

A Bayesian approach for the estimation of weight matrices in spatial autoregressive models

Tamás Krisztin & Philipp Piribauer

To cite this article: Tamás Krisztin & Philipp Piribauer (2022): A Bayesian approach for the estimation of weight matrices in spatial autoregressive models, Spatial Economic Analysis, DOI: [10.1080/17421772.2022.2095426](https://doi.org/10.1080/17421772.2022.2095426)

To link to this article: <https://doi.org/10.1080/17421772.2022.2095426>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 22 Jul 2022.



[Submit your article to this journal](#)



Article views: 70



[View related articles](#)



[View Crossmark data](#)

A Bayesian approach for the estimation of weight matrices in spatial autoregressive models

Tamás Krisztin ^a and Philipp Piribauer ^b

ABSTRACT

We develop a Bayesian approach to estimate weight matrices in spatial autoregressive (or spatial lag) models. Datasets in regional economic literature are typically characterized by a limited number of time periods T relative to spatial units N . When the spatial weight matrix is subject to estimation severe problems of over-parametrization are likely. To make estimation feasible, our approach focusses on spatial weight matrices which are binary prior to row-standardization. We discuss the use of hierarchical priors which impose sparsity in the spatial weight matrix. Monte Carlo simulations show that these priors perform very well where the number of unknown parameters is large relative to the observations. The virtues of our approach are demonstrated using global data from the early phase of the COVID-19 pandemic.

KEYWORDS

estimation of spatial weight matrix, spatial econometric model, Bayesian Markov chain Monte Carlo (MCMC) estimation, Monte Carlo simulations, COVID-19 pandemic

JEL C11, C21, C23, C51

HISTORY Received 2 September 2021; in revised form 11 May 2022


1. INTRODUCTION

Spatial econometrics deals with the study of cross-sectional dependence and interactions among (spatial) observations. A particularly popular spatial econometric model is the spatial autoregressive (or spatial lag) specification, where spatial interdependence between observations is governed by a so-called spatial weight matrix. The spatial weight matrix is typically assumed non-negative, row-standardized and exogenously given, with spatial weights based on some concept of neighbourhood. Geographic neighbourhood is often preferred due to exogeneity assumptions. However, when relying on geographical information, several competing approaches exist for constructing the weight matrix (for a thorough discussion, see LeSage & Pace, 2009). Recently, Kelejian and Piras (2014), Qu and Lee (2015), Han and Lee (2016), and Hsieh and Lee (2016) use alternative measures, such as (socio-)economic proximity. Another strand of the literature focuses on the uncertainty associated with the choice of neighbourhood structures by selecting or combining alternative weight matrices (e.g., Debarsy & LeSage, 2018; Piribauer & Cuaresma, 2016).

CONTACT Tamás Krisztin  krisztin@iiasa.ac.at

^aInternational Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

^bAustrian Institute of Economic Research (WIFO), Vienna, Austria

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/17421772.2022.2095426>.

Since direct estimation of a spatial weight matrix requires estimating at least $(N - 1)N$ parameters (ignoring the other model parameters), only few approaches target direct estimation of spatial weight matrices. Recently, Ahrens and Bhattacharjee (2015) and Lam and Souza (2020) tackled this problem through Least Absolute Shrinkage and Selection Operator (LASSO)-based approaches (Tibshirani, 1996), which involve (a priori) expert knowledge about the interactions between spatial units, while allowing the final estimates of the spatial weights to slightly deviate from it.¹ However, for regional economic panels, where the time dimension T is often limited relative to the number of spatial observations N , estimation results in a deleterious proliferation of the number of parameters.

In this paper we describe a novel and flexible Bayesian approach for estimation of spatial weight matrices. Our definition of spatial weight matrices fulfils the typical assumptions employed in the vast majority of the spatial econometric literature. The resulting spatial weight matrices are assumed non-negative and specific requirements to identification of the parameters can be easily implemented in a Markov chain Monte Carlo (MCMC) sampling strategy. Although our primary focus is on row-standardized spatial weight matrices, weights without row-standardization are also implementable. To make our estimation approach applicable to spatial panels where the number of time periods T is limited as compared with the number of spatial units N , we focus on spatial weight matrices which are binary prior to potential row-standardization.

In this paper we primarily focus on scenarios where no a priori information on the spatial structure is available. However, we also discuss how a priori spatial information can be implemented in a very simple and transparent way. For cases where the number of unknown parameters is large relative to the number of observations, we discuss hierarchical prior set-ups which impose sparsity in the weight matrix. In a Monte Carlo study, we show that these sparsity priors perform particularly well when the number of spatial observations N is large relative to the time periods T .

We show that our approach can be implemented in an efficient Gibbs sampling algorithm, which implies that the estimation strategy can be easily extended to other spatial econometric specifications. Among several others, such extensions include shrinkage estimation to avoid over-parameterization (Piribauer & Cuaresma, 2016), more flexible specifications of the innovation process (LeSage, 1997), controlling for unobserved spatial heterogeneity (Cornwall & Parent, 2017; Piribauer, 2016), or allowing for non-linearity in the slope parameters (Basile, 2008; Krisztin, 2017). Moreover, it is worth noting that the proposed approach can be easily adapted to matrix exponential spatial specifications (LeSage & Pace, 2007), spatial error specifications (LeSage & Pace, 2009), or local spillover models (Vega & Elhorst, 2015).

The rest of the paper is organized as follows. The next section outlines the panel version of the considered spatial lag model. The following section discusses the Bayesian estimation approach of the spatial weights along with several potential prior set-ups. The next section presents the Bayesian MCMC estimation algorithm and also discusses how to efficiently deal with the computational difficulties when updating the spatial weights in the MCMC sampler. The accuracy of the sampling procedure via a Monte Carlo simulation study is assessed in the following section. Next, we illustrate our approach using data on global infection rates of the very first phase of the recent COVID-19 pandemic. The final section concludes.

2. ECONOMETRIC FRAMEWORK

We consider a panel version of a global spillover spatial autoregressive model (SAR) of the form:²

$$y_t = \rho W y_t + \mu + \tau_t + Z_t \beta_0 + \varepsilon_t, \quad t = 1, \dots, T \quad (1)$$

where y_t denotes an $N \times 1$ vector of observations on the dependent variable measured at period t ; $\boldsymbol{\mu}$ and $\boldsymbol{\tau}_t$ represent parameters associated with fixed effects for the N spatial units and T time periods, respectively; Z_t is an $N \times q_0$ full rank matrix of explanatory variables, with corresponding $q_0 \times 1$ vector of slope parameters $\boldsymbol{\beta}_0$; and $\boldsymbol{\varepsilon}_t$ is a standard $N \times 1$ disturbance term $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_N)$.

The $N \times N$ matrix W denotes a spatial weight matrix and ρ is a (scalar) spatial dependence parameter. W is non-negative with $w_{ij} > 0$ if observation j is considered as a neighbour to i , and $w_{ij} = 0$ otherwise. A vital assumption is also that $w_{ii} = 0$, in order to avoid the case that an observation is assumed as a neighbour to itself. A frequently made assumption amongst practitioners is that W is row-stochastic with rows summing to unity. In this paper we mainly present results relating to row-stochastic weight matrices. However, as the decision on row-standardizing W depends on the empirical application, it is worth noting that the proposed approach may be easily adapted to problems without row-standardization of W .³

The reduced form of the SAR model is given by:

$$y_t = (I_N - \rho W)^{-1}(\boldsymbol{\mu} + \boldsymbol{\tau}_t + Z_t \boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}_t), \tag{2}$$

where $(I_N - \rho W)^{-1} = \sum_{r=0}^{\infty} \rho^r W^r$ is a so-called spatial multiplier matrix. To ensure that $(I_N - \rho W)$ is invertible, appropriate stability conditions need to be imposed. For row-stochastic spatial weight matrices, a sufficient stability condition for the spatial autoregressive parameter often employed is $\rho \in (-1, 1)$ (e.g., LeSage & Pace, 2009).

In most cases, the elements of W are typically treated as known. In the spatial econometric literature, there are various ways to construct such a spatial weight matrix. In this study we focus on the estimation of weight matrices which are binary prior to row-standardization. We therefore assume that the typical element of our spatial weight matrix can be obtained from an unknown $N \times N$ spatial adjacency matrix $\boldsymbol{\Omega}$ (with typical element ω_{ij}).⁴ We therefore define $W = f(\boldsymbol{\Omega})$, where $f(\cdot)$ denotes the row-standardization function:⁵

$$w_{ij} = \begin{cases} \omega_{ij} / \sum_{j=1}^N \omega_{ij} & \text{if } \sum_{j=1}^N \omega_{ij} > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The elements of the adjacency matrix $\boldsymbol{\Omega}$ are assumed as unknown binary indicators, which are subject to estimation. It is worth noting that the assumption of a binary $\boldsymbol{\Omega}$ covers a wide range of specifications commonly used in the literature such as contiguity, distance band or nearest neighbours (e.g., LeSage & Pace, 2009).

To alleviate further notation, we collect the respective dummy variables associated with the fixed effects along with the explanatory variables in an $N \times q$ matrix X_t with corresponding $q \times 1$ parameter vector $\boldsymbol{\beta}$. Moreover, define $Y = [y_1', \dots, y_T']$, $X = [X_1', \dots, X_T']$ and $S = I_T \otimes (I_N - \rho W)$ and $\mathcal{D} = \{Y, X\}$ denotes the data. The Gaussian likelihood $p(\mathcal{D}|\bullet)$ is then given by:

$$p(\mathcal{D}|\bullet) = \frac{1}{(2\pi\sigma^2)^{NT}} |S| \exp\left[-\frac{1}{2\sigma^2} (SY - X\boldsymbol{\beta})'(SY - X\boldsymbol{\beta})\right]. \tag{4}$$

When the elements of the spatial weight matrix are subject to estimation, the number of unknown parameters is likely much larger than the number of observations. Since spatial economic panels often feature limited T relative to N , the proposed estimation approach has to address the issue of over-parametrization. We discuss different ways to tackle this problem. First and foremost, one may reduce the dimensionality of the problem by imposing a priori

information on spatial weights or assuming symmetry of the spatial neighbourhood structure. Alternatively, we consider hierarchical prior set-ups which impose sparsity in the weight matrix.

When estimating spatial weights in addition to the spatial and slope parameters, identification issues are more complicated as compared with models assuming exogenous spatial weights. We therefore follow De Paula et al. (2019), who provide a thorough discussion on parameter identification for rather general SAR model specifications. As mentioned above, we consider spatial weight matrices which are non-negative and $w_{ii} = 0$ for all i . Further standard assumptions include $\sum_{j=i}^N |\rho w_{ij}| < 1 \forall i$, $|\rho| < 1$ and $\|W\| < C$ for some positive $C \in \mathbb{N}$, as well as $\beta_0 \rho \neq 0$. As an additional identifying assumption, it is important that the main diagonal elements of W^2 are not proportional to a vector of ones.⁶ Sufficient conditions for global identification are fulfilled if we make the additional assumption of $\rho > 0$ (De Paula et al., 2019, corollary 3). Without this additional restriction on ρ , De Paula et al. (2019) show that a strongly connected spatial network for global identification is needed. Since strong a priori information on the spatial weight matrix is often not available (or desired), we therefore assume $\rho \in (0, 1)$ and only consider positive spatial autocorrelation, which is a typical assumption for empirical applications.⁷

3. BAYESIAN ESTIMATION OF W

In this paper we use a Bayesian estimation approach to obtain estimates and inference on the unknown quantities ρ , β , σ^2 , as well as the elements of Ω . After eliciting suitable priors for the unknown parameters, we employ a computationally efficient MCMC algorithm.

Let $p(\omega_{ij} = 1)$ denote the prior belief in including the ij th element of the spatial weight matrix. Conversely, for a proper prior specification the prior probability of exclusion is then simply given by $p(\omega_{ij} = 0) = 1 - p(\omega_{ij} = 1)$. With Ω_{-ij} denoting the elements of the neighbourhood matrix without ω_{ij} , the posterior probabilities of $\omega_{ij} = 1$ and $\omega_{ij} = 0$ conditional on all other parameters are given by:

$$\begin{aligned} p(\omega_{ij} = 1 | \Omega_{-ij}, \beta, \sigma^2, \rho, D) &\propto p(\omega_{ij} = 1) |S_1| \exp \left[-\frac{1}{2\sigma^2} (S_1 Y - X\beta)' (S_1 Y - X\beta) \right] \\ p(\omega_{ij} = 0 | \Omega_{-ij}, \beta, \sigma^2, \rho, D) &\propto p(\omega_{ij} = 0) |S_0| \exp \left[-\frac{1}{2\sigma^2} (S_0 Y - X\beta)' (S_0 Y - X\beta) \right], \end{aligned} \quad (5)$$

where S_1 and S_0 are given by S through updating the spatial weight matrix W via setting $\omega_{ij} = 1$ and $\omega_{ij} = 0$, respectively.⁸ Using the law of total probability, it is straightforward to show that the resulting conditional posterior for ω_{ij} is Bernoulli:

$$p(\omega_{ij} | \Omega_{-ij}, \beta, \sigma^2, \rho, D) \sim \text{BER} \left(\frac{\bar{p}_{ij}^{(1)}}{\bar{p}_{ij}^{(0)} + \bar{p}_{ij}^{(1)}} \right), \quad (6)$$

with $\bar{p}_{ij}^{(1)} = p(\omega_{ij} = 1 | \Omega_{-ij}, \beta, \sigma^2, \rho, D)$ and $\bar{p}_{ij}^{(0)} = p(\omega_{ij} = 0 | \Omega_{-ij}, \beta, \sigma^2, \rho, D)$ given in equation (5). Since the conditional posterior follows a convenient and well-known form, efficient Gibbs sampling can be employed.

A Bayesian estimation framework requires elicitation of a prior on Ω . Obvious candidates are independent Bernoulli priors on the unknown indicators ω_{ij} :

$$p(\omega_{ij}) \sim \text{BER}(\underline{p}_{ij}), \quad (7)$$

where \underline{p}_{ij} denotes the prior inclusion probability of ω_{ij} , $p(\omega_{ij} = 1) = \underline{p}_{ij}$. Conversely, the prior probability of exclusion then simply takes the form $p(\omega_{ij} = 0) = 1 - \underline{p}_{ij}$.

A natural prior choice would involve setting $\underline{p}_{ij} = \underline{p} = 1/2$ for $i \neq j$, and 0 otherwise, which implies that each off-diagonal element in Ω has an equal prior chance of being included. However, in many cases a researcher has possible a priori information on the underlying structure of the spatial weight matrix. The following stylized examples demonstrate how to incorporate such information in a flexible and straightforward way.

Figure 1 illustrates the flexibility of prior elicitation for Ω in the case of a ‘linear city’ with $N = 15$ equidistant regions. Case A shows a prior specification without any prior uncertainty on the elements of W by setting $\underline{p}_{ij} = 1$ if i and j are considered as neighbours and 0 otherwise.

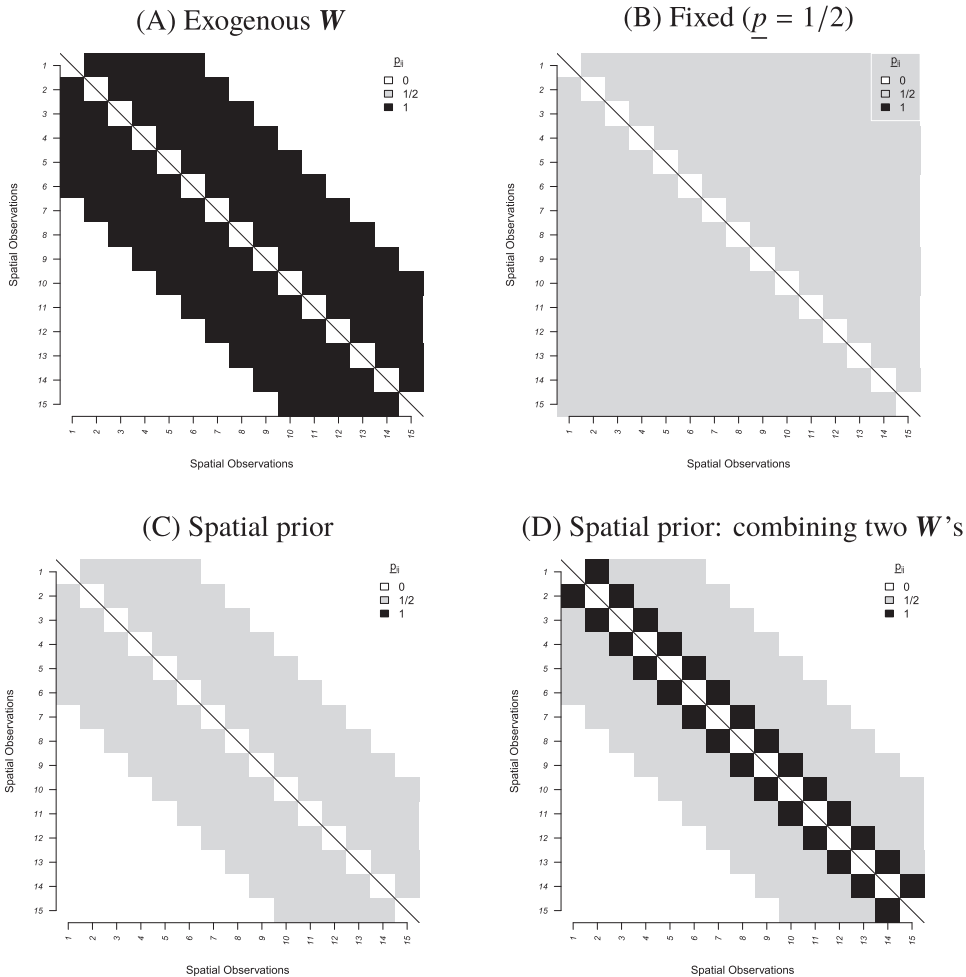


Figure 1. Some stylized prior examples for W in a linear city.
 Notes: Alternative prior set-ups for a linear city of $N = 15$ spatial observations. Case A shows a prior specification without any prior uncertainty on the spatial links. This set-up implies an exogenous W and no estimation of the weights is involved. Case B involves no spatial prior information and each element has a prior probability of inclusion $\underline{p}_{ij} = 1/2 \forall i \neq j$. Case C shows uncertainty of the linkages in W only within a certain spatial domain. Case D is a stylized prior specification considering uncertainty among two (or more) weight matrices, with setting $\underline{p}_{ij} = 1$ in regions where the two matrices overlap.

In this case, no estimation on the spatial links is involved and the model reduces to a standard SAR model with an exogenously given W (in this example, a distance band specification).

Case B depicts the opposite case where no prior spatial information is available. Specifically, this case considers full estimation of all $N^2 - N$ potential links with respective prior inclusion probability $\underline{p}_{ij} = 1/2$ for $i \neq j$.

Subplots C and D in Figure 1 depict prior set-ups where a priori spatial information is available to the researcher, but associated with uncertainty. Case C illustrates a prior where the general spatial domain is assumed as being a priori known, but uncertainty over specific linkages exists. In empirical practice, spatial weight matrices based on geographical information are often viewed as being preferable due to exogeneity assumptions to (socio-)economic data. The illustrated prior specification follows this idea by still allowing for uncertainty and flexibility among the spatial neighbourhood.

Recent contributions to spatial econometric literature propose selecting (Piribauer & Cuarasma, 2016) or combining (Debarsy & LeSage, 2018) multiple exogenous spatial weight matrices. Case D follows a similar idea by depicting a mixture of a distance band and a contiguity matrix (i.e., neighbourhood if regions share a common border). The intersecting elements of the two spatial structures (resulting in a contiguity matrix) are assumed as being included by setting $\underline{p}_{ij} = 1$.

3.1. Hierarchical prior set-ups and sparsity

The prior structure in equation (7) involves fixed inclusion probabilities \underline{p} , which implies that the number of neighbours of observation i follows a Binomial distribution $\sum_{l=1}^{N-1} \omega_{il} \sim BN(N-1, \underline{p})$ with a prior expected number of neighbours of $(N-1)\underline{p}$. However, such a prior structure has the potential undesirable effect of promoting a relatively large number of neighbours. For example, when $\underline{p} = 1/2$, the prior expected number of neighbours is $(N-1)/2$, since combinations of ω_{ij} resulting in such a neighbourhood size are dominant in number.

To put more prior weight on parsimonious neighbourhood structures and therefore promote sparsity in the adjacency matrix, one may explicitly account for the number of linkages in each row of the adjacency matrix $\boldsymbol{\omega}_i = [\omega_{i1}, \dots, \omega_{iN}]$. We consider a flexible prior structure on the number of neighbours $\sum \boldsymbol{\omega}_i$ that corresponds to a beta-binomial distribution $\mathcal{BB}(N-1, \underline{a}_\omega, \underline{b}_\omega)$ with two prior hyperparameters $\underline{a}_\omega, \underline{b}_\omega > 0$. The beta-binomial distribution is the result of treating the prior inclusion probability \underline{p} as random (rather than being fixed) by placing a hierarchical beta prior on it. For ω_{ij} , the resulting prior can be written as follows:

$$p(\omega_{ij}) \propto \Gamma(\underline{a}_\omega + \sum \boldsymbol{\omega}_i) \Gamma(\underline{b}_\omega + (N-1) - \sum \boldsymbol{\omega}_i), \quad (8)$$

where $\Gamma(\cdot)$ denotes the Gamma function, and \underline{a}_ω and \underline{b}_ω are prior hyperparameters.

In the case of $\underline{a}_\omega = \underline{b}_\omega = 1$, the prior takes the form of a discrete uniform distribution over the number of neighbours. By fixing $\underline{a}_\omega = 1$, we follow Ley and Steel (2009) and anchor the prior expected number of neighbours \underline{m} via $\underline{b}_\omega = [(N-1) - \underline{m}]/\underline{m}$.

4. BAYESIAN MCMC ESTIMATION OF THE MODEL

This section presents the Bayesian MCMC estimation algorithm for the proposed modelling framework. Estimation is carried out using an efficient Gibbs sampling scheme. The only exception is the sampling step for the spatial (scalar) autoregressive parameter ρ , where we propose using a standard Griddy-Gibbs step.⁹ The sampling scheme involves the following steps:

- Set starting values for the parameters (e.g., by sampling from the prior distributions).
- Sequentially update the parameters by subsequently sampling from the conditional posterior distributions presented in this section.

The second step is repeated for B times after discarding the first B_0 draws as burn-ins.

4.1. Sampling $\boldsymbol{\beta}$ and σ^2

For the slope parameters $\boldsymbol{\beta}$ and the error variance σ^2 we use common Normal and inverted Gamma prior specifications, respectively. Specifically, $p(\boldsymbol{\beta}) \sim \mathcal{N}(\boldsymbol{\theta}, \underline{V}_\beta)$ and $p(\sigma^2) \sim \mathcal{IG}(\underline{a}_{\sigma^2}, \underline{b}_{\sigma^2})$, where \underline{V}_β , \underline{a}_{σ^2} and \underline{b}_{σ^2} denote prior hyperparameters.

The resulting conditional posterior distribution is Gaussian and of well-known form (e.g., LeSage & Pace, 2009):

$$\begin{aligned} p(\boldsymbol{\beta}|\sigma^2, \rho, \boldsymbol{\Omega}, \mathcal{D}) &\sim \mathcal{N}(\bar{\boldsymbol{b}}_\beta, \bar{\mathbf{V}}_\beta) \\ \bar{\boldsymbol{b}}_\beta &= \sigma^{-2} \bar{\mathbf{V}}_\beta \mathbf{X}' \mathbf{S} \mathbf{Y} \\ \bar{\mathbf{V}}_\beta &= (\sigma^{-2} \mathbf{X}' \mathbf{X} + \underline{\mathbf{V}}_\beta^{-1})^{-1}. \end{aligned} \quad (9)$$

The conditional posterior of σ^2 is inverted Gamma:

$$\begin{aligned} p(\sigma^2|\boldsymbol{\beta}, \rho, \boldsymbol{\Omega}, \mathcal{D}) &\sim \mathcal{IG}(\bar{a}_{\sigma^2}, \bar{b}_{\sigma^2}) \\ \bar{a}_{\sigma^2} &= \underline{a}_{\sigma^2} + N\Gamma/2 \\ \bar{b}_{\sigma^2} &= \underline{b}_{\sigma^2} + (\mathbf{S}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{S}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}). \end{aligned} \quad (10)$$

4.2. Sampling ρ

For the spatial parameter ρ , we use a standard Beta distribution (LeSage & Pace, 2009, p. 142). The conditional posterior is given by:

$$p(\rho|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Omega}, \mathcal{D}) \propto p(\rho)|\mathcal{S}| \exp\left[-\frac{1}{2\sigma^2}(\mathbf{S}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{S}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right]. \quad (11)$$

Note that the conditional posterior for ρ does not follow a well-known form and thus requires alternative sampling techniques. We follow LeSage and Pace (2009) and use a Griddy-Gibbs step (Ritter & Tanner, 1992) to sample ρ .¹⁰

4.3. Sampling the elements of the adjacency matrix $\boldsymbol{\Omega}$

As discussed in the previous section, we propose two alternative prior specifications for the unknown indicators of the spatial weight matrix ω_{ij} . First, an independent Bernoulli prior structure with fixed inclusion probabilities (7). Second, a hierarchical prior structure which treats the inclusion probabilities as random (8). After eliciting the prior, the binary indicators ω_{ij} can be sequentially sampled in random order from a Bernoulli distribution with conditional posterior given in (6).

4.4. Fast computation of the determinant terms

For the Bayesian MCMC algorithm, it is worth noting that repeated sampling from equation (6) is required. However, this requires evaluating the conditional probabilities $p(\omega_{ij} = 1|\cdot)$ and $p(\omega_{ij} = 0|\cdot)$ in equation (5). The main computational difficulty lies in the calculation of the determinants $|\mathcal{S}_0|$ and $|\mathcal{S}_1|$, which has to be carried out per Gibbs sampling step for the $N^2 - N$ unknown elements of the spatial adjacency matrix. The computational costs associated with direct calculation of these determinants steeply rises with N – in fact by a factor of $\mathcal{O}(N^3)$.

This makes direct evaluation of the determinant prohibitively expensive, especially for large values of N . To avoid direct evaluation, we provide computationally efficient updates for the determinant, allowing for estimation of models with larger sample sizes.

It is worth noting that it is not necessary to directly calculate the determinant of the $NT \times NT$ matrix S_z (with $z \in \{0, 1\}$). Only the determinant of the $N \times N$ matrix $A_z = I_N - \rho W_z$ needs to be updated, since $|S_z| = |I_T \otimes A_z| = |A_z|^T$. Here, W_z denotes the spatial weight matrix obtained by setting $\omega_{ij} = 1$ and $\omega_{ij} = 0$, respectively.

Direct evaluation of $|A_z|$ can be largely avoided, since updating ω_{ij} changes only the i -th row of A , if we do not restrict $\mathbf{\Omega}$ to be symmetric (we will address this case shortly). To illustrate, let $\mathbf{\Omega}^{(c)}$ denote the current – to be updated – spatial adjacency matrix, and $W^{(c)}$ the associated spatial weight matrix with determinant $|A^{(c)}| = |I_N - \rho W^{(c)}|$. Using the so-called matrix determinant lemma, we can efficiently calculate:

$$|A_z| = |A^{(c)} + \mathbf{v}_i \boldsymbol{\delta}'_i| = \{1 + \boldsymbol{\delta}'_i (A^{(c)})^{-1} \mathbf{v}_i\} |A^{(c)}|.$$

Where \mathbf{v}_i is an $N \times 1$ vector of zeros, except for its i -th entry, which is unity. The $N \times 1$ vector $\boldsymbol{\delta}_i$ contains the differences between the i -th row of A_z and the i -th row of $A^{(c)}$.

It becomes clear that equation (12) provides a computationally cheap way for updating the determinant $|A_z|$, conditional on $|A^{(c)}|$ and $(A^{(c)})^{-1}$. This implies that during the MCMC procedure, for each update of ω_{ij} , we have to keep track of the determinant (for which equation 12 provides a simple update) and the inverse of A_z . Direct evaluation of A_z^{-1} is – similar to direct evaluation of the determinant – prohibitively expensive for moderate to large N , since it has to be carried out for each unknown element of $\mathbf{\Omega}$. However, we can rely on the so-called Sherman–Morrison formula to avoid direct evaluation of the matrix inverse:

$$A_z^{-1} = (A^{(c)} + \mathbf{v}_i \boldsymbol{\delta}'_i)^{-1} = (A^{(c)})^{-1} - \frac{(A^{(c)})^{-1} \mathbf{v}_i \boldsymbol{\delta}'_i (A^{(c)})^{-1}}{1 + \boldsymbol{\delta}'_i (A^{(c)})^{-1} \mathbf{v}_i}. \quad (13)$$

Combining the formulas in equations (12) and (13) thus provides a numerically cheap and viable way to update the elements of the spatial adjacency matrix.¹¹

The binary nature of ω_{ij} can be exploited for additional computational gains. Either A_0 or A_1 always exactly equals $A^{(c)}$ and thus its determinant and inverse is already known. This only necessitates calculating $|A_z|$ and $(A_z)^{-1}$ for only $z = 1$ or for $z = 0$, but not both.

If a symmetric spatial adjacency matrix $\mathbf{\Omega}$ is assumed, the update process remains generally the same, however the determinant and matrix inverse updates have to be performed iteratively. In this case, both ω_{ij} and ω_{ji} (for $i \neq j$) are set to either 1 or 0. Thus, both the i -th and the j -th row of A_z differ from $A^{(c)}$. Following the notation in the non-symmetric case, let us denote the differences between these rows as $\boldsymbol{\delta}_i$ and $\boldsymbol{\delta}_j$. To obtain an update of $|A_z|$ and A_z^{-1} , we first evaluate equations (12) and (13), based on $\boldsymbol{\delta}_i$, \mathbf{v}_i , $|A^{(c)}|$, and $(A^{(c)})^{-1}$. Using the resulting determinant and matrix inverse, as well as \mathbf{v}_j , and $\boldsymbol{\delta}_j$, we again evaluate equations (12) and (13), which yield $|A_z|$ and A_z^{-1} .

5. SIMULATION STUDY

To assess the accuracy of our proposed approach, we evaluate its performance in a Monte Carlo study. Our benchmark data generating process comprises two randomly generated explanatory variables, as well as spatial unit and time fixed effects:

$$\tilde{y}_t = \tilde{\rho} \tilde{W} \tilde{y}_t + \tilde{\boldsymbol{\mu}} + \tilde{\tau}_t + \tilde{Z}_t \tilde{\boldsymbol{\beta}}_0 + \tilde{\boldsymbol{\varepsilon}}_t.$$

To maintain succinct notation, we denote the simulated values in the Monte Carlo study with a

tilde. The matrix of explanatory variables \tilde{Z}_t is defined as $\tilde{Z}_t = [\tilde{z}_{1t}, \tilde{z}_{2t}]$, where both \tilde{z}_{1t} and \tilde{z}_{2t} are normally distributed with 0 mean and variance of 1, $q_0 = 2$. The corresponding vector of coefficients is defined as $\tilde{\beta}_0 = [-1, 1]'$. The vector of residuals $\tilde{\epsilon}_t$ is generated from a normal distribution with 0 mean and $\tilde{\sigma}^2 = 0.5$. The fixed effects parameters $\tilde{\mu}$ and $\tilde{\tau}_t$ are randomly generated from a standard normal distribution.

The row-stochastic spatial weight matrix \tilde{W} is based on an adjacency matrix $\tilde{\Omega}$, which is generated from an $N/20$ nearest neighbour specification, by additionally assuming symmetry of the weight matrix prior to row-standardization.¹² The nearest-neighbour specification is based on a randomly generated spatial location pattern, sampled from a normal distribution with 0 mean and unity variance. In the Monte Carlo study we vary $T \in \{10, 40\}$ and $N \in \{20, 100\}$. Additionally, we vary the strength of spatial dependence $\tilde{\rho} \in \{0.3, 0.5, 0.8\}$.

For the Monte Carlo simulation study, we compare the following prior set-ups:

- *Fixed* ($\underline{p} = 1/2$) prior, which corresponds to the fixed Bernoulli prior specification in equation (7), where we set $\underline{p} = 1/2$.
- *Sparsity* ($\underline{m} = (N-1)/2$) prior, which is analogous to the prior set-up in equation (8), with $\underline{a}_\omega = \underline{b}_\omega = 1$. This prior set-up corresponds to a discrete uniform distribution over the number of neighbours.
- *Sparsity* ($\underline{m} = N/10$) prior set-up, which corresponds to equation (8), with $\underline{a}_\omega = 1$ and $\underline{b}_\omega = [(N-1) - \underline{m}]/\underline{m}$. We set the number of a priori expected neighbours to $\underline{m} = N/10$. This prior set-up thus imposes more sparsity in Ω as compared with the former.

For all prior specifications under scrutiny, we consider two alternative estimation set-ups by assuming that the adjacency matrix is either *symmetric* or *non-symmetric*.¹³ We moreover report the predictive performance of two alternative specifications using exogenous weight matrices. In these cases the employed weights are based on the true (symmetric) adjacency matrix by fixing the accuracy to the 99% and 95% level, respectively. We simulate such cases by randomly switching 1% and 5% of the elements in the true binary adjacency matrix $\tilde{\Omega}$, respectively. The resulting exogenous adjacency matrices thus result in exactly 99% and 95% overlap in the binary observations with the true adjacency matrix, while maintaining the same level of sparsity.

The prior set-up for our remaining parameters is as follows. We assume a Gaussian prior for β with 0 mean and a variance of 100. We use an inverse gamma prior for σ^2 with rate and shape parameters 0.01. The prior for the spatial autoregressive parameter ρ is a symmetric Beta specification with shape and rate parameters equal to 1.01. The chosen priors can thus be considered highly non-informative.

In [Table 1](#) we use several criteria to evaluate the performance of the alternative specifications. For the spatial autoregressive and the slope parameters we report the well-known root mean squared error (RMSE). For assessing the ability to estimating the spatial adjacency matrix, we use the measure of accuracy. The accuracy measure is defined as the sum of correctly identified unknown elements, divided by the number of total elements to be estimated. This measure is calculated separately for each posterior draw. The reported value is an average over all posterior draws and Monte Carlo iterations.

[Table 1](#) summarizes the results of our Monte Carlo simulation. For all combinations of N , T , $\tilde{\rho}$ under scrutiny, it presents the respective RMSE for both the slope coefficients β and the spatial autoregressive parameter. The third block shows the accuracy of the estimated adjacency matrix Ω . Lower values in terms of RMSEs indicate outperformance. Conversely, for accuracy in Ω higher values indicate outperformance. The best performance among the three employed prior scenarios within a subgroup is highlighted in bold. In addition, the last two columns in [Table](#)

Table 1. Monte Carlo simulation results.

	N	T	$\tilde{\rho}$	Non-symmetric			Symmetric			Exogenous	
				Fixed	Sparsity	Sparsity	Fixed	Sparsity	Sparsity	W	
				$\underline{\rho} = 1/2$	$\underline{m} = N/2$	$\underline{m} = N/10$	$\underline{\rho} = 1/2$	$\underline{m} = n/2$	$\underline{m} = n/10$	0.99	0.95
RMSE(β)	20	40	0.3	0.193	0.161	0.163	0.162	0.163	0.164	0.168	0.176
			0.5	0.172	0.173	0.172	0.170	0.169	0.171	0.179	0.216
			0.8	0.169	0.169	0.169	0.165	0.166	0.167	0.287	0.553
		10	0.3	0.234	0.207	0.198	0.203	0.182	0.181	0.192	0.204
			0.5	0.257	0.210	0.206	0.189	0.191	0.190	0.206	0.253
			0.8	0.217	0.216	0.217	0.205	0.204	0.206	0.371	0.658
	100	40	0.3	0.098	0.099	0.097	0.099	0.099	0.098	0.079	0.080
			0.5	0.144	0.088	0.083	0.145	0.114	0.076	0.084	0.086
			0.8	0.154	0.087	0.088	0.073	0.081	0.081	0.089	0.141
		10	0.3	0.111	0.112	0.111	0.111	0.111	0.112	0.092	0.093
			0.5	0.135	0.118	0.104	0.135	0.136	0.118	0.088	0.094
			0.8	0.346	0.143	0.140	0.254	0.102	0.102	0.100	0.151
RMSE(ρ)	20	40	0.3	0.199	0.029	0.031	0.030	0.029	0.030	0.034	0.060
			0.5	0.035	0.040	0.042	0.035	0.035	0.035	0.039	0.083
			0.8	0.021	0.021	0.022	0.018	0.018	0.018	0.084	0.177
		10	0.3	0.237	0.152	0.094	0.291	0.147	0.106	0.058	0.080
			0.5	0.155	0.060	0.054	0.109	0.053	0.051	0.059	0.114
			0.8	0.027	0.032	0.032	0.028	0.028	0.029	0.097	0.179
	100	40	0.3	0.280	0.283	0.277	0.279	0.283	0.287	0.027	0.033
			0.5	0.447	0.109	0.101	0.446	0.353	0.220	0.021	0.054
			0.8	0.148	0.044	0.047	0.049	0.024	0.024	0.034	0.097
		10	0.3	0.242	0.256	0.268	0.245	0.252	0.274	0.050	0.062
			0.5	0.373	0.176	0.141	0.371	0.391	0.404	0.041	0.074

Accuracy Ω	20	40	0.8	0.473	0.106	0.110	0.169	0.141	0.137	0.044	0.105
			0.3	0.648	0.930	0.954	0.963	0.982	0.983	0.990	0.950
			0.5	0.983	0.988	0.989	0.998	0.998	0.998	0.990	0.950
			0.8	0.995	0.995	0.995	0.999	1.000	0.999	0.990	0.950
		10	0.3	0.554	0.752	0.866	0.679	0.875	0.904	0.990	0.950
			0.5	0.734	0.898	0.931	0.915	0.962	0.967	0.990	0.950
			0.8	0.975	0.983	0.984	0.996	0.997	0.997	0.990	0.950
			0.3	0.530	0.713	0.847	0.539	0.686	0.848	0.990	0.950
	100	40	0.5	0.530	0.898	0.929	0.539	0.793	0.933	0.990	0.950
			0.8	0.847	0.966	0.966	0.978	0.977	0.977	0.990	0.950
			0.3	0.530	0.713	0.844	0.539	0.685	0.846	0.990	0.950
			0.5	0.530	0.746	0.883	0.539	0.702	0.905	0.990	0.950
		10	0.8	0.531	0.926	0.933	0.564	0.944	0.944	0.990	0.950

Note: Results are based on 1000 Monte Carlo iterations. For each, the corresponding sampling algorithms are run using 500 draws, where the initial 500 were discarded as burn-in. The values given for RMSE(β) and RMSE(ρ) correspond to the average root mean squared error over all Monte Carlo iterations. Bold values denote the best performing specification within a section (symmetric or non-symmetric). The exogenous Ω specifications correspond to classic SAR models with randomly perturbed exogenous adjacency matrices, which have an accuracy of 99% and 95% compared with the true adjacency matrix. For RMSEs, lower values indicate outperformance. Conversely, for the accuracy indicators of Ω , higher values indicate outperformance.

1 show the results for the benchmark SAR models using exogenous randomly perturbed adjacency matrices with accuracy fixed at the 99% and 95% levels, respectively.

Intuitively, the precision of the estimation improves as the number of observations NT increases in proportion to the number of unknown parameters.¹⁴ The results in Table 1 largely confirm this intuition. The performance indicators for both ρ and Ω also clearly improve for high levels of spatial autocorrelation ($\rho = 0.8$). In scenarios where the number of unknown parameters is smaller than the number of observations our approach even manages to outperform both rather hard benchmarks using exogenous spatial weight matrices close to the true Data Generating Process (DGP). This relative outperformance appears particularly pronounced when the strength of spatial dependence ρ is large. In these settings, symmetric specifications (which resemble the true DGP) even manage to produce accuracy in the adjacency matrix close to unity.

Particularly interesting results appear in the most challenging Monte Carlo scenarios, where the number of unknown parameters is particularly large relative to the number of observations ($N = 100$ and $T = 10$). In these scenarios, the number of parameters to be estimated exceeds the number of observations by a factor of more than 10. In these cases, prior specifications without using shrinkage appear to fail estimating the underlying spatial structure by producing rather poor accuracy measures. However, when employing sparsity priors, Table 1 reveals that our approach still manages to produce relatively accurate predictive results. In the existence of pronounced spatial autocorrelation, the sparsity specifications even manage to closely track the predictive performance of the rather tough exogenous benchmarks.

Note that the symmetric specifications (where we impose $\omega_{ij} = \omega_{ji}$) typically outperform their non-symmetric counterparts due to their resemblance to the true DGP. However, for settings where the number of unknown parameters is smaller than the number of observations both scenarios track each other closely. Among the alternative prior specifications under scrutiny, Table 1 shows rather similar results (no clear best specification emerges) in scenarios where N is small relative to T . However, for particularly over-parametrized settings (high N and low T) the proposed *sparsity* priors particularly outperform the *fixed* set-ups. Specifically, even in the scenario with $N = 100$ and $T = 10$, the sparsity priors still perform comparatively well.¹⁵

6. EMPIRICAL ILLUSTRATION

To illustrate our proposed approach using real data, we estimate spatial panel specifications based on country-specific daily infection rates in the very early phase of the coronavirus pandemic. We use the COVID-19 data set provided by Johns Hopkins University (Dong et al., 2020). The database contains information on (official) daily infections for a large panel of countries around the globe. For the empirical illustration, we focus on the very beginning of the outbreak by using data from 17 February to 20 April 2020.

The starting date of our sample marks the beginning of the pandemic in major countries, such that large parts of Asia, Europe and North America can be included.¹⁶ The choice of the end date is motivated by the results of Krisztin et al. (2020), where the degree of spatial dependence among infections rates becomes insignificant after 20 April, when the majority of countries in the sample implemented lockdown policies.

For the empirical application we use data for the following countries: Australia (AUS), Bahrain (BHR), Belgium (BEL), Canada (CAN), China (CHN), Finland (FIN), France (FRA), Germany (DEU), Iran (IRN), Iraq (IRQ), Israel (ISR), Italy (ITA), Japan (JPN), Kuwait (KWT), Lebanon (LBN), Malaysia (MYS), Oman (OMN), Republic of Korea (KOR), Russian Federation (RUS), Singapore (SGP), Spain (ESP), Sweden (SWE), Thailand (THA), United Arab Emirates (ARE), United Kingdom (GBR), United States of America (USA) and Viet Nam (VNM).

By including a biweekly time lag, our resulting panel thus comprises $N = 27$ countries across the globe for a period of $T = 19$ days.¹⁷ We follow the work of Guliyev (2020), Krisztin et al. (2020) and Han et al. (2021), among others, and use panel versions of a spatial growth specification for the country-specific COVID-19 infections:

$$y_t = \boldsymbol{\mu} + \tau_t + \rho W y_{t-r} + x_{t-14} \boldsymbol{\beta} + Z_{t-14} \boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}_t, \tag{14}$$

where $y_t = x_t - x_{t-14}$, and x_t is an $N \times 1$ vector comprising the (logged) daily number of official cases per 100,000 inhabitants per country for time period $t = 1, \dots, T$.¹⁸ $\boldsymbol{\mu}$ and τ_t represent fixed effects for the countries and the time periods, respectively. W denotes the spatial weight matrix with spatial autoregressive parameter ρ as defined before. We again primarily focus on row-stochastic weight matrices. Results based on spatial weight matrices without row-standardization are presented in the Appendix in the supplemental data online.

We also consider alternative model specifications using contemporaneous as well as temporal lags of the spatial lag ($W y_{t-r}$ with $r \in \{0, 14\}$). A plethora of recent studies exploit the contemporaneous spatial information ($r = 0$) for modelling the spread of COVID-19 infections (e.g., Guliyev, 2020; Han et al., 2021; Jaya & Folmer, 2021; Kosfeld et al., 2021; Krisztin et al., 2020). Using contemporaneous spatial information appears reasonable when the primary interest lies in quantifying spatial co-movements of infection rates. However, for many questions of interest, a temporal spatial lag $W y_{t-r}$ ($r > 0$) might be an interesting alternative since it reflects the notion that the spatial process of virus transmission takes some time to manifest (Elhorst, 2021; Mitze & Kosfeld, 2022). Since our proposed estimation approach can be easily applied to these alternative specifications, we provide estimates for both specifications.¹⁹

In addition to the Initial infections variable x_{t-14} , matrix Z_{t-14} contains three explanatory variables on a daily basis. Several studies emphasize the importance of climatic condition on the COVID-19 virus spread. For a survey on the effects of climate on the spread of the COVID-19 pandemic, see Briz-Redón and Serrano-Aroca (2020). We therefore use daily data on the country specific maximum measured temperature (Temperature) and precipitation levels (Precipitation) as additional covariates. Both variables stem from a daily database of

Table 2. Estimation results for benchmark specifications.

	$W y_t$				$W y_{t-14}$			
	Fixed		Sparsity		Fixed		Sparsity	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Initial infections	-0.8761	0.0117	-0.9244	0.0117	-0.9533	0.0126	-0.9911	0.0114
Stringency	-0.4566	0.0736	-0.5661	0.0451	-0.2503	0.0858	0.0616	0.0410
Precipitation	0.0365	0.0339	-0.0444	0.0335	0.0541	0.0608	0.0483	0.0511
Temperature	-0.0014	0.0015	-0.0016	0.0015	-0.0032	0.0026	-0.0017	0.0025
ρ	0.6319	0.0129	0.5592	0.0101	0.9618	0.0110	0.9481	0.0139
σ^2	0.0187	0.0013	0.0209	0.0014	0.0401	0.0034	0.0516	0.0036
Average no. of neighbours	7.8370		3.6083		4.2849		2.8082	
Fixed effects	Yes		Yes		Yes		Yes	
N	27		27		27		27	
T	19		19		19		19	

Note: Posterior quantities based on 5000 MCMC draws, where the first 2500 were discarded as burn-ins. Values in bold denote significance under a 90% posterior credible interval.

country-specific data, which was compiled via the Dark Sky API.²⁰ As a third variable, we also include the well-known stringency index (Stringency) put forward by Hale et al. (2020), which summarizes country-specific governmental policy measures to contain the spread of the virus. In this application, we use the biweekly average of the reported stringency index. Since all these influences arguably require some time to be reflected in the official infection figures, we use a biweekly lag of 14 days (in accordance with r in alternative variants).²¹

Table 2 presents a summary of the estimation results. The left part shows results for specifications using a contemporaneous spatial lag Wy_t , while the right part summarizes results for the case Wy_{t-14} .

For each specification, the first rows contain the posterior mean and standard deviations for the slope parameters followed by estimates of ρ and σ^2 . Posterior quantities which appear significantly different from zero using a 90% posterior credible interval are depicted in bold. Table 2

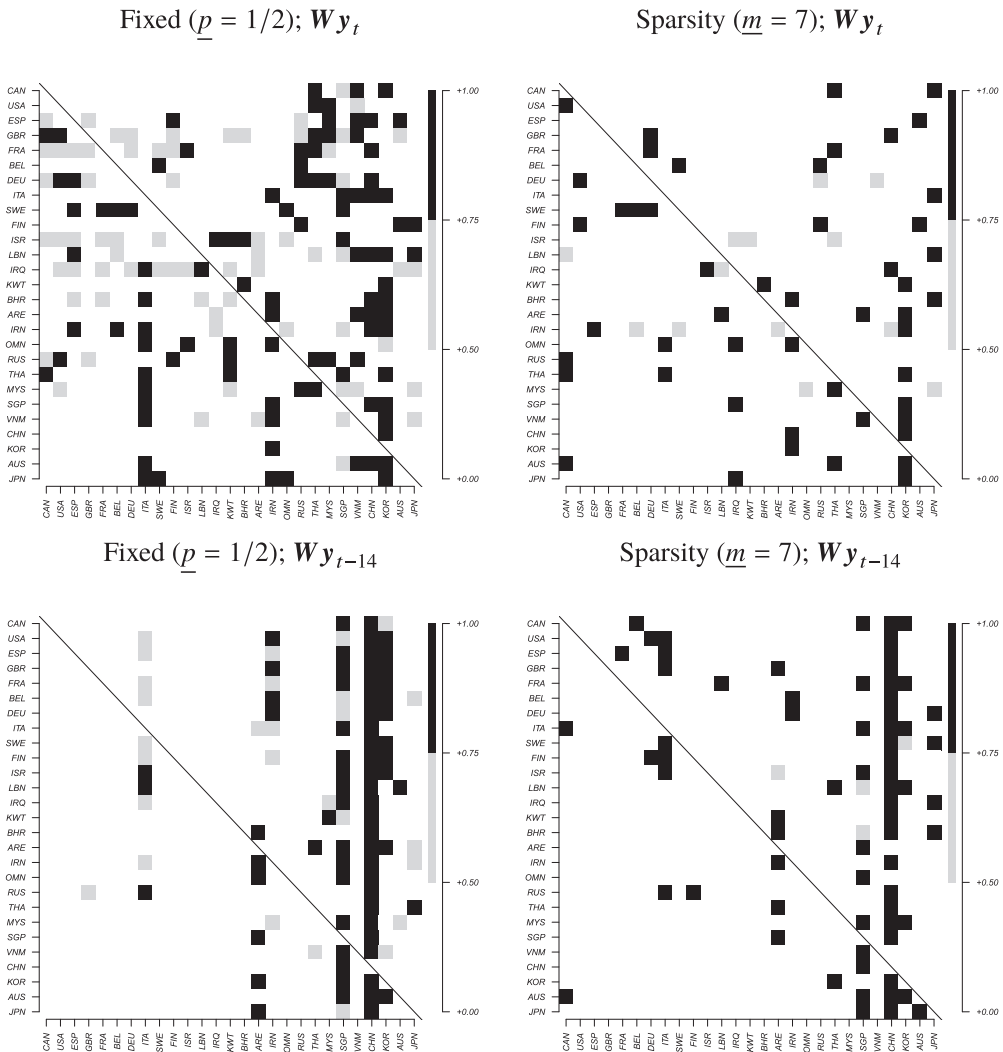


Figure 2. Posterior inclusion probabilities for benchmark specifications. Note: Posterior inclusion probabilities of spatial links are based on 5000 MCMC draws. Inclusion probabilities 0.50–0.75 (little evidence for inclusion) are coloured grey. Strong evidence for inclusion (> 0.75) coloured black.

moreover presents the average posterior expected number of neighbours, which is given by the average row sum of the matrix of posterior inclusion probabilities based on $p(\omega_{ij} = 1|\mathcal{D})$. This measure can be viewed as a measure of sparsity in the estimated matrix of linkages. All specifications moreover contain fixed effects for both N and T .²²

Table 2 shows rather similar ρ and σ^2 posterior quantities for the flat and the sparsity prior. However, there appear some marked differences between the specifications Wy_t and Wy_{t-14} . In all cases, spatial dependence appears strong and precisely estimated, but appears particularly high in the temporal lag specification Wy_{t-14} . However, Table 2 similarly reveals higher estimates for the nuisance parameter σ^2 for the temporal spatial lag models. It shows rather precise and negative coefficients for the initial infections variable, indicating conditional convergence patterns. For most model variants, Table 2 moreover suggests a significant negative impact of the stringency index on infection growth. The majority of the slope parameter estimates associated with the variables temperature and precipitation appear more muted and insignificant. Overall,

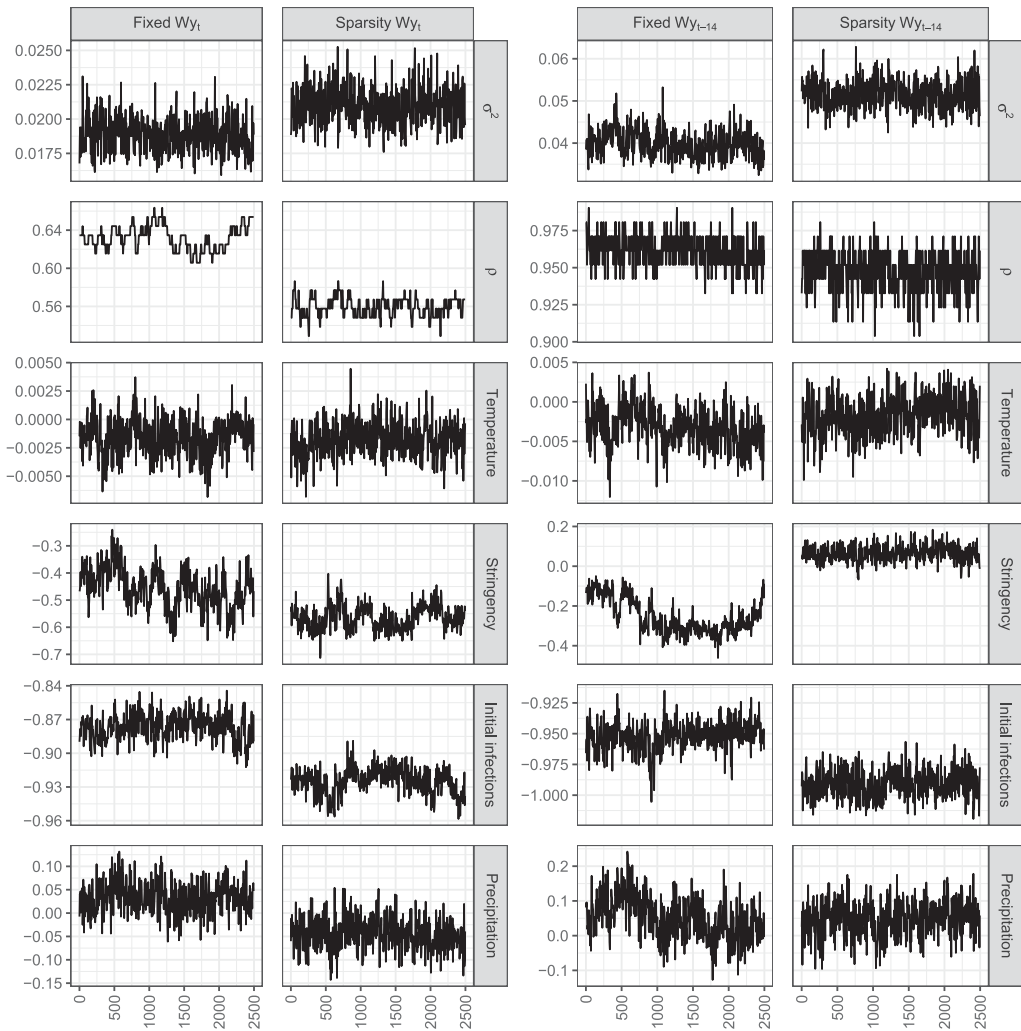


Figure 3. Trace plots for benchmark specifications.

Note: Posterior draws are based on 5000 MCMC draws, where the first 2500 were discarded as burn-ins.

Table 2 moreover clearly demonstrates that a hierarchical prior set-up can enforce sparsity in the resulting adjacency matrix. Both sparsity specifications result in an average number of neighbours smaller than the models with fixed prior specifications.

Figure 2 depicts the posterior inclusion probabilities $p(\omega_{ij} = 1|\mathcal{D})$ for the considered specifications. To better visualize the results we have reordered the countries by their longitudes, starting with Canada and the United States and ending with Southeast Asian countries, Australia and Japan. Clusters along the main diagonal thus roughly indicate geographic spatial linkages. For the sake of visualization, we distinguish between negligible evidence for inclusion (< 0.50 ; white colour), moderate evidence ($0.50 - 0.75$; grey colour), and strong evidence (> 0.75 ; black colour).

The two upper plots in Figure 2 depict posterior inclusion probabilities $p(\omega_{ij} = 1|\mathcal{D})$ for the specifications involving a contemporaneous spatial lag Wy_t , while the lower part shows temporal spatial lag specifications Wy_{t-14} . In both cases, the left subplots present results based on independent prior inclusion probabilities of $\underline{p} = 1/2$. The right plots are based on sparsity priors using $\underline{m} = 7$. The columns in the subplots indicate marginal posterior importance of the countries as predictors of coronavirus infections in linked countries. Conversely, rows depict the countries to be predicted. The results using sparsity priors generally produce similar patterns as the fixed prior specifications and clearly demonstrate its ability of dimension reduction in the connectivity structure. For the contemporaneous spatial lag specification (upper plots), Figure 2 suggests a slightly more pronounced regional dependency structure as compared with the temporal spatial lags. Moreover, it reveals marked spill-out effects from Asian countries, as well as from Iran and Italy.²³

Results based on a biweekly temporal spatial lag Wy_{t-14} show even more pronounced spill-out effects from Asian countries (most notably China, Republic of Korea and Singapore).²⁴ For European countries, results similarly suggest Italy as a further important source country of spatial virus transmission. The estimated spatial linkages are thus in close agreement with the actual origins of the overall virus transmission for the very early period of the global outbreak of the pandemic.

To showcase convergence of the posterior MCMC chains, Figure 3 depicts trace plots for ρ , σ^2 , and slope parameters. Overall, the trace plots show rather good mixing and convergence properties. Convergence of the chains have moreover been checked using the diagnostics proposed by Geweke (1992) implemented in the R package coda (Plummer et al., 2006). Results moreover appear rather robust concerning alternative modelling frameworks. Estimation results of these alternative specifications are presented in the Appendix in the supplemental data online.²⁵

7. CONCLUSIONS

In this paper we propose a Bayesian approach for estimation of weight matrices in spatial econometric models. A particular advantage of our approach is the simple integration into a standard Bayesian MCMC algorithm. The proposed framework can therefore be adapted and extended in a simple and computationally efficient way to cover a large number of alternative spatial specifications prevalent in recent literature. Our approach may thus be easily extended to cover inter alia non-Gaussian models such as spatial probit (LeSage et al., 2011) or logit specifications (Krisztin & Piribauer, 2021), local spillover models (Vega & Elhorst, 2015), or spatial error models (LeSage & Pace, 2009).

Our approach does not necessarily rely on specific prior information for the spatial linkages. Spatial information, however, can be easily implemented in a flexible and transparent way. We moreover motivate the use of hierarchical priors which impose sparsity in the resulting spatial weight matrix. These sparsity priors are particularly useful in applications where the number

of unknown parameters exceeds those of the observations. The virtues of our approach comes at the price that we focus on spatial neighbourhood structures which are binary (prior to row-standardization). However, this assumption is implicitly assumed in many spatial applications in the regional economic literature where spatial weight matrices are constructed based on concepts of contiguity, distance band or nearest neighbours.

Based on Monte Carlo simulations, we show that our approach appears particularly promising when the number of spatial observations N is large relative to the time dimension T , which is a rather common characteristic of data sets in the regional science literature. We moreover demonstrate the usefulness of our approach using real data on the outbreak of the COVID-19 pandemic. The results of this empirical application corroborate the findings in the Monte Carlo simulation study that the proposed approach performs well even in the cases of high over-parametrization.

ACKNOWLEDGMENTS

This article is based on an earlier working paper.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

FUNDING

Tamás Krisztin was supported by funds from the Oesterreichische Nationalbank (Austrian National Bank) [grant number 18690]; and Philipp Piribauer was supported by the Austrian Science Fund (FWF)[grant number ZK 35].

NOTES

¹ Ahrens and Bhattacharjee (2015) consider the case of sparsity in the spatial weights by employing shrinkage towards the zero matrix.

² We also consider specifications with a spatial lag of the temporally lagged dependent variable. Sampling strategies for these cases are presented in the Appendix in the supplemental data online.

³ For thorough discussions of the implications of row-standardization, see Plümper and Neumayer (2010) and Liu et al. (2014).

⁴ Equation (3) implies some observations may have zero neighbours. However, priors on the number of neighbours can be easily elicited to rule out such situations. Moreover, a researcher might easily abstain from row-standardization by neglecting the transformation in equation (3).

⁵ The function $f(\cdot)$ may simply be dropped when considering models without row-standardization of W .

⁶ The most obvious case where this assumption would be violated is a fully connected W with $w_{ij} = 1/N$ for all $i \neq j$.

⁷ These assumptions can be checked during estimation by using standard rejection sampling techniques in the MCMC sampling steps (e.g., LeSage & Pace, 2009; Koop, 2003). Rejection sampling rejects draws of parameter combinations that do not fulfil these assumptions.

⁸ To reduce the dimensionality of the parameter space, an interesting alternative might be the assumption of a symmetric Ω , which halves the number of free elements in the spatial weight matrix. This assumption can be imposed by simply simultaneously updating $\omega_{ij} = \omega_{ji}$, respectively.

⁹ A random walk Metropolis–Hastings step for ρ might be employed as an alternative.

¹⁰ Since the support for ρ is limited, the Griddy–Gibbs approach (or sometimes inversion approach) relies on univariate numerical integration techniques of the conditional posterior for ρ and uses the cumulative density function for producing draws of ρ . A Metropolis–Hastings step may be used as a standard alternative, but these typically produce less efficient draws with poorer mixing properties (see also LeSage & Pace, 2009).

¹¹ Note the implication that an update of ρ necessitates a direct evaluation of the determinant $|A|$ and the matrix inverse A^{-1} , as in this case no convenient equations exist. An update of ρ , however, has to be performed only once per Gibbs step, as opposed to the $N^2 - N$ updates necessary for Ω , thus justifying the relatively higher computational costs.

¹² More specifically, $\hat{\Omega} = (\hat{\Omega}_0 + \tilde{\Omega}_0)/2$, where $\hat{\Omega}_0$ is a $N/20$ nearest neighbour adjacency matrix.

¹³ However, a direct comparison of the results between symmetric and non-symmetric specifications does not appear reasonable, since the adjacency matrix in the data-generating process is assumed symmetric.

¹⁴ The number of unknown parameters amounts to $N^2 + T + q_0 + 2$ and $N(N - 1)/2 + N + T + q_0 + 2$ for non-symmetric and symmetric spatial weight matrices, respectively.

¹⁵ Figure A4 in the Appendix in the supplemental data online illustrates the convergence properties of a random Monte Carlo sample for the case of $N = 20$ and $T = 10$. This case was chosen because it is similar to the settings in the empirical applications.

¹⁶ Countries without any (official) infections in the starting period have been excluded from the sample. Moreover, we exclude India as a clear outlier from the sample due to its particular small (official) infection rates throughout the observation period.

¹⁷ With a biweekly time lag, the dependent variable thus captures data from 2 to 20 April ($T = 19$). For a better comparison, we have fixed the time period captured by y_t for all alternative specifications. Moreover, it is worth noting that a notable earlier starting date would result in relatively few (cross-sectional) observations. However, our results are rather robust when considering a longer time horizon.

¹⁸ The spatial growth regression in (14) may be alternatively specified in levels rather than in log-differences by setting $y_t = x_t$. Results using this alternative specification are very similar and are presented in the Appendix in the supplemental data online.

¹⁹ In the special case of $r > 0$, computational efficiency is tremendously increased, as no log-determinant calculations are required in the MCMC algorithm. The sampling strategy for these cases is presented in the Appendix in the supplemental data online.

²⁰ See <https://www.kaggle.com/datasets/vishalvoseph/weather-dataset-for-covid19-predictions/>.

²¹ As robustness checks, we also tried a shorter lag length of one week. The estimated spatial structures appeared very similar to the biweekly benchmarks. All these additional robustness checks, along with the R codes, are available from the authors upon request.

²² For the benchmark specifications, the number of unknown parameters and observations amounts to 753 and 513, respectively.

²³ The regional dependency structure appears particularly pronounced when a level specification of the infection dynamics is imposed. Sensitivity checks based on this alternative specification are presented in Figure A1 in the Appendix in the supplemental data online.

²⁴ When comparing the results, it is important to note that for all specifications under scrutiny, we have fixed the time period in the dependent variable (y_t from 2 to 20 February, i.e., $T = 19$). The biweekly temporal spatial lag specification thus inherently comprises spatial information prior to the period in y_t .

²⁵ Estimates when using a smaller time lag of seven days also appear very similar. Results along with the R codes used are available from the authors upon request.

ORCID

Tamás Krisztin  <http://orcid.org/0000-0002-9241-8628>

Philipp Piribauer  <http://orcid.org/0000-0002-3414-3101>

REFERENCES

- Ahrens, A., & Bhattacharjee, A. (2015). Two-step lasso estimation of the spatial weights matrix. *Econometrics*, 3(1), 128–155. doi:10.3390/econometrics3010128
- Basile, R. (2008). Regional economic growth in Europe: A semiparametric spatial dependence approach. *Papers in Regional Science*, 87(4), 527–544. doi:10.1111/j.1435-5957.2008.00175.x
- Briz-Redón, Á., & Serrano-Aroca, Á. (2020). The effect of climate on the spread of the COVID-19 pandemic: A review of findings, and statistical and modelling techniques. *Progress in Physical Geography: Earth and Environment*, 44(5), 591–604. doi:10.1177/0309133320946302
- Cornwall, G. J., & Parent, O. (2017). Embracing heterogeneity: The spatial autoregressive mixture model. *Regional Science and Urban Economics*, 64, 148–161. doi:10.1016/j.regsciurbeco.2017.03.004
- De Paula, Á., Rasul, I., & Souza, P. (2019). Identifying network ties from panel data: Theory and an application to tax competition. *arXiv preprint arXiv:1910.07452*.
- Debarsy, N., & LeSage, J. (2018). Flexible dependence modeling using convex combinations of different types of connectivity structures. *Regional Science and Urban Economics*, 69, 48–68. doi:10.1016/j.regsciurbeco.2018.01.001
- Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20(5), 533–534. doi:10.1016/S1473-3099(20)30120-1
- Elhorst, J. P. (2021). The dynamic general nesting spatial econometric model for spatial panels with common factors: Further raising the bar. *Review of Regional Research*, 1–19. doi:10.1007/s10037-021-00163-w
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (Eds.), *Bayesian Statistics* (Vol. 4, pp. 169–193). Oxford: Clarendon Press.
- Guliyev, H. (2020). Determining the spatial effects of COVID-19 using the spatial panel data model. *Spatial Statistics*, 38, 100443. doi:10.1016/j.spasta.2020.100443
- Hale, T., Petherick, A., Phillips, T., & Webster, S. (2020). Variation in government responses to COVID-19. *Blavatnik School of Government Working Paper*, 31, 2020–2011. doi:10.1038/s41562-021-01079-8
- Han, X., & Lee, L. F. (2016). Bayesian analysis of spatial panel autoregressive models with time-varying endogenous spatial weight matrices, common factors, and random coefficients. *Journal of Business & Economic Statistics*, 34(4), 642–660. doi:10.1080/07350015.2016.1167058
- Han, X., Xu, Y., Fan, L., Huang, Y., Xu, M., & Gao, S. (2021). Quantifying COVID-19 importation risk in a dynamic network of domestic cities and international countries. *Proceedings of the National Academy of Sciences*, 118, 31.
- Hsieh, C. S., & Lee, L. F. (2016). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics*, 31(2), 301–319. doi:10.1002/jae.2426
- Jaya, I. G. N. M., & Folmer, H. (2021). Bayesian spatiotemporal forecasting and mapping of COVID-19 risk with application to west Java province, Indonesia. *Journal of Regional Science*, 61(4), 849–881. doi:10.1111/jors.12533
- Kelejian, H. H., & Piras, G. (2014). Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. *Regional Science and Urban Economics*, 46, 140–149. doi:10.1016/j.regsciurbeco.2014.03.001
- Koop, G. (2003). *Bayesian econometrics*. John Wiley & Sons Ltd.

- Kosfeld, R., Mitze, T., Rode, J., & Wälde, K. (2021). The COVID-19 containment effects of public health measures: A spatial difference-in-differences approach. *Journal of Regional Science*, 61(4), 799–825. doi:10.1111/jors.12536
- Krisztin, T. (2017). The determinants of regional freight transport: A spatial, semiparametric approach. *Geographical Analysis*, 49(3), 268–308. doi:10.1111/gean.12125
- Krisztin, T., & Piribauer, P. (2021). A Bayesian spatial autoregressive logit model with an empirical application to European regional FDI flows. *Empirical Economics*, 61(1), 231–257. doi:10.1007/s00181-020-01856-w
- Krisztin, T., Piribauer, P., & Wögerer, M. (2020). The spatial econometrics of the coronavirus pandemic. *Letters in Spatial and Resource Sciences*, 13(3), 209–218. doi:10.1007/s12076-020-00254-1
- 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. doi:10.1080/07350015.2019.1569526
- LeSage, J. P. (1997). Bayesian estimation of spatial autoregressive models. *International Regional Science Review*, 20(1–2), 113–129. doi:10.1177/016001769702000107
- LeSage, J. P., Kelley Pace, R., Lam, N., Campanella, R., & Liu, X. (2011). New Orleans business recovery in the aftermath of Hurricane Katrina. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(4), 1007–1027. doi:10.1111/j.1467-985X.2011.00712.x
- LeSage, J. P., & Pace, R. K. (2007). A matrix exponential spatial specification. *Journal of Econometrics*, 140(1), 190–214. doi:10.1016/j.jeconom.2006.09.007
- LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. CRC Press.
- Ley, E., & Steel, M. F. (2009). On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics*, 24(4). doi:10.1002/jae.1057
- Liu, X., Patacchini, E., & Zenou, Y. (2014). Endogenous peer effects: Local aggregate or local average? *Journal of Economic Behavior & Organization*, 103, 39–59. doi:10.1016/j.jebo.2014.03.025
- Mitze, T., & Kosfeld, R. (2022). The propagation effect of commuting to work in the spatial transmission of COVID-19. *Journal of Geographical Systems*, 24(1), 5–31. doi:10.1007/s10109-021-00349-3
- Piribauer, P. (2016). Heterogeneity in spatial growth clusters. *Empirical Economics*, 51(2), 659–680. doi:10.1007/s00181-015-1023-y
- Piribauer, P., & Cuaresma, J. C. (2016). Bayesian variable selection in spatial autoregressive models. *Spatial Economic Analysis*, 11(4), 457–479. doi:10.1080/17421772.2016.1227468
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6(1), 7–11.
- Plümper, T., & Neumayer, E. (2010). Model specification in the analysis of spatial dependence. *European Journal of Political Research*, 49(3), 418–442. doi:10.1111/j.1475-6765.2009.01900.x
- Qu, X., & Lee, L. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, 184(2), 209–232. doi:10.1016/j.jeconom.2014.08.008
- Ritter, C., & Tanner, M. A. (1992). Facilitating the Gibbs sampler: The Gibbs stopper and the Griddy–Gibbs sampler. *Journal of the American Statistical Association*, 87(419), 861–868. doi:10.1080/01621459.1992.10475289
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x
- Vega, H. S., & Elhorst, J. P. (2015). The SLX model. *Journal of Regional Science*, 55(3), 339–363. doi:10.1111/jors.12188