, S. (2018). Causal ordering and inference on acyclic networks. *Empirical Economics 55*, 213–32.

[6] Bhattacharya, A., Pati, D., Pillai, N. S. and Dunson, D. B. (2015). Dirichlet–Laplace priors for optimal shrinkage. *Journal of the American Statistical Association 110*, 1479–90.

[7] Blei, D. M., Kucukelbir, A. and McAuliffe, J. D.(2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association 112*, 859–77.

[8] Case, A. (1991). Spatial Patterns in Household Demand. *Econometrica 59*, 953–65.

[9] Cliff, A. D. and Ord, J. K. (1973). *Spatial Autocorrelation*. London: Pion.

[10] Debarsy, N., Ertur, C. and LeSage, J.P., (2012). Interpreting dynamic space–time

panel data models. *Statistical Methodology 9*, 158–171.

[11] Fox, J. (1979). Simultaneous equation models and two-stage least squares. *Sociological methodology 10*, 130–50.

[12] Gibson, H. D., Hall, S. G. and 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*, 371–85.

[13] Gibson, H. D., Hall, S. G., Gefang,D., Petroulas, P. and 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*, 1454–70.

[14] Gefang, D., Koop, G. and Poon, A. (2020). Computationally efficient inference in large Bayesian mixed frequency VARs. *Economics Letters 191*, 109120.

[15] Gefang, D., Koop, G. and Poon, A. (2022). Forecasting using variational Bayesian inference in large vector autoregressions with hierarchical shrinkage. *International Journal of Forecasting*.

[16] Hall, S. G., Gefang, D. and Tavlas, G. S. (2023). A test to select between spatial weighting matrices. *Journal of Spatial Econometrics 4*.

[17] Krisztin, T. and Piribauer, P. (2023). A Bayesian approach for the estimation of weight matrices in spatial autoregressive models. *Spatial Economic Analysis 18*, 44–63.

[18] Krock, M., Kleiber W. and Becker, S. (2021). Nonstationary modeling with sparsity for spatial data via the basis graphical lasso. *Journal of Computational and Graphical*

Statistics 30, 375–89.

[19] Lam, C. and Souza, P. C. (2019). Estimation and selection of spatial weight matrix in a spatial lag model. *Journal of Business and Economic Statistics 38*, 693–710.

[20] Lee, L-F. and Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics 154*, 165–85.

[21] Li, Y., Craig, B. A. and Bhadra, A. (2019). The graphical horseshoe estimator for inverse covariance matrices. *Journal of Computational and Graphical Statistics 28*, 747–57.

[22] Liu, X. and Saraiva, P. (2019). GMM estimation of spatial autoregressive models in a system of simultaneous equations with heteroskedasticity. *Econometric Reviews 38*, 359–85.

[23] Loaiza-Maya, R., Smith, M. S., Nott, D. J. and Danaher P. J. (2021). Fast and accurate variational inference for models with many latent variables. *Journal of Econometrics. 230*, 339–62.

[24] Makalic, E. and Schmidt, D. F. (2016). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters 23*, 179–82.

[25] Ormerod, J. T. and Wand, M. P. (2010). Explaining variational approximations. *The American Statistician 64*, 140–53.

[26] Piribauer, P., Glocker, C. and Krisztin, T.(2023). Beyond Distance: The Spatial Relationships of European Regional Economic Growth. Available at SSRN 4391999.

[27] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society,* Series B *58*, 267–88.

[28] Wang, H. (2012). Bayesian graphical lasso models and efficient posterior computation. *Bayesian Analysis 7*, 867–86.

[29] Yang, K. and Lee, L. F. (2017). Identification and QML estimation of multivariate and simultaneous equations spatial autoregressive models. *Journal of Econometrics 196*, 196–214.

[30] Zellner, A., and Theil, H. (1962). Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations. *Econometrica 30*, 54–78.

Appendix: Data Sources

(a) Spreads: monthly data from Statistical Data Warehouse, European Central Bank

(b) Ratings: monthly data from Fitch, Standard and Poor's and Moody's

(c) The ratio of government debt to GDP: quarterly data from Thomson Reuters Datastream

(d) Real GDP growth: quarterly data from Thomson Reuters Datastream

(e) Current account balance as a percentage of GDP: Either monthly or quarterly data from Thomson Reuters Datastream

(f) Harmonised Consumer Price Index: monthly data from Thomson Reuters Datstream

(g) Political uncertainty: quarterly data from IFO World Economic Survey. Two variables have been used in order to create the variable political uncertainty. 1) The present climate for foreign investors, political stability (which was discontinued), which was updated with information from the new variable, 2) Political instability, using an appropriate mapping for the reported values of each variable.