Bayesian Variable Selection in Spatial Autoregressive Models

Crespo Cuaresma, Jesus; Piribauer, Philipp

DOI:
10.57938/aa3effd2-1312-4f27-bee3-5e8e74f45f99

Published: 01/07/2015

Document Version:
Publisher's PDF, also known as Version of record

Document License:
Unspecified

Department of Economics
Working Paper No. 199

Bayesian Variable Selection in Spatial Autoregressive Models

Philipp Piribauer
Jesus Crespo Cuaresma

July 2015

Bayesian variable selection in spatial autoregressive models*

Philipp Piribauer†

Jesús Crespo Cuaresma‡

Abstract

This paper compares the performance of Bayesian variable selection approaches for spatial autoregressive models. We present two alternative approaches which can be implemented using Gibbs sampling methods in a straightforward way and allow us to deal with the problem of model uncertainty in spatial autoregressive models in a flexible and computationally efficient way. In a simulation study we show that the variable selection approaches tend to outperform existing Bayesian model averaging techniques both in terms of in-sample predictive performance and computational efficiency.

Keywords: spatial autoregressive model, variable selection, model uncertainty, Markov chain Monte Carlo methods

JEL Codes: C18, C21, C52

This version: July 2015

*The research carried out in this paper was supported by funds of the Oesterreichische Nationalbank (Oesterreichische Nationalbank, Anniversary Fund, project number: 16244).

†Corresponding author: Philipp Piribauer, Vienna University of Economics and Business (WU), Welthandelsplatz 1, 1090 Vienna, Austria. E-mail: philipp.piribauer@wu.ac.at.

‡Vienna University of Economics and Business (WU), Wittgenstein Centre for Demography and Global Human Capital (WIC), International Institute for Applied Systems Analysis (IIASA) and Austrian Institute of Economic Research (WIFO).

1 Introduction

The rigorous treatment of model uncertainty in models for spatially correlated data has recently received considerable attention in the econometric literature. Bayesian model averaging techniques have gained momentum in empirical research for regional economic growth (see, for example, LeSage and Fischer 2008, Crespo Cuaresma and Feldkircher 2013, Crespo Cuaresma et al. 2014 and Piribauer and Fischer 2014), where often alternative theoretical frameworks are not necessarily mutually exclusive (Raftery 1995). This problem, which is usually referred to as the *open-endedness* of growth theories (Brock and Durlauf 2001), implies that inference based on individual specifications may overestimate the precision of the estimates obtained. Bayesian model averaging techniques have thus been employed to deal with the problem of model uncertainty by producing parameter inference where model specification uncertainty is integrated out.

Standard Bayesian model averaging methods involve the computation of marginal likelihoods, which do not have closed form solutions for spatial autoregressive (SAR) models (see LeSage and Parent 2007). If many covariates are potential candidates to be part of the specification, model averaging of SAR models thus leads to a severe computational burden, since the calculation of the corresponding marginal likelihoods require numerical integration techniques. Crespo Cuaresma and Feldkircher (2013) propose the use of spatial eigenvector filtering techniques (see, for example, Griffith 2000a and Griffith 2000b) in order to reduce the models to linear representations, albeit losing information about the spatial autoregressive parameter. Piribauer and Fischer (2014), on the other hand, propose the use of posterior model weights based on the Bayesian information criterion and maximum likelihood estimates of the matrix exponential specification (see LeSage 2007) of global spatial spillover effects.

Recent developments in the spatial econometric literature focus on more flexible model specifications as those entertained by Crespo Cuaresma and Feldkircher (2013) or Piribauer and Fischer (2014). Vega and Elhorst (2013), for example, allow for parametrized spatial weight matrices, while other contributions include multiple weight matrices (see Elhorst et al. 2012). Han and Lee (2013) put forward a Bayesian approach that integrates varying orders of neighbors in order to account for spatial linkages across units. This approach relies on smoothness priors (Shiller 1973) to impose a stochastic structure of decay on higher-order neighbors. Apart from these various ways to specify spatial spillovers, a large strand of literature considers simultaneous space-time panel data models (see, for example, Debarsy et al. 2012 and Parent and LeSage 2012). Since Bayesian model averaging techniques for spatial autoregressive models rely on the calculation of marginal likelihoods, the consideration of model uncertainty in the context of such flexible model specifications is barely feasible.

In this paper, we introduce two Bayesian model selection approaches for spatial econometric model specifications. The first approach uses stochastic search variable selection (SSVS) priors put forward by George and McCulloch (1993) and George and McCulloch (1997), which have been widely applied especially in the time-series literature (see George et al. 2008 and Koop and Korobilis 2010). SSVS priors rely on a

mixture of normal priors, where the choice of the prior hyperparameters may have severe effects on posterior inference. A second approach, based on work by Kuo and Mallick (1998), overcomes this problem through a simple reparametrization of the model. Both approaches can be implemented using a Gibbs sampling algorithm in a straightforward way. This feature thus allows for a computationally efficient treatment of model uncertainty even in very flexible spatial modelling frameworks.

In a simulation study, we compare the performance of both variable selection approaches proposed with Bayesian model averaging techniques advocated by LeSage and Parent (2007) both in terms of in-sample predictive performance as well as computational efficiency. The simulation study shows that the in-sample performance of both variable selection approaches outperforms Bayesian model averaging, particularly for large sample sizes. Moreover, the Monte Carlo experiments show that the proposed approaches present a clear advantage with respect to Bayesian model averaging for spatial autoregressive models in terms of computational efficiency. This is largely attributable to the fact that these techniques do not require the computation of marginal likelihoods. Furthermore, both of the variable selection methods proposed in this paper can be implemented in a straightforward way using Gibbs sampling. This allows for an efficient assessment of the problem of model uncertainty in flexible model specifications such as dynamic spatial panels or spatial models involving multiple or parameterized spatial weight matrices.

The remainder of the paper is organized as follows. Section 2 briefly outlines the Bayesian model averaging methodology for spatial autoregressive models introduced by LeSage and Parent (2007). Section 3 and 4 presents the model selection approaches along with their implementation in a Markov chain Monte Carlo (MCMC) sampler. Section 5 presents Monte Carlo simulation results to investigate the performance of our model selection approaches compared to standard Bayesian model averaging for spatial autoregressive models. Section 6 concludes.

2 Model uncertainty in spatial autoregressive models

Consider a SAR model specification of the form[1]

$$\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{\iota}_n \alpha + \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{2.1}$$

where $\boldsymbol{y}$ denotes an $n \times 1$ vector of observations on the dependent variable, $\boldsymbol{\iota}_n$ is an $n \times 1$ vector of ones with a corresponding scalar parameter α and $\boldsymbol{X}$ is a $n \times q$ matrix of explanatory variables, with $\boldsymbol{\theta}$ being the corresponding $q \times 1$ vector of parameters. The n-element vector $\boldsymbol{\varepsilon}$ is assumed to be distributed as $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega})$, with $\boldsymbol{\Omega} = \sigma^2 \boldsymbol{\Sigma}$. Without

[1] We are aware of the fact that the existing literature in spatial econometrics increasingly relies on spatial econometric specifications other than the SAR. A popular extension widely applied in empirical practice is the spatial Durbin model (SDM). It is, however, worth noting that a SDM specification can easily be formulated by simply adding $\boldsymbol{WX}$ as extra regressors and all the forthcoming formulae also apply for the SDM (see LeSage and Pace 2009).

loss of generality, we may assume $\boldsymbol{\Sigma} = \boldsymbol{I}_n$.[2] $\boldsymbol{W}$ denotes an $n \times n$ row-stochastic and non-negative spatial weight matrix which describes the linkages between the observations in the sample. The entries in $\boldsymbol{W}$ are assumed given. Specifically, $[W]_{ij} > 0$ for $i \neq j$ if observation i and j are considered neighbors. Conventionally, $[W]_{ii} = 0$ for all $i = 1, \ldots, n$, since no observation is considered as being a neighbor to itself. The scalar spatial parameter $\rho \in (-1, 1)$ reflects the magnitude of spatial dependence among the observations.

We consider model uncertainty in terms of carrying out posterior inference on the parameters of equation (2.1) in presence of a potentially large number of alternative combination of variables in $\boldsymbol{X}$.

2.1 Bayesian model averaging

Bayesian model averaging approaches propose to carry out inference using weighted averaged posterior distributions across alternative models in the model space $\mathcal{M}$. $\mathcal{M}$ captures models formed by all subsets of potential covariates in $\boldsymbol{X}$ and alternative spatial weight matrices. Without loss of generality, for the exposition we consider only covariate uncertainty within the class of linear specifications, that is, we fix $\boldsymbol{W}$ and assume that q potential (non-constant) explanatory variables are available. The cardinality of $\mathcal{M}$ in our case is thus 2^q, the number of combinations of q variables in groups of sizes 1 to q.

Characterizing a particular model $M_m \in \mathcal{M}$ for $m = 1, \ldots, 2^q$ by its parameter vector $\boldsymbol{\psi} = \begin{bmatrix} \alpha & \boldsymbol{\theta}' & \rho & \sigma^2 \end{bmatrix}'$, the posterior density for $\boldsymbol{\psi}$ conditional on model M_m, is given by

$$p(\boldsymbol{\psi}|M_m, \mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{\psi}, M_m)p(\boldsymbol{\psi}|M_m)}{p(\mathcal{D}|M_m)}, \tag{2.2}$$

where $p(\mathcal{D}|\boldsymbol{\psi}, M_m)$ is the likelihood, $p(\boldsymbol{\psi}|M_m)$ denotes the prior for $\boldsymbol{\psi}$ and $\mathcal{D}$ denotes the data. Non-informative prior specifications on α and σ^2 are often used in the literature, since these parameters are common to all models in $\mathcal{M}$. For the slope parameters, $\boldsymbol{\theta}$, g-priors (Zellner 1986) are often used, since they lead to convenient analytical posteriors in the case of the normal linear regression model. Fernández et al. (2001) propose the use of the unit information prior or the risk inflation criterion (two popular fixed g-priors), depending on the size of the sample as compared to the number of potential regressors. Liang et al. (2008) and Feldkircher and Zeugner (2009), on the other hand, propose the use of a hierarchical prior on the slope parameters, governed by a prior on the shrinkage factor. For the spatial autoregressive parameter ρ, uniform or beta priors over the interval [-1,1] are standard choices (LeSage and Pace 2009).

[2]It is worth noting that there is also a strand of literature in spatial econometrics which emphasizes the importance of spatial dependence in the error term. Two popular specifications include the spatial error (SEM) and spatial Durbin error (SDEM) specification (see LeSage and Pace 2009). In order to maintain succinct notation, the focus of this paper lies on spatial lag dependence by setting $\boldsymbol{\Sigma} = \boldsymbol{I}_n$. However, the notation in the paper has been consciously chosen to be very general, such that the proposed approaches can also be implemented in spatial error specifications.

Since Bayesian model averaging treats the candidate models M_m as random, the posterior model probability $p(M_m|\mathcal{D})$ can be expressed as

$$p(M_m|\mathcal{D}) = \frac{p(\mathcal{D}|M_m)p(M_m)}{p(\mathcal{D})} \propto p(\mathcal{D}|M_m)p(M_m). \tag{2.3}$$

Like for any other parameter, a prior distribution across specifications, M_m, has to be elicited. A natural choice would be to use a uniform prior such that $p(M_m) = 2^{-q}$, which treats each candidate model as equally likely a priori. Alternative choices for $p(M_m)$ which allow for more flexible prior settings across model sizes are discussed in Ley and Steel (2009).

The marginal likelihood, $p(\mathcal{D}|M_m)$, in turn, is given by

$$p(\mathcal{D}|M_m) = \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty p(\mathcal{D}|\boldsymbol{\psi}, M_m)p(\boldsymbol{\psi}|M_m)\ \mathrm{d}\alpha\ \mathrm{d}\boldsymbol{\theta}\ \mathrm{d}\rho\ \mathrm{d}\sigma^2. \tag{2.4}$$

The use of g-priors allows for analytical integration of the non-spatial parameters (that is α, $\boldsymbol{\theta}$, and σ^2) in equation (2.4). However, calculating the marginal likelihood for SAR specifications involves numerical integration techniques (LeSage and Parent, 2007). The lack of analytical tractability of equation (2.4) for spatial autoregressive models implies that Bayesian model averaging of SAR specifications might lead to a severe computational burden when the model space is large and/or more flexible model specifications are considered.

Using the law of total probability, Bayesian model averaging involves producing weighted averaged parameter estimates using the posterior model probabilities $p(M_m|\mathcal{D})$ as weights. The posterior density of the parameters is thus

$$p(\boldsymbol{\psi}|\mathcal{D}) = \sum_{m=1}^{2^q} p(M_m|\mathcal{D})p(\boldsymbol{\psi}|M_m, \mathcal{D}). \tag{2.5}$$

For large q, the number of terms in equation (2.5) can be enormous, thus rendering exhaustive summation infeasible. Markov chain Monte Carlo model composition (MC^3) techniques can be used to approximate the sum in equation (2.5) (see Madigan et al. 1995). Detailed derivations of all the posterior quantities in the context of SAR specifications can be found in LeSage and Parent (2007).

2.2 Stochastic search variable selection in SAR models

George and McCulloch (1993) and George and McCulloch (1997) put forward stochastic search variable selection (SSVS) priors for linear regression models. In this section, we generalize them to SAR models. This model selection approach has the advantage that no marginal likelihood computations are needed and posterior inference can be carried out in a flexible and efficient way using Gibbs sampling.

Based on a standard SAR model specification as in equation (2.1), priors on the regression parameters $\boldsymbol{\zeta} = \begin{bmatrix}\alpha & \boldsymbol{\theta}'\end{bmatrix}'$, σ^2 and ρ have to be elicited. The SSVS prior treats

each element in $\boldsymbol{\zeta}$, ζ_l for $l = 1, \ldots, q+1$ as a realization of a mixture of to normal distribution. The prior distribution for ζ_i is given by

$$p(\zeta_l|\delta_l) \sim \delta_l \mathcal{N}(0, \underline{s}_{1,l}^2) + (1 - \delta_l)\mathcal{N}(0, \underline{s}_{0,l}^2), \tag{2.6}$$

where $\boldsymbol{\delta} = \begin{bmatrix}\delta_1 & \ldots & \delta_{q+1}\end{bmatrix}'$ denotes a vector of latent binary parameters corresponding to $\boldsymbol{\zeta}$. $\underline{\boldsymbol{s}}_0 = \begin{bmatrix}\underline{s}_{0,1} & \ldots & \underline{s}_{0,q+1}\end{bmatrix}'$ and $\underline{\boldsymbol{s}}_1 = \begin{bmatrix}\underline{s}_{1,1} & \ldots & \underline{s}_{1,q+1}\end{bmatrix}'$ denote prior hyperparameters for the variance of $\boldsymbol{\zeta}$.

By setting $\underline{s}_{1,l} \gg \underline{s}_{0,l}$ $\forall$ l, the latent binary variables δ_l govern which one of the normals in equation (2.6) apply. If $\delta_l = 1$, the corresponding covariate is considered as exerting potentially significant effects on the dependent variable $\boldsymbol{y}$. In this case, the prior variance $\underline{s}_{1,l}$ is large, such that the prior only exerts little influence on the posterior. In the case of $\delta_l = 0$, due to the very small variance $\underline{s}_{0,l}$, the respective parameter is shrunk towards zero. Hence, δ_l governs whether the lth covariate is considered as being included or excluded in the model. Similar to Bayesian model averaging techniques, a posterior inclusion probability for ζ_l can be calculated by simply evaluating the posterior probability of $\delta_l = 1$. Since the SSVS prior treats δ_l as random, equation (2.6) formulates a hierarchical prior specification over the corresponding parameters. The prior $p(\boldsymbol{\delta})$ incorporates prior information on the relative importance of candidate subsets of regressors. Following George and McCulloch (1997), we set

$$p(\boldsymbol{\delta}) = \prod_{l=1}^{q+1} \underline{g}_l^{\delta_l}(1 - \underline{g}_l)^{(1-\delta_l)}, \tag{2.7}$$

where the scalar prior hyperparameter $\underline{g}_l = \Pr(\delta_l = 1) = 1 - \Pr(\delta_l = 0)$ is the prior probability of inclusion of the lth explanatory variable. We assume that each model has an intercept term included by setting $\underline{g}_1$ equal to unity. In parallel to the assumption of a flat prior over model specifications, a natural choice for the prior inclusion of individual covariates is $\underline{g}_2 = \cdots = \underline{g}_{q+1} = \frac{1}{2}$. Alternative prior specifications of $\boldsymbol{\delta}$ are thoroughly discussed in George and McCulloch (1993).

George and McCulloch (1993) propose a semi-automatic approach for the elicitation of $\underline{\boldsymbol{s}}_0$ and $\underline{\boldsymbol{s}}_1$ which uses the estimated standard deviations $\hat{\boldsymbol{s}}$ obtained using maximum likelihood estimation. With $\underline{c_0} \ll \underline{c_1}$ denoting scalar scaling constants, $\underline{s}_{0,l}$ and $\underline{s}_{1,l}$ are simply set equal to $\underline{c_0}\hat{s}_l$ and $\underline{c_1}\hat{s}_l$, respectively.

For σ^2, an inverted gamma prior is used,

$$p(\sigma^2) \sim \mathcal{IG}(\underline{a}, \underline{b}) \tag{2.8}$$

with $\underline{a}$ and $\underline{b}$ denoting scalar prior hyperparameters. For the spatial autocorrelation parameter ρ, a uniform prior distribution may be used, $\rho \sim \mathcal{U}(-1, 1)$.

With the semi-conjugate prior setup given by equations (2.6) and (2.8), $\boldsymbol{\zeta}$ and σ^2 can be directly sampled using Gibbs sampling steps (see also LeSage 1997 and LeSage and Pace 2009). The conditional posterior distribution of σ^2 is given by

$$
\begin{aligned}
p(\sigma^2|\boldsymbol{\zeta},\boldsymbol{\delta},\rho,\mathcal{D}) &\sim \mathcal{IG}(\overline{a},\overline{b}) && (2.9)\\
\overline{a} &= \underline{a}+n/2 && (2.10)\\
\overline{b} &= \underline{b}+(\boldsymbol{A}\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{\zeta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{\zeta})/2 && (2.11)
\end{aligned}
$$

with $\boldsymbol{Z}=\begin{bmatrix}\boldsymbol{\iota}_n & \boldsymbol{X}\end{bmatrix}$ and $\boldsymbol{A}=\boldsymbol{I}_n-\rho\boldsymbol{W}$.

Following George and McCulloch (1993) and George and McCulloch (1997) the conditional posterior for the binary vector $\boldsymbol{\delta}$ is Bernoulli distributed,

$$
\begin{aligned}
\Pr(\delta_l=1|\zeta_l,\sigma^2,\boldsymbol{\delta}_{l\cdot},\rho) &\sim \text{Ber}\left(\frac{\overline{u}_{1,l}}{\overline{u}_{0,l}+\overline{u}_{1,l}}\right) && (2.12)\\
\overline{u}_{1,l} &= \frac{1}{\underline{s}_{1,l}}\exp\left(-\frac{1}{2}\frac{\zeta_l^2}{\underline{s}_{1,l}^2}\right)\underline{g}_l && (2.13)\\
\overline{u}_{0,l} &= \frac{1}{\underline{s}_{0,l}}\exp\left(-\frac{1}{2}\frac{\zeta_l^2}{\underline{s}_{0,l}^2}\right)(1-\underline{g}_l) && (2.14)
\end{aligned}
$$

where $\boldsymbol{\delta}_{l\cdot}$ denotes all elements in $\boldsymbol{\delta}$ other than δ_l.

Conditional on the other parameters, the posterior distribution for $\boldsymbol{\zeta}$ has the standard form (see Koop 2003)

$$
\begin{aligned}
p(\boldsymbol{\zeta}|\sigma^2,\boldsymbol{\delta},\rho,\mathcal{D}) &\sim \mathcal{N}(\overline{\boldsymbol{\zeta}},\overline{\boldsymbol{V}}) && (2.15)\\
\overline{\boldsymbol{\zeta}} &= \overline{\boldsymbol{V}}(\underline{\boldsymbol{V}}^{-1}\underline{\boldsymbol{\zeta}}+\boldsymbol{Z}'\boldsymbol{\Omega}^{-1}\boldsymbol{A}\boldsymbol{y}) && (2.16)\\
\overline{\boldsymbol{V}} &= (\underline{\boldsymbol{V}}^{-1}+\boldsymbol{Z}'\boldsymbol{\Omega}^{-1}\boldsymbol{Z})^{-1} && (2.17)
\end{aligned}
$$

where $\underline{\boldsymbol{\zeta}}=\boldsymbol{0}$ and $\underline{\boldsymbol{V}}$ denotes the diagonal prior variance matrix implicitly defined through the SSVS prior in equation (2.6).

The conditional posterior for the spatial autoregressive parameter ρ, however, does not correspond to a known distribution (see LeSage and Pace 2009) and is given by

$$
p(\rho|\boldsymbol{\zeta},\sigma^2,\boldsymbol{\delta},\mathcal{D})\propto|\boldsymbol{A}|\exp\Big(-\frac{1}{2\sigma^2}(\boldsymbol{A}\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{\zeta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{\zeta})\Big)p(\rho). \qquad (2.18)
$$

Sampling from the conditional posterior for ρ is amenable for a griddy Gibbs step, which involves univariate numerical integration to sample from the conditional posterior distribution (see, for example, Koch and LeSage 2015).[3] With conditional posteriors for σ^2, $\boldsymbol{\delta}$, $\boldsymbol{\zeta}$ and ρ given by equations (2.9), (2.12), (2.15) and (2.18), respectively, Markov chain Monte Carlo estimation of the model outlined above involves the following steps:

[3]Alternatively, a random-walk Metropolis-Hastings step could be implemented to draw ρ (see LeSage and Pace 2009).

1. Initiate values for the parameters to be estimated.

2. Given the most recent iteration, draw $\boldsymbol{\zeta}$ from (2.15), σ^2 from (2.9), $\boldsymbol{\delta}$ from (2.12) and ρ from (2.18).

The second is repeated for a sufficiently large number of iterations. The corresponding draws are stored after discarding a number of burn-in draws (see, for example, Koop 2003).

2.3 An alternative SSVS setting for SAR models

An alternative variable selection approach to that presented above that avoids having to elicit the mixture of normals in equation (2.6) would replace the second (low-variance) normal by a point mass at zero. In this case, however, a Markov chain Monte Carlo sampling algorithm would break down due to the zero prior variances (see George and McCulloch 1993 and Kuo and Mallick 1998). To avoid this pitfall, an expanded SAR model can be formulated by defining

$$\begin{aligned} \boldsymbol{y} &= \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{Z}\boldsymbol{\zeta} + \boldsymbol{\varepsilon} && (2.19) \\ \boldsymbol{\zeta} &= \begin{bmatrix} \beta_1\gamma_1 & \dots & \beta_{q+1}\gamma_{q+1} \end{bmatrix}', && (2.20) \end{aligned}$$

where all the variables are defined as before, with γ_j playing the role of δ_j above. The expanded SAR model treats the slope parameters $\boldsymbol{\zeta}$ as the element-wise product of the $(q+1)$-dimensional vector $\boldsymbol{\beta}$ and the vector of binary variables given by $\boldsymbol{\gamma}$.

Priors for ρ and σ^2 are elicited as above, $p(\rho) \sim \mathcal{U}(-1, 1)$ and $p(\sigma^2) \sim \mathcal{IG}(\underline{a}, \underline{b})$. For $\boldsymbol{\beta}$ a normal prior is used,

$$\boldsymbol{\beta} \sim \mathcal{N}(\underline{\boldsymbol{\beta}}, \underline{\boldsymbol{V}}_\beta) \tag{2.21}$$

where $\underline{\boldsymbol{\beta}} = \mathbf{0}$ and $\underline{\boldsymbol{V}}_\beta$ denote the prior mean and variance, respectively.

For the binary indicators γ_l, we elicit independent Bernoulli priors,

$$p(\gamma_l) \sim \text{Ber}(\underline{p}_l). \tag{2.22}$$

The prior hyperparameters $\underline{p}_l$ specify the prior probability of inclusion of the corresponding covariates. The constant term may again be forced to enter each of the considered subsets by setting $\underline{p}_1 = 1$. In order to obtain a uniform prior for the candidate models, prior inclusion probabilities would be elicited as $\underline{p}_2 = \dots = \underline{p}_{q+1} = \frac{1}{2}$.

After setting the priors for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, σ^2 and ρ, a Markov chain Monte Carlo sampler can be set up to sequentially sample from the conditional posterior distributions. It is worth noting that the conditional posteriors for σ^2 and ρ are exactly the same than in the standard SSVS framework (see equations (2.9) and (2.18), respectively). The conditional posterior distribution for $\boldsymbol{\beta}$ takes the form

$$
\begin{aligned}
p(\boldsymbol{\beta}|\boldsymbol{\gamma},\sigma^2,\rho,\mathcal{D}) &\sim \mathcal{N}(\overline{\boldsymbol{\beta}},\overline{\boldsymbol{V}}_\beta) && (2.23)\\
\overline{\boldsymbol{\beta}} &= \overline{\boldsymbol{V}}_\beta(\underline{\boldsymbol{V}}_\beta^{-1}\underline{\boldsymbol{\beta}}+\tilde{\boldsymbol{Z}}'\boldsymbol{\Omega}^{-1}\boldsymbol{A}\boldsymbol{y}) && (2.24)\\
\overline{\boldsymbol{V}}_\beta &= (\underline{\boldsymbol{V}}_\beta^{-1}+\tilde{\boldsymbol{Z}}'\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{Z}})^{-1} && (2.25)\\
\tilde{\boldsymbol{Z}} &= \begin{bmatrix}\gamma_1\boldsymbol{z}_1 & \dots & \gamma_{q+1}\boldsymbol{z}_{q+1}\end{bmatrix} && (2.26)
\end{aligned}
$$

where $\boldsymbol{z}_l$ denotes the lth column vector of matrix $\boldsymbol{Z}$.

The conditional posterior distribution of γ_l is Bernoulli,

$$
\begin{aligned}
p(\gamma_l|\boldsymbol{\beta},\boldsymbol{\gamma}_{l\cdot},\sigma^2,\rho) &\sim \text{Ber}\left(\frac{\overline{p}_{1,l}}{\overline{p}_{1,l}+\overline{p}_{0,l}}\right) && (2.27)\\
\overline{p}_{1,l} &= \underline{p}_l \exp\Big(-\frac{1}{2\sigma^2}(\boldsymbol{A}\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{\zeta}_l^{*})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{\zeta}_l^{*})\Big) && (2.28)\\
\overline{p}_{0,l} &= (1-\underline{p}_l)\exp\Big(-\frac{1}{2\sigma^2}(\boldsymbol{A}\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{\zeta}_l^{**})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{\zeta}_l^{**})\Big) && (2.29)
\end{aligned}
$$

where $\boldsymbol{\zeta}_l^{*}$ and $\boldsymbol{\zeta}_l^{**}$ denote the column vector $\boldsymbol{\zeta}$ with the lth element replaced by β_l and zero, respectively.

3 Model uncertainty in SAR specifications: A simulation study

This section presents the results of Monte Carlo experiments that investigate the performance of the model averaging and variable selection approaches outlined above. In the Monte Carlo experiments data on the potential explanatory variables $\hat{\boldsymbol{Z}} = \begin{bmatrix}\boldsymbol{\iota}_{\hat{n}} & \hat{\boldsymbol{z}}_1 & \dots & \hat{\boldsymbol{z}}_9\end{bmatrix}$ are generated from a standard normal distribution for each of the (non-constant) regressors. In the simulation study, $\hat{\boldsymbol{Z}}$ is of dimension $\hat{n}\times 10$. We consider sample sizes $\hat{n}\in\{100,1000\}$.

The dependent variable $\hat{\boldsymbol{y}}$ is generated according to the following data generating process

$$
\hat{\boldsymbol{y}} = \hat{\boldsymbol{A}}^{-1}(\hat{\boldsymbol{Z}}\hat{\boldsymbol{\zeta}}+\hat{\boldsymbol{\varepsilon}}), \tag{3.1}
$$

where $\hat{\boldsymbol{A}}^{-1}=(\boldsymbol{I}_{\hat{n}}-\hat{\rho}\hat{\boldsymbol{W}})^{-1}$ and $\hat{\boldsymbol{\varepsilon}}\sim\mathcal{N}(\boldsymbol{0},\hat{\sigma}^2\boldsymbol{I}_{\hat{n}})$. The $\hat{n}\times\hat{n}$ spatial weight matrix $\hat{\boldsymbol{W}}$ is treated as fixed using a seven nearest neighbor specification based on standard uniform draws of longitude and latitude values.[4] We consider settings based on both moderate and high levels of spatial autocorrelation by setting $\rho\in\{0.5,0.9\}$.

The true slope coefficients $\hat{\boldsymbol{\zeta}}$ in the data generating process given by equation (3.1) are chosen in a way such that the constant term and the first three (non-constant)

[4]Different nearest neighbors specifications yield qualitatively similar results. Results for alternative specifications of $\hat{\boldsymbol{W}}$ are available upon request.

covariates enter the regression with non-zero parameters. Specifically, we set 0.5, 0.3, 1.0, and -0.9 for $\boldsymbol{\iota}_{\hat{n}}$, $\hat{\boldsymbol{z}}_1$, $\hat{\boldsymbol{z}}_2$ and $\hat{\boldsymbol{z}}_3$, respectively. For the remaining six candidate covariates in $\hat{\boldsymbol{Z}}$, we set the corresponding slope parameters equal to zero. In the Monte Carlo experiments we set $\hat{\sigma}^2$ equal to the level required to yield specific signal-to-noise ratios (SNR), defined as (see Pace et al. 2011),

$$SNR = 1 - \frac{E(\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{A}}^{-2} \hat{\boldsymbol{\varepsilon}})}{\hat{\boldsymbol{\zeta}}' \hat{\boldsymbol{Z}}' \hat{\boldsymbol{A}}^{-2} \hat{\boldsymbol{Z}} \hat{\boldsymbol{\zeta}} + E(\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{A}}^{-2} \hat{\boldsymbol{\varepsilon}})}, \tag{3.2}$$

where $E(\cdot)$ denotes the expectation operator.

Prior implementation

For all selection and averaging techniques a uniform prior for ρ has been used, $p(\rho) \sim \mathcal{U}(-1, 1)$. The intercept term has been assumed to be included in all specifications. For all the three approaches under scrutiny, a uniform prior over the space of specifications has been considered by setting $p(M_m) = 2^{-q}$. In the Bayesian model averaging framework (henceforth BMA SAR) a BRIC-specification for the prior over the slope coefficients was used (see Fernández et al. 2001), along with a non-informative prior for the disturbance variance.

In the SSVS prior setting for the scaling constants $\underline{c}_0$ and $\underline{c}_1$, the values 1/100 and 100 have been chosen.[5] A very non-informative setting was elicited for the disturbance variance by setting $\underline{a} = \underline{b} = 0.001$. In order to mirror the uniform model prior in the Bayesian model averaging setup, we set $\underline{g}_2 = \cdots = \underline{g}_{q+1} = 1/2$. Moreover, $\underline{g}_1$ is set to unity in order to force the constant to be included in each potential specification.

Employing the same arguments as in the SSVS prior setting, for the expanded SAR approach we set $\underline{p}_2 = \cdots = \underline{p}_{q+1} = 1/2$ along with $\underline{p}_1 = 1$. Similarly, $\underline{a} = \underline{b} = 0.001$. Although this framework does not require priors for scaling constants, a prior for $\boldsymbol{\beta}$ has to be elicited. To ensure that the prior for the slope coefficients barely influences the outcomes, we use a rather diffuse setup for the prior variance of this parameter vector, $\underline{V}_\beta = 1000 \boldsymbol{I}_{q+1}$.

Simulation results

Posterior inference in the simulation exercise is based on 500 posterior draws, with the first 300 discarded as burn-ins. Tables 1 and 2 report the results of the Monte Carlo experiments, relating to averages over 1,000 simulated datasets for a signal-to-noise ratio of 0.2 and 0.9, respectively. True values used to generate the data are reported in the first column. Both tables report the simulation results for the three

[5] Alternative popular settings for $\underline{c}_0$ and $\underline{c}_1$, such as 1/10 and 10, or 1/5 and 5, respectively, have also been tested. The results remained qualitatively robust. It is however worth noting that a prior setup with a smaller gap between $\underline{c}_0$ and $\underline{c}_1$ resulted in larger root mean squared errors, in particular for the slope coefficients associated with the covariates not included in the data generating process.

approaches under consideration in the columns that follow: first, the approach based on the expanded spatial autoregressive model ('expanded SAR'), second, the spatial autoregressive model based on stochastic search variable selection priors ('SAR SSVS'), and finally, Bayesian model averaging of spatial autoregressive models ('SAR BMA').

For each approach, the tables report the average posterior inclusion probability ('PIP'), a measure of the importance of the variables. In Bayesian model averaging frameworks, the posterior inclusion probability of a particular explanatory variable is calculated as the sum of posterior model probabilities for models including that variable. For the expanded SAR and SSVS SAR approach, posterior inclusion probabilities are given by $\Pr(\delta_l = 1|\mathcal{D})$ and $\Pr(\gamma_l = 1|\mathcal{D})$, respectively. The tables also report the root mean squared error ('RMSE') for each potential covariate along with average timing results.

Both tables clearly show that the expanded SAR approach produces the lowest average root mean squared errors for explanatory variables not entering the data generating process. This finding generally holds true for different n, ρ, as well as signal-to-noise ratios. The outperformance over the two alternative approaches appears to be particularly large for a moderate degree of spatial autocorrelation. In these settings the expanded SAR approach is superior to Bayesian model averaging in terms of minimizing the average root mean squared errors of non-important covariates by a factor of three or more. Comparing RMSEs for the non-important variables between SAR SSVS and SAR BMA, no clear pattern can be seen in small sample sizes. However, for large sample sizes, Bayesian model averaging generally outperforms SAR SSVS.

In terms of posterior inclusion probabilities, the expanded SAR approach also appears to detect unimportant covariates better than the SAR SSVS or SAR BMA. Only in the scenario with the small sample size ($n = 100$) and low signal-to-noise ratio ($SNR = 0.2$), in this respect SAR SVSS occasionally outperforms the expanded SAR approach. For variables not entering the true data generating process, the general outperformance over SAR SSVS might be attributable to fact that the expanded SAR approach uses a prior point mass at zero for non-important covariates, thus resulting in smaller posterior inclusion probabilities. The performance in terms of posterior inclusion probabilities for these variables between SSVS and Bayesian model averaging appears to be very similar in large sample sizes. However, when considering small samples, Bayesian model averaging performs worst by producing the largest inclusion probabilities.

Turning attention to variables entering the true data generating process, the tables show that the SAR BMA approach advocated by LeSage and Parent (2007) performs particularly well in terms of minimizing the root mean squared errors for these variables. These findings become particularly apparent for small sample sizes. This outperformance, however, vanishes in large sample sizes. In large samples, the three approaches under scrutiny produce very similar in-sample predictive performances. The same results can be observed when comparing posterior inclusion probabilities. The Bayesian model averaging approach slightly, however, generally outperforms the alternative ap-

Table 1: Results of Monte Carlo experiments for $SNR = 0.2$, averaged over 1,000 simulation runs

$SNR = 0.2$														
			$\rho = 0.5$						$\rho = 0.9$					
			expanded SAR		SAR SSVS		SAR BMA		expanded SAR		SAR SSVS		SAR BMA	
	Variable	true	PIP	RMSE	PIP	RMSE	PIP	RMSE	PIP	RMSE	PIP	RMSE	PIP	RMSE
$n = 100$	$\boldsymbol{\iota}_n$	0.500	1.000	0.198	1.000	0.203	1.000	0.195	1.000	0.416	1.000	0.417	1.000	0.392
	$\boldsymbol{x}_1$	0.300	0.251	0.248	0.417	0.251	0.500	0.214	0.128	0.263	0.186	0.285	0.302	0.238
	$\boldsymbol{x}_2$	1.000	0.996	0.155	1.000	0.147	1.000	0.146	0.885	0.347	0.913	0.360	0.964	0.267
	$\boldsymbol{x}_3$	-0.900	0.999	0.143	1.000	0.142	1.000	0.141	0.832	0.371	0.869	0.376	0.921	0.299
	$\boldsymbol{x}_4$	0.000	0.031	0.026	0.045	0.066	0.104	0.054	0.056	0.064	0.073	0.125	0.141	0.107
	$\boldsymbol{x}_5$	0.000	0.032	0.033	0.043	0.062	0.110	0.059	0.035	0.017	0.031	0.048	0.093	0.039
	$\boldsymbol{x}_6$	0.000	0.029	0.012	0.026	0.045	0.116	0.039	0.047	0.084	0.035	0.103	0.109	0.108
	$\boldsymbol{x}_7$	0.000	0.034	0.019	0.047	0.055	0.127	0.051	0.050	0.034	0.033	0.056	0.117	0.067
	$\boldsymbol{x}_8$	0.000	0.028	0.014	0.042	0.046	0.130	0.064	0.047	0.055	0.037	0.070	0.110	0.091
	$\boldsymbol{x}_9$	0.000	0.033	0.029	0.036	0.059	0.111	0.057	0.052	0.056	0.039	0.118	0.117	0.090
	ρ			0.122		0.125		0.117		0.058		0.057		0.054
	σ^2			0.366		0.357		0.337		0.956		0.957		0.790
	time			1.573		1.656		12.637		1.550		1.632		12.623
$n = 1,000$	$\boldsymbol{\iota}_n$	0.500	1.000	0.063	1.000	0.063	1.000	0.062	1.000	0.081	1.000	0.082	1.000	0.082
	$\boldsymbol{x}_1$	0.300	0.993	0.052	1.000	0.048	0.998	0.049	0.873	0.110	0.950	0.089	0.944	0.084
	$\boldsymbol{x}_2$	1.000	1.000	0.044	1.000	0.044	1.000	0.045	1.000	0.069	1.000	0.069	1.000	0.068
	$\boldsymbol{x}_3$	-0.900	1.000	0.046	1.000	0.046	1.000	0.046	1.000	0.074	1.000	0.074	1.000	0.074
	$\boldsymbol{x}_4$	0.000	0.012	0.006	0.050	0.020	0.044	0.015	0.016	0.004	0.022	0.013	0.045	0.015
	$\boldsymbol{x}_5$	0.000	0.011	0.003	0.038	0.014	0.040	0.007	0.020	0.009	0.056	0.033	0.044	0.018
	$\boldsymbol{x}_6$	0.000	0.011	0.002	0.047	0.017	0.046	0.008	0.016	0.005	0.060	0.032	0.055	0.017
	$\boldsymbol{x}_7$	0.000	0.011	0.004	0.032	0.015	0.040	0.012	0.014	0.004	0.037	0.025	0.045	0.016
	$\boldsymbol{x}_8$	0.000	0.010	0.002	0.032	0.016	0.037	0.011	0.024	0.024	0.058	0.039	0.056	0.029
	$\boldsymbol{x}_9$	0.000	0.008	0.002	0.030	0.014	0.036	0.006	0.013	0.005	0.031	0.017	0.034	0.010
	ρ			0.038		0.038		0.037		0.011		0.011		0.011
	σ^2			0.105		0.105		0.102		0.219		0.222		0.217
	time			9.887		11.813		28.286		10.853		13.067		30.223

SNR stands for the signal-to-noise ratio, RMSE for the root mean squared error, and PIP for the posterior inclusion probability.

Table 2: Results of Monte Carlo experiments for $SNR = 0.9$, averaged over 1,000 simulation runs

$SNR = 0.9$														
			$\rho = 0.5$						$\rho = 0.9$					
			expanded SAR		SAR SSVS		SAR BMA		expanded SAR		SAR SSVS		SAR BMA	
	Variable	true	PIP	RMSE	PIP	RMSE	PIP	RMSE	PIP	RMSE	PIP	RMSE	PIP	RMSE
$n = 100$	$\boldsymbol{\iota}_n$	0.500	1.000	0.092	1.000	0.091	1.000	0.084	1.000	0.261	1.000	0.246	1.000	0.210
	$\boldsymbol{x}_1$	0.300	0.953	0.078	0.967	0.076	0.996	0.058	0.741	0.155	0.817	0.143	0.894	0.113
	$\boldsymbol{x}_2$	1.000	1.000	0.054	1.000	0.054	1.000	0.054	1.000	0.093	1.000	0.091	1.000	0.092
	$\boldsymbol{x}_3$	-0.900	1.000	0.051	1.000	0.049	1.000	0.049	1.000	0.082	1.000	0.082	1.000	0.080
	$\boldsymbol{x}_4$	0.000	0.013	0.003	0.031	0.012	0.124	0.017	0.018	0.007	0.032	0.024	0.126	0.028
	$\boldsymbol{x}_5$	0.000	0.015	0.012	0.070	0.025	0.121	0.022	0.016	0.005	0.027	0.013	0.112	0.020
	$\boldsymbol{x}_6$	0.000	0.011	0.003	0.044	0.021	0.118	0.018	0.013	0.003	0.032	0.015	0.118	0.020
	$\boldsymbol{x}_7$	0.000	0.012	0.007	0.045	0.025	0.135	0.024	0.022	0.013	0.040	0.031	0.119	0.033
	$\boldsymbol{x}_8$	0.000	0.012	0.004	0.026	0.013	0.117	0.018	0.019	0.007	0.041	0.034	0.120	0.032
	$\boldsymbol{x}_9$	0.000	0.014	0.005	0.036	0.016	0.116	0.018	0.028	0.027	0.055	0.037	0.140	0.039
	ρ			0.068		0.066		0.061		0.038		0.035		0.031
	σ^2			0.040		0.039		0.040		0.098		0.100		0.095
	time			1.545		1.629		12.448		1.531		1.618		12.428
$n = 1,000$	$\boldsymbol{\iota}_n$	0.500	1.000	0.025	1.000	0.024	1.000	0.025	1.000	0.047	1.000	0.048	1.000	0.048
	$\boldsymbol{x}_1$	0.300	1.000	0.016	1.000	0.017	1.000	0.017	1.000	0.027	1.000	0.026	1.000	0.026
	$\boldsymbol{x}_2$	1.000	1.000	0.015	1.000	0.015	1.000	0.015	1.000	0.028	1.000	0.028	1.000	0.028
	$\boldsymbol{x}_3$	-0.900	1.000	0.019	1.000	0.019	1.000	0.018	1.000	0.027	1.000	0.027	1.000	0.027
	$\boldsymbol{x}_4$	0.000	0.002	0.000	0.035	0.005	0.037	0.002	0.012	0.008	0.050	0.013	0.051	0.009
	$\boldsymbol{x}_5$	0.000	0.003	0.000	0.048	0.006	0.042	0.003	0.015	0.007	0.061	0.014	0.055	0.010
	$\boldsymbol{x}_6$	0.000	0.003	0.000	0.029	0.003	0.033	0.002	0.005	0.001	0.039	0.009	0.042	0.006
	$\boldsymbol{x}_7$	0.000	0.006	0.001	0.062	0.009	0.053	0.006	0.007	0.002	0.036	0.007	0.037	0.003
	$\boldsymbol{x}_8$	0.000	0.003	0.000	0.044	0.007	0.040	0.003	0.012	0.010	0.037	0.012	0.047	0.011
	$\boldsymbol{x}_9$	0.000	0.002	0.000	0.034	0.006	0.036	0.003	0.005	0.001	0.039	0.010	0.040	0.005
	ρ			0.019		0.019		0.019		0.007		0.008		0.007
	σ^2			0.013		0.013		0.013		0.030		0.030		0.030
	time			9.890		11.832		29.517		9.861		11.841		28.215

SNR stands for the signal-to-noise ratio, RMSE for the root mean squared error, and PIP for the posterior inclusion probability.

proaches when it comes to estimating the spatial disturbance parameter σ^2 and the spatial autoregressive parameter ρ.

In terms of computational efficiency, the results highlight the striking performance of the two variable selection approaches over Bayesian model averaging. Depending on the sample size, the two candidate approaches are computationally faster than Bayesian model averaging by a factor of three to eight. Since both variable selection approaches can be easily implemented in a Gibbs sampler, it is worth noting that these gains in computational efficiency are likely to become even larger when more flexible model specifications are considered as part of the model space.

4 Conclusions

Bayesian model averaging and variable selection methods are potentially useful instruments to deal with specification uncertainty in the presence of spatially autocorrelated data. While Bayesian model averaging techniques have been widely applied for spatial autoregressive models, Bayesian variable selection methods have received little attention thus far. Filling this gap in the existing literature, we compare one Bayesian model averaging and two Bayesian stochastic search variable selection methods for spatial autoregressive models using Monte Carlo simulations. Our results indicate that the existing methods aimed at assessing model uncertainty in the class of SAR models perform very well in terms of identifying important covariates in small samples. The proposed variable selection procedures, however, appear superior to model averaging techniques when it comes to excluding non-important explanatory variables. Moreover, both variable selection approaches provide very sizeable improvements in terms of computational time.

Given the computational advantage of the variable selection methods, these appear best suited to approach uncertainty in model spaces which allow for more flexibility in the specification of spatial spillovers or in the functional forms applied to the potential covariates or spatial spillovers. These approaches to model uncertainty might include multiple weight matrices (see Elhorst et al. 2012), parameterized weight matrices (Vega and Elhorst 2013), heteroskedastic disturbances (LeSage 1997) or dynamic space-time panel data models (Parent and LeSage 2012). Our results support Bayesian variable selection techniques as a central tool for further developments towards making inference in the presence of spatial autocorrelation as flexible and robust as possible to model uncertainty.

References

Brock W and Durlauf S (2001) Growth empirics and reality. *World Bank Economic Review* 15(2), 229–272

Crespo Cuaresma J, Doppelhofer G and Feldkircher M (2014) The determinants of economic growth in European regions. *Regional Studies* 48(1), 44–67

Crespo Cuaresma J and Feldkircher M (2013) Spatial filtering, model uncertainty and the speed of income convergence in Europe. *Journal of Applied Econometrics* 28(4), 720–741

Debarsy N, Ertur C and LeSage JP (2012) Interpreting dynamic space–time panel data models. *Statistical Methodology* 9(1), 158–171

Elhorst JP, Lacombe DJ and Piras G (2012) On model specification and parameter space definitions in higher order spatial econometric models. *Regional Science and Urban Economics* 42(1), 211–220

Feldkircher M and Zeugner S (2009) Benchmark priors revisited: On adaptive shrinkage and the supermodel effect in Bayesian model averaging. *IMF Working Paper* WP/09/202

Fernández C, Ley E and Steel MF (2001) Benchmark priors for Bayesian model averaging. *Journal of Econometrics* 100(2), 381–427

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

George EI and McCulloch RE (1997) Approaches for Bayesian variable selection. *Statistica Sinica* 7(2), 339–373

George EI, Sun D and Ni S (2008) Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics* 142(1), 553–580

Griffith DA (2000a) A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems* 2, 141–156

Griffith DA (2000b) Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses. *Linear Algebra and its Applications* 321, 95–112

Han X and Lee Lf (2013) Bayesian estimation and model selection for spatial Durbin error model with finite distributed lags. *Regional Science and Urban Economics* 43(5), 816–837

Koch W and LeSage JP (2015) Latent multilateral trade resistance indices: Theory and evidence. *Scottish Journal of Political Economy* 62(3), 264–290

Koop G (2003) *Bayesian Econometrics*. Chichester: John Wiley & Sons

Koop G and Korobilis D (2010) *Bayesian multivariate time series methods for empirical macroeconomics*. Now Publishers Inc

Kuo L and Mallick B (1998) Variable selection for regression models. *Sankhyā: The Indian Journal of Statistics, Series B* 60(1), 65–81

LeSage J (2007) A matrix exponential spatial specification. *Journal of Econometrics* 140, 190–214

LeSage J and Pace R (2009) *Introduction to Spatial Econometrics*. Boca Raton: CRC

Press

LeSage JP (1997) Bayesian estimation of spatial autoregressive models. *International Regional Science Review* 20(1-2), 113–129

LeSage JP and Fischer M (2008) Spatial growth regressions, model specification, estimation, and interpretation. *Spatial Economic Analysis* 3(3), 275–304

LeSage JP and Parent O (2007) Bayesian model averaging for spatial econometric models. *Geographical Analysis* 39(3), 241–267

Ley E and Steel MF (2009) On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regressions. *Journal of Applied Econometrics* 24(4), 651–674

Liang F, Paulo R, Molina G, Clyde M and Berger J (2008) Mixtures of g Priors for Bayesian Variable Selection. *Journal of the American Statistical Association* 103, 410–423

Madigan D, York J and Allard D (1995) Bayesian graphical models for discrete data. *International Statistical Review* 63(2), 215–232

Pace RK, LeSage JP and Zhu S (2011) Spatial dependence in regressors and its effect on estimator performance. *Available at SSRN 1801241* , 1–34

Parent O and LeSage JP (2012) Spatial dynamic panel data models with random effects. *Regional Science and Urban Economics* 42(4), 727–738

Piribauer P and Fischer MM (2014) Model uncertainty in matrix exponential spatial growth regression models. *Geographical Analysis* , doi: 10.1111/gean.12057

Raftery AE (1995) Bayesian model selection in social research. *Sociological Methodology* 25, 111–163

Shiller RJ (1973) A distributed lag estimator derived from smoothness priors. *Econometrica* 41(4), 775–788

Vega SH and Elhorst JP (2013) On spatial econometric models, spillover effects, and W. In *53rd ERSA Congress, Palermo, Italy*

Zellner A (1986) *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, chapter On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions. North-Holland: Amsterdam